MULTIVALENT pIII PHAGE DISPLAY LIBRARIES: SELECTED ISSUES AND APPLICATIONS

by

DANIEL ROSS WILSON

B.Sc., University of Guelph, 1990

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

Department of Microbiology and Immunology
and the Biotechnology Laboratory

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 1997

© Daniel Ross Wilson, 1997

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Microbiology and Immunology

The University of British Columbia
Vancouver, Canada

Date August 18, 1997

DE-6 (2/88)

Abstract

By the use of model Ff phage libraries displaying (i) variants of the *Plasmodium falciparum* circumsporozoite protein (CSP) immunodominant tetrapeptide repeats and (ii) peptides derived from the gene encoding *Bordetella pertussis* filamentous hemagglutinin (FHA), the application potential of phage display was illustrated and several issues related to the utility of phage display were examined. Studies with the *P. falciparum* model, together with a statistical analysis of conformations adopted by aspartate/asparagine-proline residue pairs in proteins deposited in the Brookhaven Protein Data bank, provided insight into the structure of the CSP repeats and into the epitopes recognized by a panel of anti-CSP monoclonal antibodies. These studies also served to demonstrate the technical simplicity and general utility of phage display when applied to simple models.

The more complex model, involving phage display of *B. pertussis* FHA-derived peptides, served as a model for genomic libraries. Work with this model led to the identification of fourteen antigenic regions within the large (200 kDa) FHA protein; these findings were both confirmed by and greatly extend previous antigenic analyses of FHA. This model also demonstrated that while construction of genomic phage display libraries is technically achievable, limitations imposed by phage and host biology may limit their utility and broad applicability. Difficulties encountered while exploring this model indicate that vector design is important to library construction and employment, and that attempts to display some foreign peptides on the phage surface can so profoundly affect phage/host biology that in some cases incompatible peptides will be "corrected" so as to enable their display.

These and other issues of phage/host biology and their impact on phage display are discussed, including the need for fundamental studies into the influence of the nature (composition and sequence) of a peptide on its ability to be incorporated into virions.

Table of Contents

Abstract ii

Table of Contents iii

List of Tables x

List of Figures xi

Acknowledgements xiv

Chapter 1 - Introduction 1

1.1. OVERVIEW 1

1.2. FF PHAGE BIOLOGY AND EARLY PHAGE DISPLAY 5

1.2.1. Overview 5

1.2.2. fd-tet and other *ori(-)* mutants 8

1.2.3. pIII 9

1.2.4. Development of pIII-based phage display 10

1.2.5. The first random peptide libraries 11

1.3. ROLE OF THE HOST CELL 12

1.3.1. Signal peptides 12

1.3.2. Other aspects of membrane insertion/translocation 15

1.3.3. Genomic libraries, membrane anchors and secondary signal sequences 16

1.3.4. Codon usage and context 17

1.3.5. Molecular chaperones and export targeting 18

1.3.6. Proteases 18

1.4. IMPACT OF PHAGE AND HOST BIOLOGY ON PHAGE DISPLAY 19

1.5. APPLICATIONS AND INNOVATIONS IN PHAGE DISPLAY 20

1.5.1. Random peptide libraries 20

1.5.2. Phage-displayed proteins and alternative methods of display 25

1.5.3. "Directed evolution of a protein" 29

1.5.4. Gene fragment libraries 29

1.5.5. Innovations in peptide presentation 32

1.5.6. Innovations in identification of target clones 34

1.6. THE PRESENT STUDY 35

1.6.1. Project goals 35

1.6.2. Project accomplishments in context of the concurrent work of others 36

1.6.3. Thesis layout 37

Chapter 2 - Materials and Methods 38

2.1. BACTERIAL STRAINS, PLASMIDS AND BACTERIOPHAGE 38

2.1.1. Culture of bacteria 38

2.2. General Methods 41
2.2.1. Common reagents 41
2.2.2. DNA manipulation 41
2.2.3. Agarose gel electrophoresis 43
2.2.4. DNA quantification 43
2.2.5. Oligonucleotides 43
2.3. DNA Extraction From Virion Hosts and Virions 44
2.3.1. RF DNA extraction from host cells 44
2.3.2. S.s. DNA extraction from virions 44
2.4. Sequencing 45
2.4.1. Sequencing primers 45
2.4.2. Sequencing reactions and gels 45
2.5. Host Cell Transformation and Infection 46
2.5.1. Calcium chloride transformation 46
2.5.2. Electroporation 47
2.5.3. Infection of F-piliated host cells with virions 48
2.6. Harvesting of Virions 49
2.6.1. Harvesting virions from bacterial growth on solid media 49
2.6.2. Separation of virions in culture or wash supernatant from cells by centrifugation 49
2.6.3. One-stage PEG precipitation 49
2.6.4. Two-stage PEG precipitation 50
2.6.5. CsCl density gradient centrifugation 50
2.7. Virion Quantification and Analysis 51
2.7.1. Quantification of virion particles by ultraviolet (UV) spectroscopy 51
2.7.2. Quantification of plaque-forming units (pfu) and transducing units (TU) . . . 51
2.7.3. Agarose gel electrophoresis of virions 53
2.8. Vector Construction 54
2.8.1. Conventional vector fDRW20 54
2.8.2. Conventional vectors fDRW21 and fDRW22 55
2.8.3. Amber vector fDRW5 55
2.8.4. Amber vector fDRW613 56
2.8.5. Amber vector fDRW613C 59
2.8.6. Amber vector fDRW70 59
2.8.7. Conventional vector series fDRW8*nn* 60
2.9. Vector Stability Issues 61
2.9.1. Initial studies of frameshift vectors 61
2.9.2. Evaluation of *E. coli* host strains for fDRW5 propagation (Table 7-I) 64
2.9.3. Comparative analysis of frameshift and amber vector stability (Figure 7-15) 64

2.9.4. Effects of SupF and wild-type pIII on pseudorevertant production 66
2.9.5. Followup analysis of pseudorevertants and comparison of fUSE5 with fDRW5 (Figures 7-24 to 7-26) 67

2.10. CONSTRUCTION OF A LIBRARY OF VARIANT PEPTIDES DERIVED FROM THE *P. FALCIPARUM* CIRCUMSPOROZOITE PROTEIN (CSP-LIBRARY) 69
2.10.1. Construction of a clone displaying PNANPNANPNA 69
2.10.2. Construction of a first CSP-library 69
2.10.3. Unsuccessful construction of a second, proline-biased CSP-library 71
2.10.4. Successful construction of a third, proline-biased CSP-library (Table 7-II) . . 72

2.11. *B. PERTUSSIS FHAB* RESTRICTION FRAGMENT LIBRARIES 73
2.11.1. Construction and analysis of *B. pertussis fhaB Sau*3AI fragment (FHA-S) libraries 73
2.11.2. Construction and analysis of *B. pertussis fhaB Hha*I/*Hin*P1I/*Hpa*II restriction fragment (FHA-H/H/H) libraries (Figures 7-17 to 7-22) 76
2.11.3. Followup study of FHA-H/H/H library clones (Figures 7-21 and 7-22) 81

2.12. CONSTRUCTION OF *B. PERTUSSIS FHAB* DNASE I FRAGMENT LIBRARIES 84
2.12.1. Construction and analysis of FHA series 70 libraries (Table 7-V) 84
2.12.2. Construction of FHA series 80 libraries (Table 7-VI) 86

2.13. ANTIBODIES AND IMMUNOLOGICAL REAGENTS 87
2.13.1. Rabbit pAbs produced against wild-type phage f1 87
2.13.2. Mouse mAbs produced against *P. falciparum* CSP 88
2.13.3. Rabbit pAbs produced against *B. pertussis* FHA 89
2.13.4. Secondary antibodies 92
2.13.5. Immunochemicals, recombinant CS protein and related material 92

2.14. BIOPANNING (AFFINITY SELECTION OF TARGET CLONES) 93
2.14.1. Biopanning random peptide libraries with α-CSP mAbs 93
2.14.2. Biopanning FHA-70 and -80 libraries (Figure 7-30) 94

2.15. PLAQUE LIFTS AND RELATED METHODS EMPLOYED TO IDENTIFY ANTIBODY-REACTIVE CLONES IN FHA-70 AND -80 LIBRARIES 96
2.15.1. Methodology issues 96
2.15.2. Identification of antibody-reactive clones in FHA-70 libraries and assessment of the effect of library amplification on the fraction of antibody-reactive clones recovered 98
2.15.3. Identification of antibody-reactive clones in FHA library 80-A 102
2.15.4. Large-scale propagation of selected FHA-70 and -80 clones 103
2.15.5. Assessment of biopanning enrichment (Figure 7-31) 105

2.16. ELISA METHOD "A": IMMOBILIZED VIRIONS. 106
2.16.1. Methodology 106
2.16.2. ELISA of CSP-library clones (Chapter 4) 107

2.16.3. Preliminary ELISA of random peptide library clones with α-CSP mAbs . . 109
2.16.4. Initial ELISA and BCA assays of FHA-70 and -80 clones (Figures 6-3 and 6-4) 110
2.16.5. Competition ELISA and related BCA assays of FHA-70 and -80 clones . . 112

2.17. ELISA METHOD "B": IMMOBILIZED ANTIBODIES 113
2.17.1. Capture ELISA of selected CSP-library clones (Figure 4-7) 113
2.17.2. Capture ELISA of random peptide library biopan eluates (Figure 5-2) 114

2.18. IMMUNOBLOTS 114
2.18.1. Assessment of binding to nitrocellulose of FHA-70 and -80 clones (Figure 6-4) and immunoblots of antibody-reactive clones (Figure 6-6, assay (i)) 114
2.18.2. Immunoblots of antibody-reactive FHA-70 and -80 clones (Figure 6-6, (ii)) 115

2.19. ANALYSIS OF ASX-PRO TURNS 116

2.20. PREDICTED SIGNAL PEPTIDE CLEAVAGE 116

Chapter 3 - The "Asx-Pro turn" as a local structural motif stabilized by alternative patterns of hydrogen bonds, and a consensus-derived model of the sequence Asn-Pro-Asn 120

3.1. ABSTRACT 120

3.2. INTRODUCTION 120

3.3. ANALYTICAL METHODS 122
3.3.1. Molecular databases and datasets 122
3.3.2. Computer software 124
3.3.3. Hydrogen bonds 124

3.4. RESULTS 129
3.4.1. Residues not preceding proline favor $i \leftarrow i+4$ H-bonds; proline favors $i \leftarrow i+3$ H-bonds 129
3.4.2. Short polar sidechains in residues preceding proline favor $i \leftarrow i+3$ H-bonds 133
3.4.3. Asx-Pro favors a combination of sidechain-backbone and backbone-backbone H-bonds 134
3.4.4. Similar structures are formed by a variety of Asx-Pro H-bond conformations 134
3.4.5. Asx-Pro turns have diverse roles 137
3.4.6. Backbone and sidechain geometry of Asx-Pro turns are not unusual 138
3.4.7. A consensus-derived model for Asn-Pro-Asn agrees with experimental data 139
3.4.8. The proportion of turns *versus* non-turns increases with structure resolution 142

3.5. DISCUSSION 143

Chapter 4 - Recognition of phage-expressed peptides containing Asx-Pro sequences by monoclonal antibodies produced against *Plasmodium falciparum* circumsporozoite protein 145

4.1. ABSTRACT 145

4.2. INTRODUCTION 145

4.3. RESULTS 149

4.3.1. ELISA overview 149

4.3.2. Variant peptides recognized by Group 4 mAbs 152

4.3.3. Similar effects of residue substitutions on recognition by Group 4 mAbs . . 152

4.3.4. Bivalent Pf1B2.2 and Pf5A4.1 *versus* decavalent Pf2F1.1 156

4.4. DISCUSSION 160

4.4.1. Importance of peptide conformation and tetrapeptide cadence 160

4.4.2. Peptide acetylation may confound epitope analysis 161

4.4.3. Is peptide length important? 163

4.4.4. Structural implications of the recognition of variant peptides by Group 4 mAbs 164

4.4.5. Multiple Asx-Pro repeats as immunogens 164

4.4.6. Concluding remarks 165

Chapter 5 - Biopanning random peptide libraries (RPLs) with monoclonal antibodies raised against *Plasmodium falciparum* circumsporozoite protein suggests inherent practical limitations of RPLs 166

5.1. ABSTRACT 166

5.2. INTRODUCTION 166

5.3. RESULTS 168

5.3.1. Progressive enrichment for target clones with each round of biopanning . . 169

5.3.2. Screening of 96 clones yielded only a limited set of sequences 169

5.4. DISCUSSION 173

5.4.1. Practical issue 173

5.4.2. Biopanning results are essentially unpredictable 174

Chapter 6 - Antigenic analysis of *Bordetella pertussis* filamentous hemagglutinin using phage display libraries and rabbit α-FHA polyclonal antibodies 176

6.1. ABSTRACT 176

6.2. INTRODUCTION 176

6.2.1. *Bordetella pertussis* vaccines 176

6.2.2. Role of FHA in pathogenesis 178

6.2.3. Rationale for this study 181

6.3. RESULTS 181

6.3.1. Library construction and evaluation . . . 181
6.3.2. Selection of clones for characterization . . . 182
6.3.3. Sequencing . . . 184
6.3.4. Preliminary ELISA . . . 187
6.3.5. Immunoblots and ELISA . . . 190
6.3.6. Signal peptide cleavage predictions . . . 197

6.4. DISCUSSION . . . 200
6.4.1. Comparison with other studies . . . 200
6.4.2. Candidate immunogens for eliciting protective antibodies . . . 200
6.4.3. Issues related to phage display technology . . . 205
6.4.4. Related studies and concluding remarks . . . 207

Chapter 7 - Issues in vector stability and library construction . . . 209

7.1. ABSTRACT . . . 209

7.2. INTRODUCTION . . . 209
7.2.1. The expected utility of fUSE*n* vectors . . . 209
7.2.2. The outcome . . . 212

7.3. RESULTS . . . 212
7.3.1. Early experimentation with frameshift vectors . . . 212
7.3.2. A library of variants of the *P. falciparum* CSP immunodominant repeats, and development of amber vector fDRW5 . . . 219
7.3.3. A *B. pertussis fhaB Sau*3AI fragment library constructed with conventional vectors . . . 224
7.3.4. Amber vector fDRW613, for *Sau*3AI fragment libraries . . . 228
7.3.5. Comparative assay of frameshift and amber vector pseudorevertant production . . . 230
7.3.6. Is a gene fragment library possible? *B. pertussis fhaB* restriction fragment libraries constructed with frameshift vector fUSE1 . . . 234
7.3.7. A strategy for gene fragment and genomic libraries . . . 243
7.3.8. Design of amber vector fDRW70 . . . 244
7.3.9. *B. pertussis fhaB* DNase I fragment libraries constructed with fDRW70 . . . 250
7.3.10. *B. pertussis fhaB* DNase I fragment libraries constructed with fDRW8*nn* vectors . . . 252
7.3.11. Assessment of FHA-70 and -80 libraries with α-FHA polyclonal antibodies . . . 255

7.4. DISCUSSION . . . 260
7.4.1. General findings . . . 260
7.4.2. Vector issues . . . 262
7.4.3. Concluding remarks . . . 267

Chapter 8 - Summary and General Discussion 269

Literature Cited 274

Appendix 297

List of Tables

Chapter 2

Table 2-I. *E. coli* strains used in these studies. 39
Table 2-II. Plasmids. 40
Table 2-III. Phage display vectors. 41
Table 2-IV. Evaluation of host strains and media for plaque formation by amber vectors. 52
Table 2-V. Virions employed in biopanning FHA-70 and -80 libraries. 95
Table 2-VI. Predicted signal peptidase cleavage of antibody-reactive FHA-70 and -80 clones. 117

Chapter 3

Table 3-I. Effect of proline on backbone hydrogen bonding of preceding residue. . . . 130

Chapter 4

Table 4-I. Summary of α-CSP mAb recognition of CS repeat-derived peptides. 161

Chapter 6

Table 6-I. Peptides displayed by FHA-70 and -80 libraries 181
Table 6-II. Summary of antibody-reactive clones selected from FHA-70 libraries. 182
Table 6-III. Summary of antibody-reactive clones identified in FHA library 80-A by plaque lifts. 183
Table 6-IV. FHA regions grouped by patterns of recognition by antibody 201

Chapter 7

Table 7-I. Evaluation of *E. coli* host strains for fDRW5 propagation 223
Table 7-II. Construction and assessment of third CSP-library. 224
Table 7-III. Characteristics of FHA-S libraries. 226
Table 7-IV. FHA-H/H/H libraries: distribution of *Hha*I/*Hin*P1I/*Hpa*II fragment sizes. . 235
Table 7-V. Construction of *B. pertussis fhaB* DNase I fragment (FHA-70) libraries. . . 250
Table 7-VI. Construction of *B. pertussis fhaB* DNase I fragment (FHA-80) libraries. . . 254

List of Figures

Chapter 1

Fig. 1-1. Wild-type Ff phage and multivalent pIII phage display 2

Chapter 2

Fig. 2-1. Growth curves and estimated doubling times (T_d) of *E. coli* strains 40
Fig. 2-2. fDRW6*nn* amber vector series for constructing *Sau*3AI restriction fragment libraries 57
Fig. 2-3. Flow chart for experiment summarized in Figure 7-15 65
Fig. 2-4. Partial digest of pAS100 with *Sau*3AI 73
Fig. 2-5. Relative mobility (R_f) of *Hae*III fragments of FHA-H/H/H library clones 80
Fig. 2-6. Host cell doubling times (T_d) for FHA-H/H/H library clones 81
Fig. 2-7. Flow chart for followup study of FHA-H/H/H library clones 82
Fig. 2-8. Effect of α-FHA sera purification and biotinylation on recognition of FHA . . . 91
Fig. 2-9. Alternative methods of applying nitrocellulose to lawns 97
Fig. 2-10. Plaque lifts of CSP-library and wild-type f1 virions 98
Fig. 2-11. Plaque lifts of biopanned FHA-80 library clones 99
Fig. 2-12. Evaluation of Immulon plates for ELISA 105
Fig. 2-13. BCA assay for quantification of virions bound to Immulon plate wells: effect of incubation time 107
Fig. 2-14. Titration of recognition of CSP-library clones by α-CSP mAbs 108
Fig. 2-15. Control for BCA assay 113

Chapter 3

Fig. 3-1. Local H-bonds in Asx-Pro sequences, Dataset A 126
Fig. 3-2. Local H-bonds in Asx-Pro-Asx sequences, Dataset C 127
Fig. 3-3. Effect of proline on backbone hydrogen bonding 131
Fig. 3-4. Local H-bonds in Asx-Pro, Ser-Pro and His-Pro sequences 135
Fig. 3-5. Alternative hydrogen bonding patterns in Asx-Pro turns 136
Fig. 3-6. Superimposed wireframe models of Asx-Pro turns 137
Fig. 3-7. Ramachandran plot of ϕ,ψ dihedral angles in Asx-Pro turns and other proline-containing sequences 138
Fig. 3-8. Sidechain χ_1 angles of Asx in Asx-Pro turns 139
Fig. 3-9. Superimposed α-carbon traces of Asx-Pro helix caps and flanking residues . . 139
Fig. 3-10. Local H-bonds in Asx-Pro-Asx sequences 140
Fig. 3-11. Superimposed wireframe models of Asx-Pro-Asx peptides 141
Fig. 3-12. H-bonds in Asn-Pro-Asn sequences that form Asx-Pro turns 142
Fig. 3-13. Influence of structure resolution on identification of Asx-Pro turns 143

Chapter 4

Fig. 4-1. Postulated hydrogen bonds within the sequence Asn-Pro-Asn-Ala 147
Fig. 4-2. Phage-displayed CSP-related peptides 148

Fig. 4-3. ELISA with recombinant CS protein and phage-displayed native peptides . . . 150
Fig. 4-4. ELISA with phage-displayed native peptides and single-residue variants 151
Fig. 4-5A. Group 4 mAb Pf1B2.2 ELISA with phage-displayed native peptides as well as one-, two- and three-residue variants 153
Fig. 4-5B. Group 4 mAb Pf5A4.1 ELISA with phage-displayed native peptides as well as one-, two- and three-residue variants 154
Fig. 4-5C. Group 4 mAb Pf2F1.1 ELISA with phage-displayed native peptides as well as one-, two- and three-residue variants 155
Fig. 4-6. Comparison of ELISA results for Group 4 mAbs 156
Fig. 4-7. Group 4 mAb Pf1B2.2, comparison of ELISA methods A and B 157
Fig. 4-8. Relative quantities of virions bound to microtiter plate wells 158
Fig. 4-9. Effect on ELISA signal of varying the quantity of virions 159
Fig. 4-10. Inhibition of ELISA signal by preincubation of α-CSP mAbs with $(NPNA)_3$ peptide 159

Chapter 5

Fig. 5-1. Overview of 16 biopans 167
Fig. 5-2. Apparent enrichment for target clones with multiple rounds of biopanning . . 168
Fig. 5-3. Initial screen of clones selected from third round of biopanning with Pf2A10 170
Fig. 5-4. Initial screen of clones selected from third round of biopanning with Pf1B2.2 171

Chapter 6

Fig. 6-1. Sequences of antibody-reactive FHA-70 and -80 clones 185
Fig. 6-2. Distribution of antibody-reactive FHA-70 and -80 clones 187
Fig. 6-3. Early ELISA of FHA-70 and -80 clones 188
Fig. 6-4. Uniform binding of FHA-70 and -80 clones to nitrocellulose 190
Fig. 6-5. Example immunoblot of FHA-70 and -80 clones 191
Fig. 6-6. Immunoblots of antibody-reactive FHA-70 and -80 clones 192
Fig. 6-7. Variable binding of FHA-70 and -80 clones to Immulon plates 194
Fig. 6-8. Inhibition of ELISA signal by preincubation of α-FHA antibodies with purified FHA 195
Fig. 6-9. Predicted signal peptide cleavage of antibody-reactive FHA-70 and -80 clones 198
Fig. 6-10. Positional comparison of antibody-reactive clones with results of other studies 199
Fig. 6-11. Sequence overlaps in region I and XI clones 202

Chapter 7

Fig. 7-1. Phage display vectors provided by G. P. Smith 211
Fig. 7-2. Early efforts to prepare linear fragments of fUSE5 RF DNA 213
Fig. 7-3. Early preparations of fUSEn RF DNA 214
Fig. 7-4. Pseudorevertant production by fUSEn frameshift vectors 216
Fig. 7-5. Use of chloramphenicol-containing medium to reduce s.s. DNA species 217
Fig. 7-6. Effects of S1 nuclease and restriction endonucleases on fUSE5 DNA preparations 217
Fig. 7-7. Elution of s.s. DNA with 1.3M NaCl in Qiagen column wash 218

Fig. 7-8. Construction of phage displaying CSP-related peptides 219
Fig. 7-9. Conversion of frameshift vector fUSE5 to amber vector fDRW5 220
Fig. 7-10. fDRW5 RF DNA preparations for constructing CSP-peptide libraries 221
Fig. 7-11. Assessment of clonal bias in first CSP-library 221
Fig. 7-12. Conventional vectors fDRW20, fDRW21 and fDRW22 225
Fig. 7-13. Assessment by agarose gel electrophoresis of FHA-S library virions 226
Fig. 7-14. Design of amber vector fDRW613 and two-amber vector fDRW613C 229
Fig. 7-15. Frameshift and amber vector stability as determined by pseudorevertant production 231
Fig. 7-16. Effects of SupF and wild-type pIII on pseudorevertant production. 234
Fig. 7-17. Screen for virion-producing transformants in FHA-H/H/H libraries 235
Fig. 7-18. Virion production by selected FHA-H/H/H library clones 236
Fig. 7-19. Assessment of influence of insert size on virion production by selected FHA-H/H/H library clones 237
Fig. 7-20. Correlation of virion production with host cell doubling time, selected FHA-H/H/H library clones 238
Fig. 7-21. Virion production by 1° and 2° clones of FHA-H/H/H library members 239
Fig. 7-22. Characterization of selected FHA-H/H/H library 1° and 2° clones 242
Fig. 7-23. Design of amber vector fDRW70 244
Fig. 7-24. Stability of fUSE5 and amber vectors, as determined by pseudorevertant production 245
Fig. 7-25. RF DNA of fUSE5 and fDRW5 propagated in *E. coli* LE392 247
Fig. 7-26. Pseudorevertants of frameshift and amber vectors 248
Fig. 7-27. A screen for virion-producing clones in FHA-70 libraries 251
Fig. 7-28. A novel class of fDRW70 pseudorevertant 252
Fig. 7-29. Design and construction of conventional vector series fDRW8*nn* 253
Fig. 7-30. Output from a single round of biopanning FHA-70 and -80 libraries 256
Fig. 7-31. Effect of biopanning on fraction of antibody-reactive clones recovered from FHA-70 libraries: unenriched library *versus* biopanning output 257
Fig. 7-32. Effect of library amplification on fraction of antibody-reactive clones recovered from FHA-70 libraries 258

Acknowledgements

I have incurred a debt of gratitude to many individuals during my studies and research.

Many in the B. B. Finlay laboratory have provided practical advice and assistance with my project, and I thank them for this. I wish to specifically thank

- Brett Finlay, my supervisor, for the opportunities he provided;
- Annette Siebers, for generous provision of antibodies, DNA fragments and other material that made the FHA study possible;
- Sharon Ruschkowski, for her cheerful commitment to ensuring that materials I required were made available at unreasonably short notice;
- Nancy Carpenter and Mitchell Uh, for indispensible technical assistance that substantially shortened the duration of my project;

as well as my supervisory committee

- R. E. W. Hancock, N. Reiner and R. A. J. Warren, each of whom contributed to the success of my project;

and other individuals, including

- G. P. Smith (University of Missouri) and J. Scott (Simon Fraser University) for bacterial strains, phage vectors and random peptide libraries;
- R. A. Wirtz (Walter Reed Army Institute of Research) for monoclonal antibodies and recombinant circumsporozoite protein;
- U. Hobohm, G. Vriend and R. Hooft (EMBL, Germany), P. Walters and M. Stahl (University of Arizona, U.S.A.) and R. Sayle (University of Edinburgh, Scotland) for computer software;
- L. L. Walsh (University of Illinois, U.S.A.) for her useful "Annotated PDB File Listing";
- L. McIntosh (U. B. C.) for his critical reading of an early draft of material presented in Chapter 3.

The studies reported in Chapter 4 were made possible by materials developed or provided by R. S. Nussenzweig and New York University (mAb Pf2A10), Y. Charoenvit and the Naval Medical Research Institute (mAb PvNSV3) and SmithKline Beecham (R32tet32 and PvNS1v20 recombinant proteins). MAbs used in this study were produced under the United Nations Development Programme/World Bank/WHO Special Programme for Research and Training in Tropical Diseases Grant 880068.

My studies were made possible by a 1967 Science and Engineering Scholarship awarded by the Natural Sciences and Engineering Research Council of Canada. Much of the research was supported by an operating grant to B. B. Finlay from the Canadian Bacterial Diseases Network Centre of Excellence.

Chapter 1

Introduction

1.1. Overview

The publication in 1990 of three studies describing peptides displayed on the surface of recombinant bacteriophage brought to the attention of many scientists a novel, simple and powerful approach to the problem of identifying or optimizing ligands for antibodies and other biomolecules. In these model studies, three groups of researchers had each constructed a vast library of $>10^7$ six- or fifteen-residue random sequence peptides surface-displayed as N-terminal fusions to a minor coat protein (pIII) of an Ff (M13, f1 and fd) bacteriophage (Figure 1-1). By technically simple affinity selection methods, they identified novel sequences that bound specifically to cognate antibodies (Cwirla *et al.* 1990; Scott & Smith 1990) or to streptavidin (Devlin *et al.* 1990).

Several aspects of Ff phage biology made these studies possible and the concept powerful. First, because Ff phage are small and large numbers can be readily accommodated in a small volume ($>>10^{12}$ mL^{-1}), constructing libraries of the required diversity ($>10^7$ unique clones; accomplished by "shotgun" cloning of degenerate oligonucleotides into the 5' region of gIII[1], the gene encoding pIII; Figure 1-1C) was technically achievable. Second, because a virion is structurally simple (Figure 1-1B), the displayed peptides were presented in a comparatively simple (with respect to many expression systems) physicochemical environment, allowing successful affinity selections (called "biopanning" by Scott & Smith 1990) with minimal

---

[1]Commonly, phage proteins are identified with Roman numerals prefixed with a lower case "p" (*e.g.*, pIII, pVIII) while their encoding genes are identified simply with Roman numerals (*e.g.*, "gene III", "gene VIII"). For clarity and convenience, I have employed the prefix "g" to identify phage genes (*e.g.*, "gIII", "gVIII").

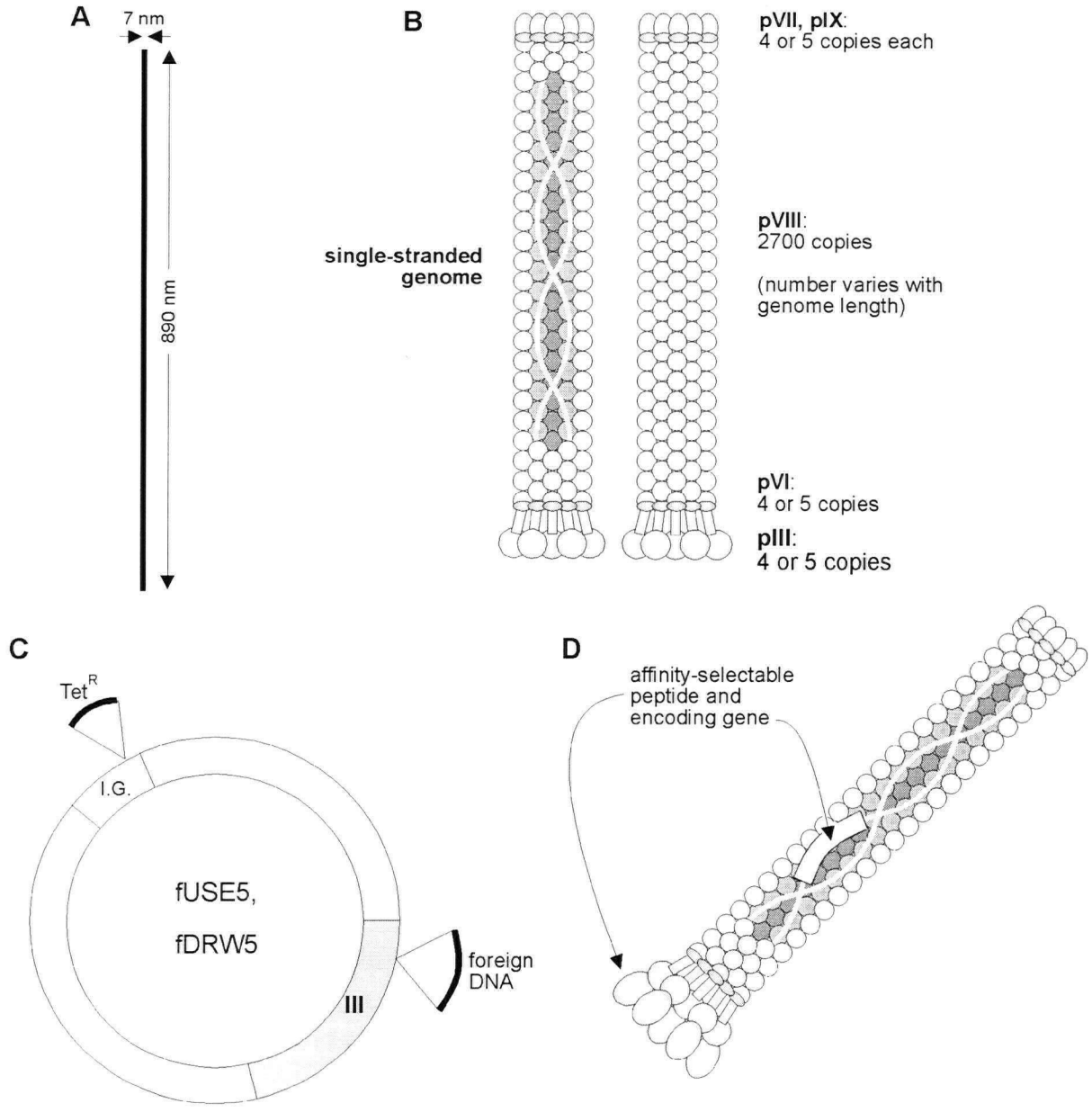


Fig. 1-1. Wild-type Ff phage and multivalent pIII phage display. **A**, **B**. Wild-type filamentous phage that infect F-piliated *Escherichia coli* (Ff phage). **A**. Ff phage are narrow, flexible rods of the indicated dimensions. **B**. Cartoon illustrates inherent simplicity of virion structure: a single-stranded circular genome is surrounded by ~2700 copies of the major coat protein pVIII and 4 or 5 copies of each of four species of minor coat proteins, including pIII which binds to a host cell F-pilus. **C**. Multivalent pIII phage display vectors fUSE5 and fDRW5 and others employed in this study are derivatives of fd-tet (Zacher *et al.* 1980) and accordingly possess a tetracycline-resistance gene (Tet^R) inserted into the major intergenic (I.G.) region. Inserting foreign DNA into gIII, at a point corresponding to the pIII N-terminal signal peptidase cleavage site, results in the display of a foreign peptide as an N-terminal fusion to each mature pIII molecule, as illustrated in **D**. As described in the text, a key feature of phage display libraries is that an affinity selectable target peptide and its encoding DNA are packaged together (**D**).

background. Finally and most importantly, each phage-displayed peptide was physically linked to the gene encoding its production: affinity selection of a target peptide corresponded to cloning its encoding gene (Figure 1-1D). Importantly, the displayed peptides appeared to have little effect on phage biology, either during the processes that lead to incorporation of pIII into a virion, or on the ability of pIII to initiate the infection cycle by adsorbing to the tip of a host cell F-pilus (Parmley & Smith 1988; Scott & Smith 1990). As well, certain technical aspects of the design of some vectors (to be discussed later) appeared to ensure that libraries free of nonrecombinants could be constructed and to allow that large peptides or proteins might be successfully displayed.

Given the power of the concept, the demonstration of its success and the technical merits of certain vectors, it thus seemed possible that phage display libraries (PDLs) could be applied to a variety of research problems in several disciplines, including pathogenesis. Early studies bore this out. Bass *et al.* (1990) soon demonstrated pIII-display of the 191-residue disulfide bond-containing human growth hormone in a properly folded form, while McCafferty *et al.* (1990) showed that single-chain Fv antibodies could be displayed on pIII. Later, Kang *et al.* (1991) showed that functional Fab fragments could be displayed as fusions to the major coat protein pVIII, while Roberts *et al.* (1992a) showed that a library of randomized pIII-displayed variants of bovine pancreatic trypsin inhibitor could be employed to identify a variant possessing $>10^6$-fold greater affinity for human neutrophil elastase than the parental peptide.

It was in context of these early successful studies that my thesis project began. My general goal was that of exploring, by means of relatively simple models the feasibility of exploiting PDLs in the study of pathogenesis, with the eventual goal of constructing phage-displayed genomic libraries that could, given an adequate molecular probe, be used to identify pathogen-encoded proteins with a specific role in pathogenesis.

Concurrently, there has been growth in both numbers and diversity of applications as well as advances in methodology. A sampling of papers published during the past six years shows that epitope characterization (45 papers) and other applications (30 papers) of random peptide libraries (RPLs) are popular. Studies related to display of large polypeptides or proteins, including affinity-maturation studies such as the "directed evolution of a protein" theme of Roberts *et al.* are also common (34 papers). Studies concerning methodology, such as conformational constraint of displayed peptides (17 papers), are more limited. Although the first PDL was constructed with gene fragments (Smith 1985), few studies (seven papers) have explored this idea further.

A number of reviews of phage display have been published, including general overviews (Wells & Lowman 1992; Barbas 1993; Smith & Scott 1993; Clackson & Wells 1994; Perham *et al.* 1995), reviews of RPLs (Scott 1992; Lane & Stephen 1993; Cortese *et al.* 1994, 1995, 1996; Scott & Craig 1994; Burritt *et al.* 1996) and of display of proteins (Dunn 1996), as well as more focussed reviews concerned with "directed evolution" and affinity-maturation themes (O'Neil & Hoess 1995), display of enzymes (Soumillion *et al.* 1994a) and antibodies (Marks *et al.* 1992; Winter *et al.* 1994), design of DNA-binding proteins by use of zinc finger libraries (Choo & Klug 1995) and applications in neurobiology (Bradbury & Cattaneo 1995).

Thus, our initial optimism has been largely justified by successful application of phage display to a variety of biological questions. At the same time, as reviewed later and illustrated in this thesis, it has become clear that not every application attempted is successful and that phage biology may play a larger role in this than initially believed. As with most expression systems, the principal issues of phage display relate to the nature of the displayed foreign peptide or protein and its influence on successful expression and incorporation into virions. Given my goal of assessing the feasibility of constructing phage-display genomic libraries,

which requires the display of diverse collections of heterologous peptides, it becomes especially important to understand the ways in which phage and host cell biology are likely to hinder this objective.

The following sections review filamentous bacteriophage biology, related aspects of host cell biology, and phage display applications with the goals of (i) providing a general overview of these subjects, (ii) illustrating ways in which phage/host biology *might* be expected to hinder phage display, and (iii) providing specific and relevant examples of difficulties experienced by others.

1.2. Ff Phage Biology and Early Phage Display

1.2.1. Overview

M13, f1 and fd[2] (reviewed in Rasched & Oberer 1986; Model & Russel 1988; Russel 1991, 1994,1995) are long, narrow bacteriophages of about 7 nm diameter and 900 nm length (Figure 1-1A). They are collectively called Ff phage because of their filamentous appearance and almost total dependence on the F-pilus for infection of their *E. coli* host. The Ff virion consists of a single-stranded (s.s.) DNA genome of roughly 6400 nucleotides packaged in a tube comprised of ~2700 copies of the major coat protein pVIII and closed at the ends by four or five copies of each of four species of minor coat proteins (Figure 1-1B). While most bacterial viruses assemble in the host cell cytoplasm and are released by cell lysis, Ff phage do not kill their hosts but rather are continuously extruded through the host cell membrane in a process that couples assembly with export. The process is efficient and virion titers commonly exceed 10^{12} mL^{-1} after overnight culture in a suitable host.

The infection cycle begins when pIII (the adsorption protein, seen as a knob-on-stem or

---

[2]Except where noted, material presented in this section is derived from the cited reviews.

"lollipop" in electron micrographs) adsorbs to the tip of a host cell F-pilus. Through a process that is not yet completely understood, the virion is brought to the cell surface, the s.s. genome is delivered to the host cell cytoplasm, and virion proteins are deposited in the cytoplasmic membrane for re-use during subsequent virion production. Immediately after infection, host polymerases employ the s.s. genome (the (+) strand) as a template for synthesis of a complementary or (-) strand, thereby creating a double-stranded (d.s.; or replicative form, RF) phage genome. Synthesis of the (-) strand is initiated at *ori(-)*, one of several sites within the major intergenic region. Because of the d.s. nature of two adjacent stem loops formed at *ori(-)*, the sequence and conformation of *ori(-)* bears some resemblance to the (double-stranded) -35 and -10 regions of RNA promoters; these similarities are important for initiation of primer synthesis by RNA polymerase (Higashitani *et al.* 1997).

Following (+) strand complementation, host-mediated transcription from the plasmid-like RF molecule leads to expression of ten genes encoding the coat proteins (pIII, pVI, pVII, pVIII, pIX), assembly/export proteins (pI and pIV) and proteins involved in DNA replication and its control (pII, pV and pX). As presently understood, the relative levels of these proteins are mainly determined by a mechanism that, while simple in concept, is extraordinarily complex in detail. In simplistic terms, the genome is organized into two main coding regions, one for proteins required in large quantities (such as pVIII) and possessing a relatively strong initial promoter; the other for proteins required in fewer numbers (such as pIII) and possessing a less efficient initial promoter. Multiple promoters within each region create a cascade effect whereby 3' genes are transcribed more frequently than 5' genes. A host mechanism that degrades mRNA transcripts from the 5' end magnifies the cascade effect, with the net result that a finely tuned balance of synthesis of the ten phage proteins is achieved with little apparent feedback control (see Fulford & Model 1988a,1988b for exceptions to this).

Synthesis of the (+) strand is initiated by pII, which nicks the (+) strand at *ori(+)*, located in the major intergenic region, allowing the freed 3' end to serve as a primer for rolling circle replication on the (-) strand template. As the strand is completed, pII ligates the molecular ends to form a closed circle. The fate of new (+) strands varies. Early in infection these serve as templates for the formation of additional RF molecules in the manner described earlier. As the levels of the s.s. DNA binding protein (SSB) pV increase (as a function of time, increases in RF DNA copy number, and consequent increase in RF-directed gV mRNA synthesis), s.s.-to-RF conversion is arrested as newly-synthesized (+) strands become sequestered by pV.

Virion assembly and export take place in the host cell membrane. Following synthesis, all five coat proteins are inserted into the inner membrane (Endemann & Model 1995). While pVI, pVII and pIX lack signal peptides, pIII and pVIII are synthesized with N-terminal signal peptides that are cleaved upon membrane insertion. Anchored by C-terminal hydrophobic regions, both pIII and pVIII are oriented with their N-termini in the periplasm. Membrane insertion of pVIII is Sec-independent (Pugsley 1993; Kuhn 1995), while pIII may require the Sec apparatus (Peters *et al.* 1994). In current models, pI and pIV form a gated, multimeric export channel with pI and pIV localized to the inner and outer membranes respectively. Assembly/export is initiated by interactions among (i) a "packaging" signal within the intergenic region, (ii) export channel protein pI, (iii) host thioredoxin, and (iv) membrane-inserted minor coat proteins pVII and pIX. In a process that requires ATP hydrolysis and in which proton motive force is also important (Feng *et al.* 1997), phage are extruded through the pI/pIV export channel with the concomitant replacement of pV SSB by pVIII. Virion assembly is completed by addition of pVI and pIII to the end of the extruding virion.

1.2.2. fd-tet and other *ori(-)* mutants

Surprisingly, mutants with defects in *ori(-)* are viable. Kim *et al.* (1981) constructed a number of viable M13 mutants containing deletions in *ori(-)*, while Zacher *et al.* (1980; see also Smith 1988) constructed the cloning vector fd-tet (the parent of many phage display vectors) by splicing a 2.8 kbp tetracycline-resistance gene into *ori(-)* at a position that interrupts one of the two stem loops required (Higashitani *et al.* 1997) for primer synthesis. In the M13 mutants and fd-tet, initiation of (-) strand replication is delayed and proceeds slowly by an unidentified alternative mechanism (Kim *et al.* 1981; Smith 1988). Relatively recently, M13 mutants with *ori(-)* defects have been shown to induce the *E. coli* SOS response (Higashitani *et al.* 1992, 1995), presumably because of delays in (-) strand synthesis or related events leading to accumulation of s.s. DNA, the apparent SOS system inducer (Walker 1996). Considering the phenotype similarities of M13 *ori(-)* mutants and fd-tet, it seems likely that fd-tet also induces the SOS response.

Ff mutations that block phage assembly while allowing DNA replication to continue are normally lethal to the host cell because of intracellular accumulation of phage DNA and gene products (Smith 1988). In fd-tet, because of its *ori(-)* defect, the RF DNA copy number is reduced about 13-fold, with the result that while RF DNA levels increase in morphogenetically defective mutants of fd-tet, these levels do not exceed those of wild-type phage (Smith 1988). As described later, Parmley and Smith (1988) exploited the viability of fd-tet mutants in designing phage display vectors.

Compared to wild-type Ff, fd-tet plaques are extremely small (for an example, see Materials and Methods, Figure 2-10) and turbid, virion production after overnight culture is reduced four-fold, and yields of RF and s.s. DNA are reduced more than 10-fold (Smith 1988). M13 *ori(-)* mutants have similar plaque phenotypes and reduced virion yields (Kim *et al.* 1981).

Infectivity (infectious units per physical particle) is reduced in fd-tet to as little as 2.5% of wild-type levels (Smith 1988) although infectivities as high as 50% of wild-type levels have been reported for the fd-tet derivative fUSE2 (Parmley & Smith 1988).

1.2.3. pIII

pIII is required not only for its role in F-pilus adsorption but also for terminating virion assembly and stabilization of the virion particle. Delays in supplying pIII during assembly/export lead to production of multilength virion particles (polyphage) containing two or more unit-length phage genomes (Model & Russel 1988). Supernatants of wild-type Ff phage routinely contain about 5% polyphage.

The knob-on-stem appearance of pIII in electron micrographs reflects its two functional domains. The N-terminal two-thirds of the molecule mediates virion adsorption while the C-terminal one-third functions in virion assembly and structure (Armstrong *et al.* 1981; Nelson *et al.* 1981; Crissman & Smith 1984; Model & Russel 1988). Thus, mutants unable to produce at least the C-terminal portion of pIII produce virions containing multiple unit-length genomes (polyphage), while mutants producing pIII that lacks 169 residues from the middle of the N-terminal domain (mature pIII contains 406 residues) produce normal numbers of unit-length virions (monophage) but are non-infectious (Crissman & Smith 1984). pIII contains two sets of repeats of GGGSE (in one-letter amino acid code) or related variants. The second and longer of these repeats occurs about two-thirds (Beck & Zink 1981) from the N-terminus of the mature protein. Given its high content of glycine and polar residues (deemed important in linker design, as reviewed in Alfthan *et al.* [1995]) and its position (which matches roughly with the estimated size of the N-terminal domain), it seems likely that these repeats serve as a flexible linker connecting the N-terminal knob and C-terminal stem.

Ff phage-infected cells exhibit a number of membrane-associated phenotypes including

accumulation of intracellular membranes and increased leakage of periplasmic proteins. Many of these effects are mediated by pIII, especially by its N-terminal 98 amino acids (Boeke *et al.* 1982; Rampf *et al.* 1991). pIII additionally mediates resistance to Ff superinfection and several other phenotypes (Boeke *et al.* 1982).

1.2.4. Development of pIII-based phage display

The first PDL was a gene fragment library constructed by Smith (1985), who cloned fragments of *Eco*RI endonuclease and methylase enzymes genes into a natural *Bam*HI site of gIII of wild-type phage f1, creating a small library of virions displaying peptides in the middle of pIII. Although identification of a single clone possessing a 171-bp insert reactive with α-*Eco*RI antibody demonstrated that "phage display" held promise, several problems related to phage biology required attention. Plaques of the identified clone were very small or barely detectable, the stationary phase titer was 100 times lower than wild-type levels, infectivity was 25-fold lower, the incidence of polyphage was higher, and recombinants in which the original insert was lost or altered were found to possess a marked selective advantage during virion propagation.

Later, Parmley and Smith (1988) showed that it was possible to affinity select rare target virions displaying peptides of interest from a "mock library"containing a 10^8-fold excess of virions lacking the target peptide. Their success derived in part from improvements in vector design that reflected several aspects of phage biology. First, the cloning site was moved from mid-gIII to a position corresponding to the N-terminus of mature pIII, with the expectations that displayed peptides would interfere less with pIII adsorption function and enjoy greater conformational freedom. Secondly, fd-tet (*versus* wild-type f1) was employed as a parent for new vectors, with the expectation that the reduced gene expression and virion production characteristic of fd-tet (Smith 1988) would diminish the host-toxicity of difficult-to-export

peptides (Smith & Scott 1993). Finally, knowing from earlier studies that pIII-deficient mutants of fd-tet were viable, Parmley and Smith developed a "frameshift" vector (fUSE1) in which a gIII frameshift at the cloning site prevented production of pIII and thus of infectious virions, unless gIII possessed a foreign insert able to restore the gIII reading frame. This would allow construction of libraries where each infectious virion could be considered a recombinant, a clear advantage (i) in eliminating overgrowth of nonrecombinants during library propagation, and (ii) assessing cloning success and the "completeness" of a library.

Some problems remained. The already-low infectivity of fd-tet (Smith 1988) was further reduced in certain fUSE1 recombinants. Thus, while a 20-bp insert (derived from a synthetic T7 RNA polymerase promoter sequence) had no apparent effect on infectivity, 54-bp (*P. falciparum* CSP repeats) and 71-bp inserts (from *E.coli lacZ*) reduced infectivity by 50% (compared to the non-frameshift vector fUSE2), while a 335-bp insert (from *lacZ*) reduced infectivity 20-fold.

1.2.5. The first random peptide libraries

The three earlier-mentioned RPL applications of Cwirla *et al.* (1990), Devlin *et al.* (1990) and Scott and Smith (1990) validated Parmley and Smith's (1988) proposal that large libraries of phage-displayed peptides could be used to identify epitopes, mimotopes (Geysen *et al.* 1987) or other mimetic peptides. Some of this success would appear to derive from continued consideration of phage biology. Thus, Scott and Smith, in constructing their RPL of 4×10^7 hexapeptides (~69% of the 20^6 possible sequences), employed a new vector (fUSE5) that incorporated several features intended to facilitate peptide display. To minimize problems in signal peptidase processing of recombinant pIII, the first two residues following the signal peptide cleavage site were retained (with Asp substituting for the normal Glu in Ala-Glu) in their clones; these were followed by a random hexapeptide sandwiched between Gly-Ala and

Gly-Ala-Ala-Gly-Ala "linker" peptides. Other features of fUSE5, including a gIII frameshift, ensured that non-recombinant virions were excluded from their library. Cwirla *et al.* (1990) also used an fd-tet-derived frameshift vector to construct their hexapeptide RPL; curiously, they inserted random peptides immediately following the signal peptide cleavage site. In the RPL of Devlin *et al.* (1990), constructed without a frameshift vector, 15-residue peptides were separated from the signal peptide cleavage site by Ala-Glu and from the remainder of pIII by Pro_6.

In spite of these and the previously-noted early successes with display of proteins such as human growth hormone, bovine pancreatic trypsin inhibitor variants and antibodies, questions remained concerning whether phage display would find broad applicability. Indeed, a review of the issues involved might predict that many applications might not succeed, being limited by the biology of the host cell.

1.3. Role of the Host Cell

As with other applications involving production of engineered proteins, peptide or protein display by Ff phage requires host cell processes for both synthesis and export of recombinant pIII and pVIII molecules. Thus the same cellular processes that sometimes interfere with recombinant technology generally can be expected to interfere with phage display. Importantly, phage display involves N-terminal fusions to phage coat proteins pIII or pVIII, and a number of studies have shown that the N-terminal mature region of a protein plays a role in its export.

1.3.1. Signal peptides

Both pIII and pVIII are synthesized with signal peptides. Indeed, much of our understanding of signal peptide processing has derived from studies of this process in pVIII.

Both processing by signal peptidase and subsequent virion production are affected in some pVIII mutants in which residues *following* the cleavage position -- *i.e.*, residues +1, +2, ..., of the mature protein -- are altered. Thus, while pVIII preprotein cleavage is efficient when Ala at position +1 is replaced with Phe, Leu, Val, Gly or Ser, cleavage is reduced by substituting Cys and abolished by Pro or Thr. Glu→Leu and Glu→Tyr substitutions at position +2 also retard pVIII processing and, in the case of Leu, affect the kinetics of virion production (Boeke *et al.* 1980; Russel & Model 1981). Although most residue substitutions in positions +1 to +5 have no apparent effect on virion production (Iannolo *et al.* 1995; Williams *et al.* 1995), certain substitutions (most notably by Tyr) can reduce virion titers dramatically, to $<10^2$ mL^{-1} culture supernatant (Iannolo *et al.* 1995). Whether this was due to hindered signal peptidase activity was not determined.

Early difficulties (*e.g.*, Greenwood *et al.* 1991; Felici *et al.* 1991; see also Perham *et al.* 1995) in employing pVIII for multivalent display of peptides larger than six amino acids raised the question of whether this was due to issues such as (i) a "size" restriction related to virion assembly (*e.g.*, steric interference in pVIII-pVIII interactions) or export through the pI/pIV channel, or (ii) issues related to pVIII synthesis and processing. Using X-ray diffraction studies and model building, Malik *et al.* (1996) have recently shown that it is theoretically possible to display much larger structures on pVIII than it has been thus far possible to accomplish, suggesting that a size restriction of practical significance does not exist. Importantly, they also showed that (for the small number of recombinants analysed) the degree of success in display of peptides on pVIII correlated better with efficiency of preprotein processing by signal peptidase, rather than with the size of the displayed peptide.

Studies of the N-terminal mature region of pIII are limited to a single report (Peters *et al.* 1994), showing that (i) that positively charged (Arg) residues within the N-terminal decamer

of an inserted peptide reduced virion titers in a dose-related manner from ~10^{10} mL^{-1} to as little as 10^3 mL^{-1}, and (ii) that *prlA* (*secY*) suppressors could restore virion production of most defective clones 10^6- to 10^7-fold to wild-type levels. Charged residues at the N-terminal mature region have been shown to hinder translocation of other recombinant proteins (e.g., Li *et al.* 1988; Yamane & Mizushima 1988). It has been suggested (i) that these disrupt a net dipole formed around the hydrophobic domain of the signal peptide and required for efficient translocation (reviewed in Boyd & Beckwith 1990; Izard & Kendall 1994) and (ii) that clusters of these residues may, by conforming to the "positive inside" rule derived by von Heijne (1994) from study of periplasmic and cytoplasmic loops of membrane proteins, serve to convert a signal peptide to a membrane anchor.

The demonstrated role of the N-terminal mature region is apt to be important in phage display, for peptides displayed in nearly all RPLs are entirely contained within a postulated N-terminal "critical 30-residue export initiation domain" (Andersson & von Heijne 1991; Nielsen *et al.* 1997; also see Summers & Knowles 1989). Notably, the popular display vector fAFF1 (Cwirla *et al.* 1990) allows foreign peptide insertion at, rather than a few residues away from (as in fUSE5, Scott & Smith 1990), the +1 position. Although Burritt *et al.*(1996) have suggested that there is no *compelling* evidence that this is problematic, data reviewed above suggest otherwise. Moreover, a survey by Nielsen *et al.* (, 1997 #196; see also von Heijne 1986) of prokaryotic signal peptides and flanking sequences reveals a marked positional distribution of amino acids around the signal peptidase cleavage site, including not only amino acid preferences in preprotein positions (-1, -2, ...) but also in the first positions of the mature protein. Thus, Pro is relatively common in positions -6 to -4 and +2 to +6 but rare (mol% <0.5[3]) in positions -3 to +1. Ala

[3]summary data derived from raw data employed in the study by Nielsen *et al.* and provided by the authors via their FTP server at ftp://virus.cbs.dtu.dk/pub/signalp.

(mol%=38.3) is the most common residue at +1, while Asp (16.2%) and Glu (15.0%) predominate at +2. Notably, these or similar residue pairs initiate mature pIII of wild-type Ff phage (Ala-Glu) and of recombinants in two frequently employed RPLs, those of Devlin *et al.* (1990, Ala-Glu) and Scott and Smith (1990, Ala-Asp).

The possible effects of defects or delays in pIII or pVIII signal peptide cleavage are interesting. Because both molecules possess hydrophobic domains at their N- and C-termini (signal peptide and membrane anchor, respectively), they could readily be converted to polytopic membrane proteins by recombinant inserts that affected signal peptidase activity. Considering that phage gene expression is crudely regulated, an anticipated result would be accumulation within the host membrane of pIII (which at normal levels mediates a number of membrane effects) or pVIII (~2700 copies are required per virion), with consequent pleiotrophic effects on host viability.

1.3.2. Other aspects of membrane insertion/translocation

The mature protein sequence can influence more than signal peptidase activity. For example, the normally Sec-independent membrane insertion of pVIII becomes Sec-dependant when the periplasmic loop (transiently formed between the hydrophobic core of the signal peptide and the C-terminal membrane anchor) is increased in size from 20 to 118 residues, and in a related way, Sec-dependance of a leader peptidase construct (studied as a model membrane protein rather than for its physiological role) increases concomitantly with gradual increases in size -- from 25 to 65 residues -- of its periplasmic loop (reviewed in von Heijne 1994). Several studies (reviewed or reported in von Heijne 1994; Cao *et al.* 1995) point to a role for both negatively and positively charged residues in markedly influencing the efficiency and success of membrane translocation. Thus, increasing the negative charge of the periplasmic loop in pVIII and leader peptidase constructs results in a corresponding increase in dependance on the

electrochemical membrane potential for efficient translocation or, in the case of some pVIII constructs, on the Sec apparatus (Cao *et al.* 1995). Finally, inserting as few as four or five tandem Asp and Glu residues at the beginning or middle of the periplasmic loop of pVIII can inhibit membrane insertion or processing by leader peptidase to as little as 18% of wild-type levels (Cao *et al.* 1995).

In context of these findings, it is interesting that Grihalde *et al.* (1995) constructed a 30-residue RPL in which variable residues followed the N-terminal mature region sequence DYKDDDDKAETA. This sequence includes an epitope tag (underlined residues) incorporated with the idea that loss of the tag would be diagnostic for N-terminal proteolysis of the pIII fusions. Surprisingly, the authors noted without comment that only 5-10% of virions displayed an intact tag and that infectivity (the fraction of physical particles that yield productive infections) of pooled library virions was only 0.4%. Conceivably, their findings may reflect a cellular response to problems in membrane translocation of the charge-rich sequence. Indeed, the neural networks developed by Nielsen *et al.* (1997) for predicting signal peptides conclude[4] that the pIII preprotein sequence encoded by the Grihalde *et al.* construct lacks a signal peptide.

1.3.3. Genomic libraries, membrane anchors and secondary signal sequences

A related problem is likely to arise in genomic PDLs, for a surprisingly large fraction (20%) of random chromosomally-derived sequences code for peptides that can serve as N-terminal signal peptides (Kaiser *et al.* 1987). Given also that (i) signal peptides and membrane anchors are difficult to distinguish by prediction algorithms (Nielsen *et al.* 1997), suggesting a relative ease of interconversion, and (ii) sequence requirements for transmembrane segments

---

[4]Prediction was made using the WWW server (http://www.cbs.dtu.dk/) of the Center for Biological Sequence Analysis using the Grihalde *et al.* preprotein sequence VKKLLFAIPLVVPFYSHS-DYKDDDDKAETA, where '-' indicates the cleavage site in wild-type pIII. The sequence yielded a cleavage site score well below (66% of) the minimum default cutoff value.

include only minimal length (>12 residues) and hydrophobicity (Pugsley 1993), then a potentially large fraction of genomic PDL clones might be expected to contain motifs that hinder translocation, serve as anchors, or yield unexpected cleavage of pIII or pVIII fusion products. Presumably, the fraction of clones affected will increase with gene fragment size. In terms of library "completeness" (the fraction of coding sequences successfully displayed on a virion surface), effects will likely be greatest in libraries of low redundancy (*i.e.*, there is little sequence overlap among clones) such as those constructed with restriction fragments, and least where redundancy is greatest, such as those derived from random DNase I digestion (Anderson 1981) of chromosomal DNA.

1.3.4. Codon usage and context

Codon usage and translational context in *E. coli* (reviewed briefly in Smith & Smith 1996; Berg & Silva 1997) may adversely affect synthesis of recombinant pVIII (especially; in view of the ~2700 copies required per virion) and pIII (less so; only 4-5 copies per virion), particularly for genomic PDLs derived from genomes with substantially different G+C content from that of *E. coli*, and with correspondingly different codon preferences. Codon usage within the *Bordetella pertussis* (%G+C = 66-70 *versus* ~48-52% for *E. coli*; Pittman 1984) *fhaB* gene, which was employed in phage display studies presented in this thesis, provides a relevant example. Analysis of 196 *fhaB* Arg codons shows that the codon most common in *E. coli* (CGU) tends to be avoided, and while this is largely compensated for by use of an alternate *E. coli*-favoured codon (CGC), *fhaB* employs other Arg codons rarely used by *E. coli*. These include 15 occurrences of AGG, the rarest of *all* sense codons (with a similarly rare cognate tRNA) employed by *E. coli* under varying growth conditions (Dong *et al.* 1996) and known to adversely affect recombinant gene expression in *E. coli* (reviewed in Kane 1995). Notably, *E. coli* avoids not only AGG but also the AG-G triplet (where '-' denotes a boundary between

codons), for reasons that appear related to translation (Smith & Smith 1996; Berg & Silva 1997). That the AG-G triplet accounts for fewer than 4% (*versus* an expected 25%) of a dataset of 742 *E. coli* tetranucleotides (Smith & Smith 1996) but 38% of a similar *fhaB* dataset (my data not shown) suggests difficulties in phage display of *fhaB*-derived peptides.

1.3.5. Molecular chaperones and export targeting

Nascent and completed proteins interact with a variety of host molecules with roles in folding and export. These include molecular chaperones such as GroEL/GroES as well as SecB and other components of the Sec apparatus (Pugsley 1993; Hartl 1996; Mayhew & Hartl 1996; Murphy & Beckwith 1996). While it is not yet clear how cytoplasmic proteins are distinguished from those destined for export, the mere attachment of a signal peptide to a normally cytoplasmic protein does not ensure its export (Summers & Knowles 1989; Andersson & von Heijne 1991). Rather, the protein must possess certain general motifs (patterns of hydrophobicity, positively-charged and flexible polypeptide segments) that allow maintenance, by molecular chaperones, of an unfolded state required for membrane translocation.

Although chaperone-assisted phage display of antibodies (see, for example, Lah *et al.* 1994; Soderlind *et al.* 1995) and of recombinant proteins generally (reviewed in Wall & Pluckthun 1995) have been explored (with mixed results), little is understood regarding how export of recombinant pIII and pVIII are likely to be influenced by molecular chaperones. Conceptually, fusion of small peptides to large, normally exported molecules such as pIII (424 residues) should have little effect on export competence. Conversely, fusion of a large, normally cytoplasmic protein to the relatively smaller pVIII (50 residues) would be expected to be pose greater difficulties.

1.3.6. Proteases

Protein degradation (reviewed in Miller 1996) is an essential and highly active component

of *E. coli* metabolism. Although much of this degradation is used to regulate levels of functional proteins, much also is concerned with eliminating damaged proteins. While the ways in which targets for degradation are identified are not well understood, certain kinds of alterations tend to increase the rate of degradation of otherwise stable proteins, including the production of longer or shorter versions of a normal protein, or amino acid substitutions that lead to exposure of structural elements not normally accessible in a native protein. Proteins and proteases have evolved together, it has been suggested, such that protein evolution (and thus recombinant technology) is constrained by proteolytic machinery effectively designed to recognize "non-native" elements. In context of these ideas, recombinant pIII and pVIII, appearing abnormal, may be candidates for degradation.

Although there is some evidence, illustrated by later examples, that protein degradation is an issue in phage display, the issue is seldom examined. Indeed, our understanding of the extent to which protein degradation occurs may be obscured by the common practice of inferring the sequence of a displayed peptide from that of its encoding insert into gIII or gVIII, without corresponding assessment (such as by SDS-PAGE and Western blotting) of pIII or pVIII fusion products.

1.4. Impact of Phage and Host Biology on Phage Display

Although in some applications of recombinant gene expression, host biology-related problems with the quantity and quality of exported proteins can be (or, out of practical necessity, must be) ignored, such matters assume greater importance with PDLs. This is because much of the utility of a PDL will derive from unbiased and reliable display of all possible encoded proteins. While issues such as library completeness, bias and clonal competition have been raised from time to time in applications of phage display, such as in explaining unexpected results, few studies have addressed the issues directly. As suggested by

material reviewed in the preceding sections, there is evidence that phage and host biology may interfere with successful phage display of certain peptides and proteins to a greater extent than commonly recognized. In reviewing applications of phage display, the following section provides a sense of (i) the degree to which such interference is in issue, and (ii) the extent to which it has been examined.

1.5. APPLICATIONS AND INNOVATIONS IN PHAGE DISPLAY

1.5.1. Random peptide libraries

(a) Library diversity and bias. RPLs displaying peptides of lengths 6 to 38 (36 residues randomized) have been described. A concern, particularly with longer peptides, is whether a library can contain all possible sequences. Although a hexapeptide library containing all of the 20^6 (6.4×10^7) possible clones can be constructed with only moderate difficulty (*e.g.*, Scott and Smith [1990] plated transformants from multiple electroporations on 64 large NUNC culture dishes to construct their 4×10^7-member library) and larger libraries are achievable, a practical upper limit ($\sim 10^9$ unique clones) dictates that libraries of peptides longer than seven residues risk being "incomplete".

"Completeness" is not essential, for libraries of limited diversity yield useful results. Thus, although the library of Devlin *et al.* (1990) displays only $1/10^9$ of the 20^{15} possible 15-residue peptides, Motti *et al.* (1994) successfully employed it to identify a MAb-reactive peptide that, although it shared only six (of 13) residues with the native antigen, could be used to induce antigen-reactive antibodies. Similarly, Schellekens *et al.* (1994) employed the Devlin library to identify a MAb-reactive peptide that shared only three residues with the native antigen (glycoprotein D of herpes simplex virus type 1, HSV-1) but, when used as an immunogen (as a BSA-hapten conjugate), nevertheless protected mice against subsequent HSV-

1 challenge. Still, lack of diversity may have limited these successes for in each study only a single sequence was isolated from among the 39 (Motti *et al.* 1994) or 5 (Schellekens *et al.* 1994) antibody-reactive clones analysed.

A related issue concerns over- and under-representation of specific amino acids. Much of this is unavoidable and reflects the bias inherent in the genetic code. Commonly, RPLs are constructed with only 32 codons (*e.g.*, of the form NNK, where K = G or T) with the effect of reducing inherent bias from a maximum amino acid molar ratio of 6:1 (*e.g.*, Arg *versus* Met) to 3:1. Of greater concern is bias due to biological selection against specific residues or sequences, such as position-specific bias against residues in the early mature region of pIII or pVIII fusion products. Realistic assessment of such bias requires that relatively large numbers of randomly selected library members be analysed. Except for a limited but otherwise careful study by DeGraaf *et al.* (1993) showing no apparent bias, such analyses have not been reported.

Crude measures of bias, possibly indicative of sequence- or position-specific effects, can be derived from analysis of the overall (*i.e.*, position-independent) distributions of amino acids in a small number of randomly selected clones. Cursory analyses suggest that bias is not extreme. Cwirla *et al.* (1990) analysed 52 randomly selected clones and found that the amino acid content of the displayed peptides was not entirely random (a χ^2 test on their data yields $p < 0.01$), Gly and Lys occurring ~1.8-fold more often than expected, and Pro about half as often as expected. Similar over- and under-representation of Gly and Pro has been found in other RPLs (Jellis *et al.* 1993; Burritt *et al.* 1996). χ^2 analyses of data published to show *lack* of bias surprisingly show that Gly occurred 2.1-fold more often than expected in one RPL (Dybwad *et al.* 1995a) and that less than one-third the expected Cys occurred in a second RPL (Dybwad *et al.* 1995b). Blond-Elguindi *et al.* (1993) also found bias favouring Gly and strongly disfavouring Cys in their RPL, as well as position-specific effects. Also, and in a manner

consistent with their expected influence on membrane insertion and signal peptide cleavage, Asp and Glu were slightly more abundant towards the N-terminus, while the reverse was true for Lys. Interestingly, Kay *et al.* (1993) found apparent selection against odd numbers of cysteines in their 38-residue RPL and reasoned that an unpaired cysteine could form a bond with one of the eight cysteines of native pIII, with a subsequent effect on phage infectivity. Zhong *et al.* (1994) have also found a bias against unpaired cysteines.

(b) Representative results - biopanning with antibodies. An example of a straightforward and successful application of RPLs is provided in a study by Conley *et al.* (1994), who biopanned a 15-residue RPL of limited diversity (9×10^7 unique clones) with an anti-HIV-1 (human immunodeficiency virus-1) mAb. Conventional biopanning methods (Smith & Scott 1993) followed by immunoscreening (colony lifts of phage-secreting host colonies) of putative positive clones yielded 20 mAb-reactive clones that could be classified into four consensus groups displaying XXDKW (8 clones), $XLD^R/_KW$ (5), $EXD^R/_KW$ (4) and $ELD^R/_KW$ (3) and subsequently used to identify ELDKW as the native epitope.

Not all biopanning experiments are so successful. For example, after three rounds of biopanning the hexapeptide RPL of Scott and Smith (1990) with α-keratin mAbs, Böttger and Lane (1994) found that only one of seven mAbs had substantially enriched for target clones. In later studies, Böttger *et al.* (1995b) had greater success with a 20-residue RPL than with a 12-residue RPL; nevertheless, the small number of 20-residue sequences recovered and their limited sequence similarity with the native protein were not enough to identify an epitope. Rather than a consensus sequence, Sioud *et al.* (1994) found a compositional "theme" after biopanning a RPL with TNF-α (tumor necrosis factor) human autoantibodies. Thus, of 63 affinity selected clones chosen for analysis, 46 displayed Ser- and Pro-rich sequences A<u>SS</u>LLA<u>SSP</u> (20 clones), N<u>SSP</u>YLNTK (16) or <u>PQSP</u>G<u>SS</u>F<u>P</u> (10) and the remaining 17

clones also contained Ser or Pro.

Some biopannings of RPLs with antibodies (or with other ligates) fail to yield meaningful results (Lane & Stephen 1993; Smith & Scott 1993) or yield misleading results. Smith and Scott (1993) have noted that two consensus sequences arise repeatedly with a variety of antibodies and other ligates. These, PWXWL (X is usually A or E) and GDWVFI, presumably bind some other component of the system (possibly streptavidin, Roberts *et al.* 1993). Indeed, phage-displayed peptides that bind to these components have themselves been identified by biopanning RPLs. These include streptavidin (Devlin *et al.* 1990; Kay *et al.* 1993), biotin (Saggio & Laufer 1993), mouse IgG Fc region (Kay *et al.* 1993) and plastic (certain Tyr- and Trp-rich sequences; Adey *et al.* 1995). Comprehensive studies (Böttger *et al.* 1995a; Stephen *et al.* 1995) of 6-, 12- and 20-residue RPLs biopanned with a total of 24 mAbs suggest other reasons for a lack of success in some biopannings, including that the epitopes recognized by some mAbs, such as the "linear but discontinuous" $KKX_{1-4}STSX_{0-4}HXK$ variants of a p53 epitope, are substantially longer than those provided in libraries commonly employed.

(c) Biopanning with non-antibody molecules. Similar results derive from panning RPLs with molecules other than antibodies. That is, as illustrated below, the sequences of the affinity selected peptides match those of the native ligate in varying degrees or not at all.

Thus, "good" matches were found by Cheadle *et al.* (1994) and Sparks *et al.* (1994), who independently biopanned RPLs with fusion proteins containing the Src homology 3 (Src SH3) domain and identified a common motif (RPLPXXP) corresponding to a previously identified Src SH3 domain binding motif. Similar success was enjoyed by Healy *et al.* (1995), who biopanned 6- and 15-residue RPLS with the integrin $\alpha_v\beta_3$ and identified a diverse collection of sequences containing the RGD triplet. In contrast, Smith *et al.* (1993) biopanned a hexapeptide RPL with S-protein (a fragment of bovine pancreatic ribonuclease) and identified

a sequence motif, $^{F}/_{Y}NF^{E}/_{V}{}^{I}/_{V}{}^{L}/_{V}$, that bore little resemblance to the natural ligand. An even less-defined but important motif was found by Blond-Elguindi *et al.* (1993) who, by biopanning 8- and 10-residue RPLs with the molecular chaperone BiP, identified a heptameric motif (best described as $\Theta^{W}/_{X}\Theta X\Theta X\Theta$, where Θ = Trp, Leu, Phe or other large hydrophobic amino acid) that fits well with the role of BiP in recognizing partially folded polypeptides.

Studies by Kay *et al.* (1993) illustrate several aspects of biopanning RPLs. Panning their 38-residue RPL with streptavidin yielded two classes of sequences, the first providing a consensus sequence ($HP^{Q}/_{M}X$, X = a non-polar residue) similar to that previously identified by phage display (Devlin *et al.* 1990), the second providing no clear motif. Intriguingly, clones of the non-motif class bound to streptavidin with ~5-fold greater affinity than selected $HP^{Q}/_{M}X$ clones. The importance of context and a possible advantage of disulfide-mediated constrained presentation was suggested by the finding that the affinity selected sequence *CHPQAC* (* denotes potential disulfide bond partners) bound 100-fold less efficiently after treatment with a reducing agent.

(d) Substrate phage. "Substrate phage" are a class of RPLs pioneered by Matthews and Wells (1993) to identify peptide substrates of proteases. In a study that illustrates their use, Matthews *et al.* (1994) constructed a substrate phage library by inserting a randomized pentapeptide "substrate" flanked by di-alanine residues between (i) a binding domain (a variant of human growth hormone, hGH) and (ii) pIII. After being immobilized on hGH receptor molecules, the library was treated with furin to release clones bearing a furin substrate. Six rounds of enrichment led to a consensus motif which was used as a "seed" for a second series of randomized libraries. Furin selection from these libraries yielded a more defined motif, $^{L}/_{P}RRF^{K}/_{R}RP$.

1.5.2. Phage-displayed proteins and alternative methods of display

Much of the success in phage display of larger peptides and proteins has been attributed to the development of alternative vectors. These can be classified (Smith 1993) according to (i) the choice of pIII or pVIII as carrier molecule, (ii) the number of fusion products displayed per virion, and (iii) the vector strategy employed. In the Type 3 (or multivalent pIII) systems commonly employed in RPLs, vectors are constructed by inserting a cloning site into gIII with the result that all 4-5 copies of pIII display a foreign peptide (Figure 1-1C and D). In Type 3+3 ("monovalent" or phagemid) display, recombinant gIII is carried on a phagemid (a plasmid containing the Ff origins of replication and packaging signal). Superinfection of phagemid-harboring cells by a helper phage (an Ff phage with a defective packaging signal) results in synthesis of phagemid s.s. DNA and its preferential packaging into infectious virion-like particles comprised of both wild-type and recombinant pIII. Recombinant gIII expression is commonly controlled so that, on average, no more than one recombinant pIII molecule is incorporated per phagemid particle. In Type 33 display, the genes encoding wild-type and recombinant pIII are included in a single phage genome; not uncommonly, recombinant gIII is inserted into the multiple cloning site of an M13 cloning vector. Type 8, 8+8 and 88 vectors are the pVIII multivalent and oligovalent counterparts to the multivalent and monovalent pIII vectors.

The perceived relative advantages of these systems relate to two main issues. The first concerns intrinsic affinity *versus* functional avidity: affinity selection from monovalent PDLs, it is believed, leads to identification of target clones with higher affinity for their cognate molecule than selection from multivalent libraries where multiple low-affinity interactions create high functional avidity. The second issue reflects phage and host biology. Here the belief is that reducing the copy number of a pIII or pVIII fusion product can improve the odds

that an intrinsically toxic or difficult-to-export product will be successfully displayed on the virion surface. Accordingly, monovalent vectors are often employed for display of larger polypeptides or proteins. Type 8 display is uncommon because of early difficulties in displaying peptides longer than five or six residues. In contrast, proteins as large as antibody Fab fragments (Kang *et al.* 1991) have been displayed with type 8+8 and 88 display (Smith 1993).

(a) Type 3 display. Both low infectivity and proteolysis have been described in several multivalent pIII systems. For example, in otherwise successful display of catalytically active *E. coli* alkaline phosphatase (AP), infectivity of the AP-pIII recombinants was only 0.3-0.4% (one-fifteenth of that of the parent vector) and proteolysis of AP-pIII fusion proteins was considerable (McCafferty *et al.* 1991). In early preparations, only 5-10% of the fusions were intact; in later preparations, 30-60%. Similar degradation was seen in virions displaying the 32 kDa B chain of ricin (a lectin from castor beans): only one copy of intact B chain-pIII fusion protein was present for every three copies of proteolyzed product (Swimmer *et al.* 1992). Proteolysis was also suspected in phage displaying the 58-residue bovine pancreatic trypsin inhibitor (Roberts *et al.* 1992b). In contrast, no significant degradation products were seen in Type 3 display of correctly folded and fully active β-lactamase (Soumillion *et al.* 1994b). Human interleukin 3 (hIL3) has also been displayed in a Type 3 vector (Gram *et al.* 1993), but it is unclear what fraction of virions displayed intact hIL3.

(b) Type 3+3 display. Among early phage display papers was that describing Type 3+3 display of the 191-residue disulfide-containing human growth hormone (hGH), accomplished by fusion of hGH to a truncated pIII, the C-terminal domain ($pIII_C$) required for virion morphogenesis (Bass *et al.* 1990). Importantly, phagemid-displayed hGH-$pIII_C$ could bind to hGH receptor and was recognized by mAbs whose epitopes are sensitive to hGH conformation.

Although phagemid titers were sensitive to the ratio of recombinant *versus* wild-type pIII molecules incorporated into each particle -- indicating an undefined but possibly translocation-related effect of hGH-$pIII_C$ on phage or host biology -- this ratio could be controlled so that, for example, only one in ten particles contained a single hGH-$pIII_C$ molecule. Immunoblots showed little evidence of hGH-$pIII_C$ degradation. Other molecules displayed as Type 3+3 $pIII_C$ fusions include (i) the α subunit of the high-affinity receptor for IgE (FcεRI) in a form recognized by IgE (Scarselli *et al.* 1993) and (ii) the 23 kDa human ciliary neurotrophic factor (CTNF) in a form that bound to its natural receptor (Saggio *et al.* 1995). Degradation of the $pIII_C$ fusions was not examined in these latter papers.

Other methods of Type 3+3 display involve fusions to the complete pIII molecule. One such study (Eerola *et al.* 1994) involved display of the 33 kDa prostate specific antigen (PSA; a serine protease) in a form capable of being recognized by antibodies recognizing non-linear epitopes. Nevertheless, Western blotting showed degradation of many of the PSA-pIII fusion products. Staphylococcal protein A (SPA) has been displayed in this manner (Kushwaha *et al.* 1994) but it is unclear what fraction of phagemid particles displayed intact SPA. Both Type 3+3 and Type 3 display have been employed to display the first 176 residues of human CD4 ($CD4_{1-176}$) in a form capable of capture by gp120 of the human immunodeficiency virus (Abrol *et al.* 1994). Infectivity of the $CD4_{1-176}$ Type 3 recombinants was only ~5% of that of the parent vector.

(c) Types 33 and 88 display. In a study (Corey *et al.* 1993) exploring both Type 33 and Type 88 display, sequences encoding a signal peptide and trypsin were fused to those encoding $pIII_C$ or mature pVIII and cloned into M13mp18, resulting in production of virions with hybrid coats comprised of trypsin-$pIII_C$ or trypsin-pVIII as well as wild-type pIII and pVIII. Importantly, the displayed trypsin fusion proteins possessed near wild-type enzyme activity.

However, <20% of virions displayed fusion proteins, and *mean* infectivity (plaque-forming units per virion, independent of whether a virion displayed a recombinant protein) was less than one-quarter of that of the non-recombinant vector. These are suggestive of profound effects of the fusions on host metabolism.

The earlier and careful construction by Markland *et al.* (1991) of a Type 88 vector and its employment to display bovine pancreatic trypsin inhibitor (BPTI) in a functional form illustrates several aspects of vector design and host cell biology. In a manner similar to that described for trypsin display, Markland *et al.* inserted sequences encoding BPTI-pVIII fusions into M13mp18. To avoid homologous recombination involving identical gVIII coding sequences, and in contrast to exclusive use of wild-type gVIII sequences employed by Corey *et al.* for trypsin display, Markland *et al.* used alternative codons to construct a synthetic gene VIII for the BPTI-pVIII fusions. Their first constructs, encoding a fusion of the pVIII signal peptide, BPTI and mature pVIII, failed to produce virions displaying BPTI-pVIII fusion products. Rather, Western blots of host cell lysates showed only a single protein species with a size characteristic of unprocessed fusion protein. Replacement of the natural pVIII signal peptide with an alternative signal sequence resulted in partial processing of the fusion protein, but reasonable levels of (and greatly enhanced) processing and incorporation of BPTI-pVIII into virions could only be achieved by use of an *E. coli prlA* (*secY*) host. The 1:100 to 1:50 ratio thus achieved of BPTI-pVIII to wild-type pVIII incorporated into virions was increased to 1:30 by changing the start codon of the wild-type gene VIII from ATG (Met) to CTG (Leu). Notably, this ATG→CTG mutation also resulted in a 10-fold reduction in virion yields.

(d) Type 8+8 display. An innovative example of Type 8+8 display of an antibody Fab fragment is provided in a study by Kang *et al.* (1991). Having employed a phagemid to coexpress (i) heavy chains (as V_H-C_{H1}) as fusions between a PelB signal peptide and pVIII,

and (ii) light chains (V_L-$C_{L)}$ as fusions to a second PelB signal peptide, they expected the V_H-C_{H1}-pVIII and V_L-C_L molecules to associate in the periplasm and form a properly folded Fab fragment. Careful studies, including electron microscopy of phagemids decorated with antigen-Fab complexes, validated their expectations. Apparent adaptations of this approach, also relying on periplasmic association of components that become incorporated into virions, are reviewed in later sections.

1.5.3. "Directed evolution of a protein"

A number of phage applications have derived from a novel approach, described as "directed evolution of a protein", in which Roberts *et al.* (1992a) constructed a PDL of 1000 protease inhibitor variants derived from the sequence of wild-type bovine pancreatic trypsin inhibitor (BPTI) by limited randomization of five residue positions believed important for interaction between BPTI and human neutrophil elastase (HNE). Subsequent affinity selection with HNE-agarose yielded a variant with $>10^6$-fold greater affinity for HNE than the parental sequence.

Similar studies, often with multiple libraries of greater diversity, have yielded other peptides or proteins with altered target specificity including (i) other protease inhibitors with increased affinity for their target (Dennis & Lazarus 1994a, 1994b; Wang *et al.* 1995a), (ii) an enzyme with altered specificity for active-site ligands (Widersten & Mannervik 1995) and (iii) zinc fingers with modified specificity in one or more fingers (reviewed in Choo & Klug 1995). The popularity of PDLs as tools for *in vitro* selection of peptides and proteins with altered properties is reflected in the subheading "phage fever" employed by Clackson and Wells (1994) in their review of the field.

1.5.4. Gene fragment libraries

Constructed by Smith (1985), the first PDL was a gene fragment library displaying

fragments of *Eco*RI endonuclease and methylase enzymes in the *middle* of pIII of wild-type phage f1. Notwithstanding the publication by Smith (1992), in his widely-distributed handbook, of protocols oriented to creating such libraries, they have found surprisingly little employment. Indeed, they appear to be limited to the following few examples.

(a) Multivalent pIII-display gene fragment libraries screened with antibodies. In an early demonstration of the utility of subgenomic libraries, Bleul *et al.* (1991) cloned 200 bp DNase I-generated fragments of subgenomic DNA of the human papilloma virus type 18 (HPV-18) into the Type 3 frameshift vector fUSE1 (described earlier; see also Parmley & Smith [1988]), creating a library of ~10^5 transformants of which <2% (so it appears from their data) possessed inserts capable of directing display of a HPV-18-related peptide. Screening by plaque lifts (rather than affinity selection) with sera against HPV-18 E6 and E7 fusion proteins identified two unique 30- to 50-residue antibody-reactive peptides for each serum.

Similarly, Wang *et al.* (1995b) cloned 100-200 bp fragments of the gene encoding the bluetongue virus capsid protein VP5 into a conventional (non-frameshift) Type 3 vector, fUSE2 (described earlier; see also Parmley & Smith [1988]), creating a library of 5,500 transformants, of which only 200 were believed to display a VP5-related peptide. Biopanning with a VP5-specific mAb led to the identification of two overlapping (44- and 50-residue) MAb-reactive peptides.

Petersen *et al.* (1995) used a related approach to construct Type 3 libraries (in fUSE5, Scott & Smith 1990) of 50-200 bp gene fragments encoding peptides for (i) *Drosophila* RNA polymerase II, (ii) human 53 protein and (iii) human cytokeratin 19 protein. One round of biopanning followed by colony blot screening led to identification of epitopes recognized by mAbs raised against these proteins. Three cytokeratin 19 libraries were constructed with different peptide linkers flanking the cloned insert. Interestingly, only one of these libraries

yielded MAb-reactive clones; this library used Cys-containing linkers to present peptides in a disulfide-constrained loop.

(b) Identification of ligand-binding domains with a phagemid system. Type 3+3 display has also been employed in gene fragment libraries. In a model system exploring the feasibility of constructing prokaryotic genomic libraries, Jacobsson and Frykberg (1995) cloned 100-700 bp sonication products of *Staphyloccocus aureus* genomic DNA into a phagemid vector. Biopanning the resulting 10^7-member library with human IgG and fibronectin and subsequent screening by (i) colony blotting with these same molecules, and (ii) hybridization with oligonucleotide probes corresponding to fibronectin binding domains, led to the identification of sequences corresponding to known binding regions of staphylococcal protein A and fibronectin binding proteins. Oddly, the cloned inserts of all five fibronectin-binding clones contained +1 or -1 frameshifts, and 47 of 50 clones derived from other libraries possessed similar frameshifts.

(c) cDNA libraries by means of Jun and Fos leucine zippers. An innovation system developed by Crameri and Suter (1993) for phage display of cDNA fragments exploits the interaction of the Jun and Fos leucine zippers, in a manner conceptually similar to the approach adopted by Kang *et al.* (1991) for display of Fab fragments by means of intraperiplasmic association of antibody heavy and light chains. In the Crameri and Suter phagemid system, Jun and Fos are expressed from *pelB-jun-gIII*$_C$ and *pelB-fos*-cDNA fusions, respectively, with the result that intraperiplasmic, high-affinity interactions between Fos and Jun allow display of cDNA-derived peptides on the phagemid surface. To prevent exchange of *fos*-cDNA fusion products among phagemid particles, cysteines were added to the Jun and Fos N- and C-termini to allow intermolecular disulfide bond formation. Notably, Jun-Fos display involves C-terminal *fusions* (to Fos) rather than *insertions* (near the N-terminus of pIII), thus allowing expression

of cDNA or other gene fragments containing translation stop sites or other non-coding sequences. In a followup study, Crameri *et al.* (1994) confirmed the utility of Jun-Fos display by screening an *Aspergillus fumigatus* cDNA library with human serum IgE, thereby identifying sequences of putative allergens.

1.5.5. Innovations in peptide presentation

Several approaches have been adopted to the construction of PDLs in which peptides are displayed in a constrained, often defined, conformation. This has proved advantageous in some cases but not in others.

(a) Disulfide-constrained peptides. Because, prior to their incorporation into virions, pIII and pVIII are inserted into the inner membrane with their N-termini within the periplasm, paired cysteine residues flanking randomized sequences (*e.g.*, *CX_6C*, where * denotes a potential disulfide bond and X is any amino acid) can be used to present peptides in a disulfide-bonded (Luzzago *et al.* 1993) loop. Cys-constrained RPLs have been employed with remarkable success. In a preface to two papers (co-published in Science: Livnah *et al.* 1996; Wrighton *et al.* 1996) describing a peptide derived from such libraries, Wells (1996) wrote that the articles "are enough to reinstate one's belief in Santa Claus", for the authors had found a 20-residue disulfide-constrained peptide capable of forming dimers and subsequently dimerizing the erythropoietin receptor, leading to its activation. The isolation of brain- and kidney-targeting phage after intravenous injection of Cys-constrained RPLs into mice (Pasqualini & Ruoslahti 1996) illustrates another novel application of disulfide-constrained libraries.

Other examples illustrate the relative merits of disulfide-constrained and unconstrained libraries. McLafferty *et al.* (1993) biopanned a *$X*CX_4C*X$* RPL with an anti-β-endorphin mAb and found that the disulfide bond was required for high-affinity binding of the MAb-reactive peptides (*$CYG^G/_AFC*$). In contrast, Felici *et al.* (1993) biopanned both constrained

(*CX_9C*) and unconstrained (X_9) RPLs and, having identified three MAb-reactive peptides from each library, found that the unconstrained peptides were better recognized in immunoblots than those from the constrained library. In reviewing constrained display, Ladner (1995) has suggested that increasing the constraint imposed on a peptide segment (*e.g.*, RGD in xxxRGDxxx, *CxxRGDxxC* and *Cx*CRGDC*xC*) decreases the likelihood that it will bind to any particular target, but that if a target is found the peptide will bind more tightly and specifically, and more can be learned about the nature of the binding.

(b) Presentation on carrier molecules. Tendamistat, a 74 amino acid molecule comprised of two three-strand β-sheets, has been employed as a scaffold in which two adjacent strand-connecting loops (13 and 6 residues) can be randomized (McConnell & Hoess 1995) in Type 33 display. A perceived advantage of constrained display involving two (*versus* one) randomized loops is that it may better approximate protein surface structures involved in inter-molecular recognition, such as between discontinuous epitopes and antibody. The 61-residue "minibody" (Pessi *et al.* 1993) is similar. Engineered from an antibody V_H domain and comprised of two three-strand β-sheets, the minibody retains the regions corresponding to two adjacent hypervariable loops (H1 and H2) of the parental antibody. Its utility was demonstrated (Martin *et al.* 1994) by the identification, from a Type 3 library displaying minibodies in which one or both of these loops had been randomized, of an H2 variant that bound human interleukin-6 and inhibited its biological activity. A more potent variant was identified in a study employing Type 3+3 display of minibodies in which the H1 loop of the previously identified H2 variant was randomized (Martin *et al.* 1996).

Phage display of randomized α-helices has been explored by Bianchi *et al.* (1995), who randomized five helix positions of a Cys_2His_2 zinc finger that are most exposed in the helix and which cluster on one side. Since zinc coordination and folding are coupled in zinc fingers,

zinc-dependent binding (between a randomized finger and its cognate molecule) was used as a built-in control against structurally undefined sequences. Phage display involving randomizing solvent-accessible residues of a surface comprised of two of the three α-helices of a monovalent Fc-binding domain of staphylococcal protein A has also been explored (Nord *et al.* 1995).

1.5.6. Innovations in identification of target clones

Two innovative methods of identifying target clones exploit the concept employed by Kang *et al.* (1991) for Fab display, and by Crameri and Suter (1993) for Jun-Fos-mediated display of cDNA fragments, *viz.*, the periplasmic association of separately exported molecules to form a complex that becomes incorporated into a virion. The innovation lies in the physical separation of the two functional domains of pIII, the N-terminal ($pIII_N$) domain which plays a role in F-pilus adsorption, and the C-terminal($pIII_C$) domain which functions in virion morphogenesis.

In the "direct interaction rescue" system of Gramatikoff *et al.* (1994), an invariant "bait" molecule such as c-Jun (Gramatikoff *et al.* 1995) is cloned as an N-terminal fusion to $pIII_C$ while variant fragments such as those from a human cDNA library are cloned as C-terminal fusions to $pIII_N$. In cells harboring a cDNA fragment encoding a peptide that interacts with the Jun bait, periplasmic association of Jun-$pIII_C$ and $pIII_N$-cDNA fusion products yields infectious virions constructed with the heterodimeric complexes; in other cells, non-infectious virions containing only Jun-$pIII_C$ are produced. The similar system of Krebber *et al.* (1995) employs single-chain antibodies (scFv) fused to $pIII_C$, and infectivity is restored by cognate pairs of $pIII_N$-antigen and scFv-$pIII_C$. Notably, antigen or antibody sequences can be randomized or otherwise varied. The system is similar to that developed by Duenas and Borrebaeck (1994) to mimic clonal selection by the immune system.

1.6. THE PRESENT STUDY

1.6.1. Project goals: "Issues and Applications"

This thesis project was initiated with the goal of exploring the use of phage display in the study of pathogenesis, with a view in mind of demonstrating the feasibility of constructing genomic PDLs that could be used to identify pathogen-encoded proteins with roles in pathogenesis. The systematic working through of two model systems was intended not only to illustrate the application potential of phage display, but also to identify limitations and to explore ways around these limitations.

The simpler of these models involved constructing a 32-member library displaying variants of the immunodominant tetrapeptide repeats of the *Plasmodium falciparum* (malaria parasite) circumsporozoite protein (CSP) and using this library to, at a minimum, characterize the epitope specificities of a panel of α-CSP mAbs. Given that, as reviewed in Chapter 4, something was already known about the structure of CSP and the epitope specificities of the α-CSP mAbs, it seemed possible to both (i) confidently explore basic methodology with a model expected to yield a narrow range of results, and (ii) provide new insights into an important malaria parasite protein.

The second, more complex model involved constructing a series of PDLs displaying random peptide fragments (size-fractioned by library) of filamentous hemagglutinin (FHA), a multifaceted adhesin of the respiratory pathogen *Bordetella pertussis* and an important component of acellular pertussis vaccines (reviewed in Chapter 6). These libraries were to be used for two purposes. First, biopanning with rabbit α-FHA polyclonal antibodies and subsequent immunocharacterization of affinity selected clones was expected to provide insight into the antigenic makeup of FHA. Second, characterization of randomly selected library

members was expected to reveal something about the nature of peptides that could and could not be successfully displayed on phage. The large size of FHA (220 kDa in its mature, processed form) and its encoding gene (*fhaB,* 10 kbp) ensured that *fhaB*-derived PDLs would serve as a non-trivial models of genomic libraries.

1.6.2. Project accomplishments in context of the concurrent work of others

Concurrent with my work towards illustrating the potential of phage display and examining its limitations, it has become clear -- as the technology pioneered by Smith has been extended by others -- that phage display can be used to answer a variety of biological questions. Also evident, however, is that phage and host biology impose greater limitations than were suggested by the successes of the early RPL studies and those that shortly followed. Notably, few studies have specifically addressed phage- or host-imposed limitations. Equally noteworthy, and although phage display was pioneered with a gene fragment library, there have been few reports of gene fragment or genomic PDLs. Possibly, this reflects research objectives or a perceived lack of utility. Alternatively, it may reflect difficulties inherent in employing these libraries. Indeed, the heterologous nature (peptide fragment size and sequence diversity) of peptides in genomic PDLs sets these libraries apart from nearly all other PDLs, which characteristically display (i) peptides of great diversity but of small fixed size, as in RPLs, or (ii) large peptides or proteins in which only a small number of residues are varied within an otherwise uniform scaffold, as in affinity maturation studies or antibody display. Given this distinction, construction and employment of a genomic PDL might be seen as a kind of "extreme test" of phage display.

In this context, my study of phage display of FHA-derived peptides has been an "extreme test", particularly so because it involved the exclusive use of multivalent (Type 3) vectors and, as reviewed earlier, monovalent vectors are believed better suited to the display

of larger peptides and proteins. The vectors employed (the fUSE*n* derivatives of fd-tet) were chosen for several reasons (reviewed in Chapter 7), including that their design incorporated features deemed important or essential to the success of several components of my project. Although unexpected problems with these vectors made it difficult to meet some project objectives, in particular the characterization of randomly sampled FHA library clones, the problems themselves provided surprising insight into certain aspects of the impact of host cell biology on phage display.

1.6.3. Thesis layout

The layout of this thesis reflects the dual themes of my project, "issues" and "applications". Applications are described first. Chapter 3 does not relate directly to phage display, but rather provides a statistical analysis (based on the Brookhaven Protein Data Bank) of structures formed by Asx-Pro dipeptides, and develops a model for folding of the Asn-Pro-Asn-Ala sequences that comprise the *P. falciparum* CSP tetrapeptide repeats. This model is further developed by the experimental data of Chapter 4, derived from characterization of 32 phage-displayed variants of these repeats with a panel of α-CSP mAbs. Experiments summarized in Chapter 5, derived from biopanning RPLs with two α-CSP mAbs, serve to illustrate an inherent limitation of RPLs. Chapter 6 provides an antigenic analysis of FHA derived from biopanning a series of FHA-derived PDLs with rabbit α-FHA polyclonal antibodies. Chapter 7, in describing various difficulties encountered during the work that led up to the successful applications of Chapters 4 and 6, deals with issues of vector stability and the construction of libraries. Finally, Chapter 8 reviews my overall findings in context of their significance to phage display technology.

Chapter 2

Materials and Methods

2.1. Bacterial Strains, Plasmids and Bacteriophage

Escherichia coli strains are described in Table 2-I; growth characteristics of selected strains are shown in Figure 2-1. Plasmids are outlined in Table 2-II, phage display vectors in Table 2-III. Wild-type bacteriophage f1 (Zinder *et al.* 1963) and a random hexapeptide phage-display library, derived from the library described in Scott and Smith (1990) by amplification in *E. coli* K91, were gifts from G. P. Smith, University of Missouri. A second phage-display library of random 15-residue peptides was a gift from J. Scott, Simon Fraser University.

2.1.1. Culture of bacteria

E. coli strains were routinely cultured in LB broth or other media such as 2xYT (Sambrook *et al.* 1989) at 37°C with shaking at 125-200 rpm, or on plates (commonly LB, or other rich medium) containing 1.5% agar. For broth cultures, commonly, ≥1 mL of a 4 h to overnight "starter" culture was used as inoculum for 30-50 mL (125 mL Erlenmeyer flask) to 600 mL (2 L flask) 10-24 h cultures; these are referred to as "serial overnight culture (*n* mL)" where *n* = final culture volume. Strains harboring fUSE*n* or fDRW*n* vectors or their recombinants were cultured in media containing 20 μg tetracycline mL^{-1} (LB-Tet). As appropriate, *E. coli* K91-Kan was cultured in media containing 100 μg kanamycin mL^{-1}; strain MC1061 in media containing 25-100 μg streptomycin mL^{-1}; strains harboring pACYC184 in media containing 25 μg tetracycline mL^{-1} and 35 μg chloramphenicol mL^{-1}; strains harboring pAS100, pJB61, pNK1759 or pNK2859 in media containing 50 or 100 μg ampicillin mL^{-1}.

Table 2-I. *E. coli* strains used in these studies.

Strain	F-piliation	Amber suppression	Principal use	Source	Reference
K37[a]	HfrC	*sup*D	titration of amber[c] vector virions	ATCC[d]	ATCC 33626
K37/ pACYC184	HfrC	*sup*D	titration of amber vector virions, particularly of virions in undiluted culture supernatant containing tetracycline	this study	
K802	F-minus	*sup*E	$CaCl_2$-transformation host, vector and virion propagation	G. P. Smith, U. of Missouri	Smith & Scott 1993
K802 *recA*	F-minus	*sup*E	single experiment (Figure 7-16)	ATCC	ATCC 47026
K91[a]	HfrC	none	virion propagation and titration	G. P. Smith	Smith & Scott 1993
K91/ pNK1759	HfrC	none	single experiment (Figure 7-16)	this study	
K91/ pNK2859	HfrC	*sup*F	titration of amber vector virions	this study	
K91/pJB61	HfrC	none	single experiment (Figure 7-16)	this study	
K91-Kan[b]	HfrC	none	virion propagation and titration	G. P. Smith	Smith & Scott 1993
K91-Kan/ pACYC184	HfrC	none	virion titration, particularly for undiluted culture supernatant containing tetracycline	this study	
LE392	F-minus	*sup*E, *sup*F	$CaCl_2$-transformation host, vector propagation	ATCC	ATCC 33572
MC1061	F-minus	none	electroporation host, virion propagation	G. P. Smith	Smith & Scott 1993

[a]*E. coli* K37 and K91 are derivatives of K38 (Smith 1988).
[b]derived from *E. coli* K91 by G. P. Smith (Smith & Scott 1993); carries kanamycin resistance within a mini-transposon inserted into *lac*Z.
[c]Amber vectors are described in §7.3.2.b. See also Table 2-III.
[d]ATCC, American Type Culture Collection.

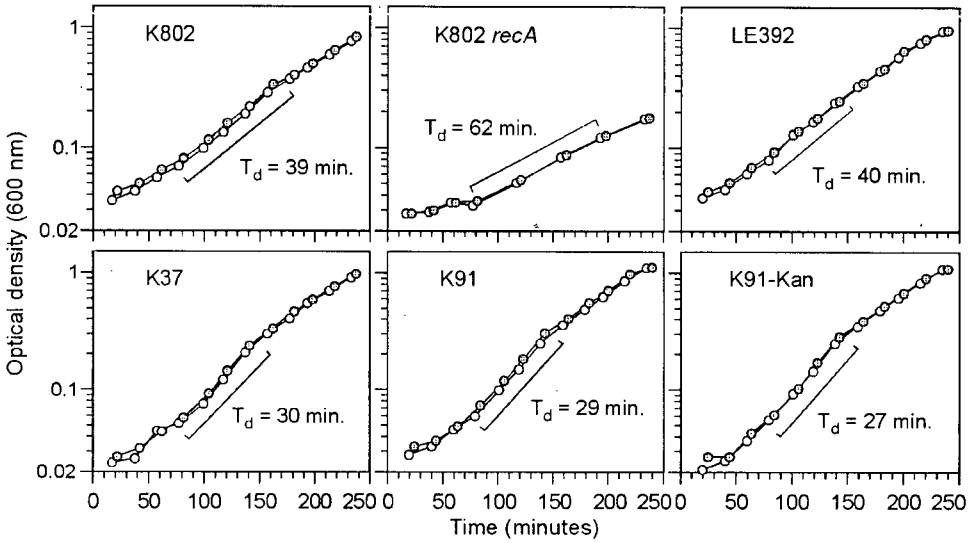


Fig. 2-1. Growth curves and estimated doubling times (T_d) of *E. coli* strains. Duplicate 0.5 mL samples from an overnight culture (2 mL LB; 37°C with shaking) of each strain were diluted 1:100 into fresh medium (50 mL LB pre-warmed to 37°C); these cultures were incubated with shaking at 37°C, and optical densities at 600 nm (OD_{600}) were measured at ~20 min. intervals using 1 mL samples (discarded) in 1 cm disposable cuvettes. Doubling time (T_d) for each duplicate was calculated on the basis of the minimum slope -- of every set of 3, 4 and 5 consecutive data points -- of the least squares linear regression of $\log(OD_{600})$ *versus* time. Reported T_d are arithmetic means for the duplicate cultures.

Table 2-II. Plasmids.

Plasmid	Description	Source	Reference
pACYC184	carries chloramphenicol resistance gene from transposon Tn9 and tetracycline-resistance gene from pSC101	New England Biolabs	Chang & Cohen 1978
pAS100	pTZ18R carrying (i) the 10 kilobase pair *B. pertussis fhaB* EcoRI fragment sub-cloned (A. Siebers) from clone C1-5 of a *Sau*3AI library of *B. pertussis* 18323 chromosomal DNA (M.J. Brennan, FDA, Bethesda, MD); this fragment is described in Delisse-Gathoye *et al.* (1990); (ii) ampicillin resistance	A. Siebers	A.S. and B. Finlay, unpublished
pJB61	pBR322 carrying (i) Ff gene III, (ii) ampicillin resistance	ATCC	ATCC 39162
pNK1759	pBR322 carrying (i) kanamycin resistance within a mini-Tn10 transposon, (ii) transposase, (iii) ampicillin resistance; designed for transposon mutagenesis	ATCC	ATCC 77352
pNK2859	pBR322 carrying (i) *sup*F within a mini-Tn10 transposon, (ii) transposase; (iii) ampicillin resistance; designed for transposon mutagenesis	ATCC	ATCC 77338

Table 2-III. Phage display vectors.

Vector	Description[a]	Source	Derived from	Reference
fDRW20	conventional[b] vector	this study	fUSE1	Figure 7-12, §2.8.1, §2.8.2
fDRW21				
fDRW22				
fDRW5	amber[c] vector		fUSE5	Figure 7-9, §2.8.3
fDRW613	amber vector		fDRW5	Figures 2-2 and 7-14, §2.8.4, §2.8.5
fDRW613C	two-amber[d] vector			
fDRW70	amber vector			Figure 7-23, §2.8.6
fDRW8*nn* series	conventional vectors			Figure 7-29, §2.8.7
fUSE1	frameshift[e] vector	G. P. Smith, U. of Missouri	fd-tet	Figure 7-1; Smith 1985; Parmley & Smith 1988; Scott & Smith 1990; Smith 1992; Smith & Scott 1993
fUSE2	conventional vector			
fUSE3	frameshift vector			
fUSE5	frameshift vector			

[a]Each of these vectors, as a derivative of fd-tet (Zacher *et al.* 1980), carries a tetracycline-resistance gene. [b]Conventional, [c]amber, [d]two-amber and [e]frameshift vectors are explained in text and figures describing these vectors (see column "Reference"). See also Chapter 7.

2.2. GENERAL METHODS

2.2.1. Common reagents

Common reagents include: **PBS**, phosphate buffered saline, 12 mM phosphate, 157 mM Na^+, 4.4 mM K^+, 140 mM Cl^-, pH 7.4; **phenol**, reagent-grade phenol equilibriated with 0.1 M Tris pH 8.0, or a similar commercial preparation; **TBS**, 50 mM Tris·Cl pH 7.5, 150 mM NaCl; **TE**, 10 mM Tris·Cl, 1 mM ethylene diamine tetraacetate (EDTA), pH 7.4 or 8.0.

2.2.2. DNA manipulation

(a) Restriction digests. Digestions of DNA with restriction endonucleases (Boehringer, New England Biolabs, Pharmacia, Promega or Stratagene) were normally carried out in manufacturer-supplied buffer at recommended temperatures in 10-20 μL volumes. In most cases for incubation temperatures ≥50°C, reaction mixtures were overlaid with a small volume of sterile mineral oil. Commonly, ≤1 μg DNA possessing one or a few restriction sites per

molecule was digested ~1 h with a large excess (4-10 U) of restriction endonuclease.

(b) Ligations. Ligations with T4 ligase (GIBCO/BRL, NEB, Promega) were carried out in manufacturer-supplied buffer in a total volume of 20 μL. Ligase units (U) reported are Weiss units.

(c) Phenol and chloroform extraction of DNA from reaction mixtures. Most commonly, an equal volume of phenol was added to a DNA-containing sample (volume commonly adjusted to 100-450 μL with TE pH 7.4 or 8.0); these were mixed by repeated inversion or gentle vortexing. After brief centrifugation, the aqueous phase was similarly extracted with chloroform. Alternatively, a DNA sample was similarly treated with a mixture of equal parts of phenol and chloroform or a 25:24:1 mixture of phenol/chloroform/isoamyl alcohol.

(d) Precipitation with ethanol. DNA samples, most commonly in volumes ≤450 μL, were precipitated by addition of 2.2 volumes reagent grade ethanol and one-tenth volume 3M sodium acetate pH 5.2-6. After overnight incubation at 4°C (most commonly) or -20°C, or at least 1 h on ice (uncommonly), samples were centrifuged at maximum speed in an Eppendorf 5415C microfuge ≥0.5 h at 4°C. After washing in 70% ethanol and centrifuging at maximum speed ≥3 min. at room temperature or at 4°C, pellets were dried at room temperature (in some cases under low vacuum) before addition of a small volume (10-20 μL) of 10 mM Tris·Cl pH 8, TE pH 7.4 or 8, other slightly alkaline buffer (as employed for restriction digests) or water. Samples were stored overnight at 4°C to allow rehydration of DNA.

(e) Precipitation with isopropanol. DNA samples such as of amber vectors digested with restriction endonucleases to excise "stuffer" fragments (for examples, see Figures 2-2, 7-9 and 7-23) were precipitated with isopropanol to eliminate these small fragments, which tend to remain in solution. After adjusting the sample volume to (typically) 200-450 μL with TE pH 7.4 or 8.0, two-thirds volume of isopropanol and one-ninth volume of 3M sodium acetate

pH 6 were added. DNA was precipitated, washed and rehydrated as described for ethanol precipitation.

2.2.3. Agarose gel electrophoresis

DNA samples were routinely analysed by agarose gel electrophoresis at 2-7 V cm^{-1} in 5 cm x 8 cm or 11 cm x 14 cm 0.7%-0.75% agarose gels containing 0.5 μg ethidium bromide (EtBr) mL^{-1}, using TAE running buffer (Sambrook *et al.* 1989) containing 0.5 μg EtBr mL^{-1}.

2.2.4. DNA quantification

(a) By UV spectroscopy. Estimates assumed that an absorbance of 1.0 at 260 nm (A_{260}) corresponded to 50 μg mL^{-1} d.s. or 33 μg mL^{-1} s.s. DNA. Relative purity of a preparation was assessed by the ratio of absorbances at 260 (A_{260}) and 280 nm. A_{260}/A_{280} ratios for d.s. and s.s. DNA preparations were commonly ~1.8 and ~1.7 respectively.

(b) By agarose gel electrophoresis. Samples estimated to contain ≤200 ng of linear DNA fragments, together with commercially purchased standards of known quantity (*e.g.*, *Hind*III- or *Bst*EII-digested bacteriophage λ DNA) were analysed by agarose gel electrophoresis and the quantities of DNA determined by comparing the intensities of fluorescence of EtBr-stained samples with those of the standards.

(c) By an agarose gel "spot" assay. Samples (~5 μL) of dilutions of DNA to be quantified, together with commercially purchased standards of known quantity, were applied to solidified 0.7% agarose (with 0.5 μg EtBr mL^{-1}) and allowed to stand ≥0.5 h. Relative quantities of DNA were determined by intensities of fluorescence of samples and standards under UV light.

2.2.5. Oligonucleotides

(a) Synthesis and purification. Most synthetic oligonucleotides were prepared by the Nucleic Acid Protein Synthesis service (NAPS) unit of the University of British Columbia (UBC) Biotechnology Laboratory; in one case, by the R.E.W. Hancock laboratory at UBC.

Following synthesis, oligonucleotides were purified by the method of Sawadogo (1991).

(b) Annealing. Roughly equimolar mixtures of pairs or other combinations of synthetic oligonucleotides to be annealed (per sample: ~25-150 μM in 50-100 μL 200 mM Tris pH 7.5, 20 mM $MgCl_2$, 500 mM NaCl) were held in a TempBlock 80°-85°C for 1.5-15 min. Samples were cooled (over a ~1-1.5 h period) by removing the TempBlock from the heat source and placing it on the bench until the temperature reached ~30°C. Samples were stored at -20°C or 4°C.

2.3. DNA Extraction From Virion Hosts and Virions

2.3.1. RF DNA extraction from host cells

Except as noted elsewhere, RF DNA of phage vectors or recombinants was extracted from 30-37.5 mL serial overnight culture (OD_{600} commonly ~ 0.9 to 1.3) using a Qiagen Tip 20 plasmid DNA extraction/purification kit, following the manufacturer's recommended protocol for extraction of plasmid DNA, except as follows. First, cell pellets in the more recent preparations were washed in STE (100 mM Tris, 0.2 mM EDTA, 200 mM NaCl, pH 8.0) as recommended (Qiagen 1993) for M13 RF DNA. Second, volumes of reagents P1, P2 and P3 (similar to those used in common alkaline lysis protocols) were scaled up (from the recommended 0.3 mL to 0.9-4 mL, with 1.8 mL commonly used in recent preparations) to compensate for increased culture volume (30-50 mL *versus* the recommended 3-6 mL).

2.3.2. S.s. DNA extraction from virions

(a) Phenol extraction of fUSEn s.s. DNA. For each of fUSE1, fUSE3 and fUSE5, s.s. DNA (packaged into virions by G. P. Smith by means of a pIII-encoding helper plasmid) was extracted from Smith-supplied virions by phenol and chloroform extraction as described by Smith (1992).

(b) Preparation of s.s. sequencing template. S.s. DNA was extracted from two-stage (most commonly; §2.6.4) or one-stage (§2.6.3) PEG-precipitated virions derived from at least

15 mL (commonly 30-37 mL) serial overnight culture (§2.1.1) using a Qiagen Tip 20 plasmid DNA extraction/purification kit following the manufacturer's recommended procedures (Qiagen 1991) for extracting s.s. DNA from M13 virions except as follows. In most cases, particularly in more recent preparations, reagents L3, L4 and L5 were scaled up from 1.0 mL to 1.8 mL. Also, Qiagen reagent L5 was occasionally replaced with the similar reagent P3 (intended for RF DNA preparations) as convenient. Finally, in many early preparations, s.s. DNA recovered from Qiagen columns was additionally phenol/chloroform extracted and ethanol precipitated; such preparations generally yielded superior sequencing chromatograms.

2.4. SEQUENCING

2.4.1. Sequencing primers

Primers for sequencing d.s. RF or s.s. DNA included the synthetic oligonucleotide 5'-CCCTCATAGTTAGCGTAACG-3' (standard "A" primer) and the alternative (and generally less satisfactory) 5'-TGAATTTTCTGTATGAGG-3' (alternative "B" primer). These primers, designed for sequencing s.s. DNA of fUSE*n* recombinants (Smith 1992), are complementary to wild-type gIII sequences (of the virion-encapsidated or "sense" strand) and prime from a position 77-86 bases (primer "A") or 18-32 bases (primer "B") downstream (*i.e.*, in the 3' direction of the "sense" strand) of fUSE*n* cloning sites. Sequencing of one set of fDRW70 recombinants possessing relatively large *B. pertussis fha*B-derived inserts (siblings of clone ID I-a; Figure 6-1) also employed the primers 5'-GCTGACCGCCTCTCCACC-3' and 5'-CCTGCGGCAACCACGGTC-3'; these are complementary to sequences within the *fha*B-derived insert of clone I-a.

2.4.2. Sequencing reactions and gels

Sequencing was performed using Applied Biosystems DNA sequencers, Applied Biosystems Taq and Taq_{FS} polymerase and DyeDeoxy Terminator cycle sequencing reagents generally in accordance with the manufacturer's recommended protocols, except as noted

below. Sequencing performed with Applied Biosystems Taq (but not Taq_{FS}) polymerase employed, for s.s. template, roughly 5-fold more (or, in one particularly successful set of sequences, 2,000-fold more) sequencing primer than recommended. Most sequencing reactions were carried out with the recommended 25 thermocycles; in some cases, this was reduced to 20 cycles. Unincorporated nucleotides were removed using Select-D G-50 (5 Prime -> 3 Prime) or Centri-Sep (Princeton Separations) spin columns. Sequencing reactions (with Applied Biosystems Taq polymerase) of CSP-library clones, most fDRW*n* vectors, some fDRW*n* pseudorevertants[5], and many *B. pertussis fhaB* restriction fragment library clones were analysed using an Applied Biosystems model 370A sequencer. Other sequencing reactions (in particular those of *B. pertussis fhaB* DNase I fragment library clones; carried out with Taq or Taq_{FS} polymerase) were analysed by the Nucleic Acid Protein Synthesis service unit of the UBC Biotechnology Laboratory using Applied Biosystems sequencers. Where sequence ambiguities existed, these were resolved by re-sequencing or by examination of sequencing chromatograms in context of documentation (Applied Biosystems Inc.) of anomalous sequence-dependant incorporation of nucleotides.

2.5. HOST CELL TRANSFORMATION AND INFECTION

2.5.1. Calcium chloride transformation

(a) $CaCl_2$-competent host cells. *E. coli* K37, K802, K91 and LE392 were cultured in LB to an OD_{600} of ~0.35-0.55; the more slowly growing (Figure 2-1) K802 *recA*, to an OD_{600} of 0.2-0.5. After chilling on ice, 20 mL culture volumes were centrifuged 10-15 min. at ~500 x g_{max} (4°C) and cell pellets were gently resuspended in 0.5 to 2 volumes cold 0.1 M $CaCl_2$. After incubating ≥20 min. on ice and centrifuging as before, the resulting cell pellets were gently resuspended in ≤1 mL cold 0.1 M $CaCl_2$ and stored (4°C) up to five days before use.

---

[5]Pseudorevertants are described later in this chapter, and discussed in Chapter 7.

(b) Transformation. Small samples (commonly unquantified, in the ng to μg range) of s.s. or RF DNA were diluted to 200 μL with TE pH 8 and chilled in glass culture tubes before combining with 200 μL cold $CaCl_2$-competent *E. coli* host cells. After incubating 30 s at 37°C, ≥30 min. on ice, 1.5-2 min. at 42°C, ≥1 min. on ice or at room temperature, 1.6 mL SOC (Sambrook *et al.* 1989) or LB (containing 0.2 μg tetracycline[6] mL^{-1} for fUSE*n* or fDRW*n* and their recombinants) was added and the cells incubated 0.5-2 h (commonly 1 h) with shaking (≤225 rpm) at 37°C before spreading (commonly as 1, 10 and 100 μL aliquots) on selective medium (LB-Tet plates for fUSE*n* or fDRW*n* or their recombinants) and incubating overnight at 37°C. In some early transformations, medium containing 0.75% agar was overlaid on spread plates (Smith 1992).

2.5.2. Electroporation

(a) Electrocompetent host cells. Several related protocols were employed for preparing electrocompetent *E. coli* MC1061. In each case, large volume cultures (typically 500 mL LB [= 1 volume] in a 2 L Erlenmeyer flask) were inoculated from ≤10 mL overnight "starter" cultures and cultured to OD_{600} = 0.45-0.6. After 15 min. incubation on ice, 460 mL culture was transferred to chilled 250 mL centrifuge bottles and centrifuged 10-15 min. at 3,800-7,500 x g_{max} (4°C) to recover cells. These were washed once or twice by resuspension in 0.5-1.0 volume ice-cold water or 1 mM HEPES pH 7 followed by centrifugation as before, and additionally washed once by resuspension in 0.03-0.04 volume ice-cold 10% glycerol and centrifugation for 10-15 min. at 4,000-6,000 x gmax (4°C). After a final resuspension of cells in 0.002-0.003 volume ice-cold 10% glycerol, 50 μL aliquots (in 500 μL Eppendorf tubes) were frozen by pushing the tubes into pulverized dry ice. These were stored at -70°C to -80°C until immediately before use.

---

[6]This concentration of tetracycline does not harm host cells but induces expression of phage-encoded tetracycline resistance (see Smith, 1992).

(b) Electroporation. Electroporations were performed using a Biorad Gene Pulser with external Pulse Controller and 0.2 cm electroporation cuvettes. For each electroporation, up to 2.5 μL of DNA in ligation or other buffer, or up to 3.5 μL of DNA in water was added to 50 μL frozen electrocompetent *E. coli* MC1061 in 500 μL Eppendorf tubes (held on pulverized dry ice until immediately before use). After each cell/DNA mixture was thawed (by holding the tube between fingers) and mixed (by flicking the tube each ~10 s during thawing) and subsequently incubated on ice ~30 s, the mixture was transferred to a chilled cuvette within a chilled holder and electroporated at 2.5 kV, 200 or 400 Ω, 25 μF. Time constants (Miller 1994) were typically 4.5-4.6 (electroporations at 200 Ω) or 8.9-9.0 (at 400 Ω) for electroporations with DNA samples in water, and 4.1-4.3 (200 Ω) or 8.3-8.7 (400 Ω) for samples in ligation buffer. Immediately after electroporation, each mixture was transferred to 2 mL LB or SOC (pre-aliquoted in glass tubes, at room temperature; with 0.2 μg tetracycline mL^{-1} for fUSE*n* or fDRW*n* vectors and their recombinants, as in §2.5.1.b) and incubated 0.5-1 h at 37°C with rapid shaking (175-225 rpm) before spreading on LB-Tet in standard petri dishes (commonly 1 μL to 100 μL aliquots) or 245 mm x 245 mm NUNC culture dishes (≤2 mL aliquot); these were incubated ≥12 h at 37°C.

2.5.3. Infection of F-piliated host cells with virions

After a small quantity (1-40 μL) of generally unquantified virions (fUSE*n*, fDRW*n* or their recombinants) in culture supernatant, LB broth, PBS or TBS were combined with 50 μL to 2 mL *E. coli* K37, K91 or K91-Kan (all F-piliated) grown to visible turbidity (OD_{600} ≤0.9, commonly ~0.2-0.3), the cell/virion mixture was incubated 20 min.-1.5 h (37°C, standing) in medium containing 0.2 μg tetracycline mL^{-1} (see §2.5.1.b) to allow infection and expression of phage-encoded tetracycline resistance before transferring to LB broth containing 20 μg tetracycline mL^{-1} or spreading on LB-Tet plates.

2.6. HARVESTING OF VIRIONS

2.6.1. Harvesting virions from bacterial growth on solid media

Virions were harvested from colonies on solid media by adding TBS or PBS (most commonly), or LB to the media surface (*e.g.*, 10-30 mL for a 245 mm x 245 mm NUNC culture dish, 3-4 mL for a standard petri plate) and washing bacterial growth from the surface with a glass spreading rod. Virions within washings were recovered by centrifugation, as below.

2.6.2. Separation of virions in culture or wash supernatant from cells by centrifugation

Virions in culture supernatant or culture plate washings (§2.6.1) were separated from host cells by centrifugation (i) ≥3 min at maximum speed (4°C or room temperature) in an Eppendorf 5415C centrifuge, for volumes ≤2 mL, or (ii) ≥10 min at 3000-7500 x g_{max} (4°C) for volumes >2 mL. In some cases supernatants were incubated ≥10 min. at 68-70°C to kill residual host cells and centrifuged as before.

2.6.3. One-stage PEG precipitation

After cultures of phage-infected *E. coli* were centrifuged (as immediately above), virions were precipitated from culture supernatant with one of two formulation of polyethylene glycol (PEG)/sodium chloride: (i) PEG formulation "A" (30% w/v polyethylene glycol 8000, 3M NaCl (Qiagen 1991) or (ii) PEG formulation "B" (16.7% w/v polyethylene glycol 8000, 3.3 M NaCl; Scott & Smith 1990; Smith 1992). Typically, virions being precipitated for preparation of sequencing template were precipitated by adding 0.2 volume of PEG "A" to culture supernatant, mixing by 100 inversions, storing on ice ≥2 h (commonly ≥4 h) or at 4°C overnight, and centrifuging as described below; virions being precipitated for use in ELISA or other assays were commonly precipitated with PEG "B". After incubation on ice or at 4°C, (i) virion/PEG samples of volume ≤1.6 mL (in 1.5 or 2 mL Eppendorf tubes) were centrifuged

0.25 to 1 h (4°C) at maximum speed in an Eppendorf 5415C centrifuge; (ii) samples of volume ≤37.5 mL, at 12,000-34,000 x g_{max} (in recent preparations, 34,000 x g_{max}) ≥0.25 h (in recent preparations, 0.5-1 h) at 4°C in 50 mL Oakridge tubes; and (iii) samples ≥37.5 mL, ≥0.5 h (typically ≥1 h) at 12,000-22,000 x g_{max} (most commonly at 15,000 x g_{max}) at 4°C in 250 mL centrifuge bottles. Pellets were resuspended in 0.02-0.1 volume PBS, TBS or Qiagen buffer L3 (§2.3.2.b) as convenient and appropriate. In most cases samples were centrifuged again to remove insoluble debris: (i) volumes ≤2 mL were centrifuged ≥3 min. at maximum speed in an Eppendorf centrifuge (4°C or room temperature); (ii) larger volumes, ≥10 min. at 3,000-6,000 x g_{max} (4°C).

2.6.4. Two-stage PEG precipitation

Except in some cases where virions were to be further purified by CsCl density gradient (§2.6.5) or were being prepared to serve as a source of s.s. sequencing template, virions precipitated once by PEG precipitation were precipitated again by addition of PEG "B", incubation on ice ≥2 h or at 4°C overnight, and centrifugation as before. After resuspension in ≤1.8 mL PBS, TBS or Qiagen buffer L3, samples were centrifuged ≥3 minutes at maximum speed (4°C or room temperature) in an Eppendorf 5415C centrifuge to recover virion-containing supernatant.

2.6.5. CsCl density gradient centrifugation

CSP-library clones and selected other phage precipitated by two-stage (most cases) or one-stage PEG precipitation were further purified by CsCl density gradient centrifugation (Smith 1992). Virion yields from (typically) 200-600 mL overnight culture were suspended in 31% CsCl (w/w) in TBS (density = 1.30 g mL^{-1}) and centrifuged 48 h at 237,000 x g_{max} (37,000 rpm) in a SW41Ti rotor (5°C) before recovering a single faintly bluish band (when illuminated by lighting from one side) as CsCl-purified virions. After washing in ~20 mL TBS and centrifugation for 4 h at 257,000 x g_{max} (4°-5°C), virions were resuspended in 0.5 to 1.5 mL

PBS or TBS.

2.7. VIRION QUANTIFICATION AND ANALYSIS

2.7.1. Quantification of virion particles by ultraviolet (UV) spectroscopy (adapted from Smith 1992; see also Yamamoto *et al.* 1970)

Dilutions of PEG-precipitated or CsCl-purified virions in TBS or PBS were quantified by deducting sample absorbance at 320 nm (A320, believed due to light scattering of impurities; Smith 1992) from absorbance at 269 nm (A_{269}) and multiplying the net value (net A_{269}) by (i) 197 μg mL^{-1} to estimate total phage protein or (ii) 6.5 x 10^{12} virions mL^{-1} to estimate numbers of physical particles as fUSE2 (Figure 7-1) equivalents. Because genome size and thus virion length and protein content of a virion increase to accommodate foreign DNA inserted into gIII (with a concomitant increase in A_{269} per virion), the latter quantity -- physical particles as fUSE2 equivalents -- nominally over-estimates the numbers of physical particles of recombinant phage. For example, quantities of recombinants possessing 600 base pair (bp) inserts are overestimated by ~6%; those with 100 bp inserts by ~1%.

2.7.2. Quantification of plaque-forming units (pfu) and transducing units (TU)

(a) Full-plate plaque assay *(adapted from Smith 1992).* After adding 400 μL F-piliated *E. coli* K37, K91, K91-Kan, or K91/pNK1759 grown to an OD_{600} of 0.6-0.9 (most commonly 0.7-0.8, corresponding to late exponential or early stationary growth phase; Figure 2-1) to 100 μL culture supernatant or serial 10-fold or other dilutions (in culture media, PBS or TBS), 3 mL medium (commonly LB) containing 0.7-0.75% soft agar held at 48°-50°C was added by pipettor or by pouring from aliquots. The cell/virion/agar mixture was poured onto plates containing LB or other medium with 1.5% agar (in most experiments these plates had been warmed to 37°C) and held at room temperature until the soft agar solidified. Plates were incubated ≥7 h (37°C) before plaque-forming units (pfu) were counted.

Pfu of wild-type phage f1, conventional vectors (*e.g.*, fUSE2), and recombinants and

pseudorevertants of amber[7] and frame-shift[8] vectors were enumerated using *E. coli* K91 or K91-Kan as convenient. Early assays to enumerate total pfu (*i.e.*, both pseudorevertant and non-pseudorevertant pfu) of amber vectors employed the amber-suppressing *E. coli* K37. Because plaques on K37 are difficult to discern (indeed, fDRW613C produces no plaques on this strain; see Table 2-IV and Figure 7-14), later assays employed *E. coli* K91/pNK1759 which produces readily discernible plaques for fDRW613C and a greater number (than K37) of discernible plaques for other amber vectors (Table 2-IV).

Table 2-IV. Evaluation of host strains and media for plaque formation by amber vectors.

	Plaque formation[a] by host strain / medium			
	E. coli K37		*E. coli* K91/pNK1759	
	LB	2xYT	LB	2xYT
fDRW613	85 ± 26	difficult to discern[b] or none discernible	388 ± 47	405 ± 37
fDRW613C	none discernible[b]		451 ± 63	447 ± 50
fDRW70	605 ± 26		1119 ± 61	1181 ± 73

[a]For each of fDRW613, fDRW613C and fDRW70, full-plate plaque assays were performed in triplicate for 100 µL samples of a dilution of overnight culture supernatant, for (i) each of the two indicated host strains, and (ii) each of the two indicated media. Body of table shows mean number of plaques ± one standard deviation.

[b]Compared to plaques on *E. coli* K91/pNK1759, those on *E. coli* K37 are faint and difficult to discern, particularly for fDRW613C.

(b) Pfu "spot" assay. Lawns of F-piliated *E. coli* were prepared in essentially the same manner as described above (§2.7.2.a) except that LB broth was substituted for dilutions of culture supernatant. After allowing the top agar to solidify, 4-5 µL samples (singly, or in duplicate or triplicate) of undiluted (requires *E. coli* K37/pACYC184 or K91-Kan/pACYC184 or another tetracycline-resistant F-piliated strain) or diluted (tetracycline-resistant strain not

[7]Virions produced by amber vectors (containing an amber codon in gIII) nominally plaque only on amber-suppressing host strains. However, mutations commonly arise in which the amber codon is changed to a sense codon (*e.g.*, encoding Tyr). These "pseudorevertants to wild-type" are accordingly able to plaque on non-amber-suppressing host strains.

[8]Virions produced by frame-shift vectors (vectors containing a frameshift in gIII) are nominally non-infectious and accordingly produce no plaques. However, mutations commonly arise in which the gIII reading frame is restored. These "pseudorevertants to wild-type" are accordingly able to produce plaques.

required) culture supernatant were applied, as "spots", to the lawn. After excess liquid had been absorbed by the media, plates were incubated ≥7 h (37°C) before plaque-forming units were counted. This "spot" assay allowed up to 24 samples to be applied to a standard plate; in many cases, 40 or more plaques could be distinguished in a single spot.

(c) TU "spot" assay. This "transducing unit" (TU) assay derives from the ability of fUSE*n* and fDRW*n* vectors and recombinants to transduce tetracycline resistance into host cells. For recombinants deficient to some degree in their abilities to infect host cells, assays of TU are in principle more sensitive than assays of pfu, since the former require only a single infection event to give rise to a colony, while the latter require multiple rounds of virion propagation and infection to give rise to a plaque (Smith 1992).

Seventy-five to 100 µL *E. coli* K91-Kan grown to visible turbidity (OD_{600} = 0.25-0.7) were combined with 5-20 µL samples of serial dilutions (commonly 5- or 10-fold in LB, PBS or TBS) of culture supernatant in sterile 96-well microtiter plates. After 0.5-2 h incubation (37°C) in medium containing 0.2 µg tetracycline mL^{-1} (see §2.5.1.b), 4-5 µL samples ("spots") were applied, singly or in duplicate, to LB-Tet plates; these were incubated overnight before tetracycline-resistant colonies were counted as TU. This "spot" assay allowed up to 24 samples to be applied to a standard plate; in many cases, 20 or more colonies (TU) could be distinguished in a single spot.

2.7.3. Agarose gel electrophoresis of virions (adapted from Griess *et al.* 1990)

PEG-precipitated virions (§2.6.4) with sample loading buffer (0.25% bromophenol blue in 30% glycerol diluted 1:5 with virion sample) were electrophoresed ≥1.5 h at ~7 V cm^{-1} in 0.7% agarose, stained ~1.5 h in 0.05% (w/v) Coomassie, 10% glacial acetic acid, and de-stained with several washes in 10% glacial acetic acid over a period of several hours, or overnight, until virion bands could be satisfactorily discerned.

2.8. VECTOR CONSTRUCTION

2.8.1. Conventional vector fDRW20

Ten fmol fUSE1 RF DNA extracted from cultures screened for pseudorevertant production, digested to apparent completion with *Pvu*II and treated with alkaline phosphatase (Boehringer), was combined with 1 μmol (thus, a 100:1 molar insert to vector ratio) of an annealed (§2.2.5.b) self-complementary 5'-phosphorylated oligonucleotide (NEB) containing a *Bgl*II restriction site (Figure 7-12). Vector and insert were ligated (1 U T4 ligase, §2.2.2.b) 25 h at 14°-16°C before electroporating (§2.5.2) one-eighth of the ligation products into *E. coli* MC1061. After 48 transformants (isolates) cultured individually overnight in 200 μL LB-Tet in wells of microtiter plates were combined into 6 pools of 8 isolates, 0.8 mL of each pool was transferred to 32 mL fresh medium and incubated overnight. Samples of *Bgl*II-digested RF DNA extracted (§2.3.1) from these cultures were analysed by agarose gel electrophoresis to determine which pool contained the largest fraction of transformants possessing *Bgl*II sites. Each of the 8 isolates comprising the pool so identified was cultured by serial overnight culture (§2.1.1, 50 mL) before harvesting (§2.3.1) RF DNA from the final cultures.

With the goal of removing multimeric inserts, unquantified (assumed to be 5-6 μg) RF DNA was digested with *Bgl*II (50 U per isolate, 2-3 h) and precipitated with isopropanol to remove excised *Bgl*II fragments. After samples were analysed by agarose gel electrophoresis, 17 fmol RF DNA of each of the five (of eight) isolates that yielded the expected restriction pattern (a single 9.4 (kbp fragment) was re-ligated (10 U T4 ligase, §2.2.2.b) for 3.5 h at 22°-23°C and electroporated (§2.5.2) into *E. coli* MC1061. Six transformants were selected for each of the 5 isolates; after overnight culture in 200 μL LB-Tet in microtiter plate wells, 5 μL culture supernatant from each transformant was applied to a lawn of *E. coli* K91-Kan/pACYC184 in a manner similar to pfu "spot" assays (§2.7.2.b). After serial overnight culture (§2.1.1, 50 mL) of four of the 17 virion-producing transformants, s.s. DNA extracted

(§2.3.2.b) from virions harvested from culture supernatants by two-stage PEG precipitation (§2.6.4) was employed as sequencing template (§2.4) to identify all of the four sequenced clones as possessing the target fDRW20 sequence.

2.8.2. Conventional vectors fDRW21 and fDRW22

In separate ligations, 11 fmol fUSE1 RF DNA extracted from cultures that screened for low pseudorevertant production and digested to apparent completion with *Pvu*II was combined with 1 μmol and 5 μmol (thus 100:1 and 500:1 molar insert:vector ratios) of an annealed (§2.2.5.b) oligonucleotide pair containing a *Bgl*II restriction site (Figure 7-12). Each mixture was ligated (10 U T4 ligase, §2.2.2.b) 22 h at 14°-18°C before electroporating (§2.5.2) one-eighth of the ligation products into *E. coli* MC1061. After culturing 80 transformants (isolates) overnight in 200 μL LB-Tet, virion-producing isolates were identified by applying 10 μL samples of culture supernatant to lawns of *E. coli* K91-Kan/pACYC184 in a manner similar to pfu "spot" assays (§2.7.2.b). After serial overnight culture (§2.1.1, 50 mL) of each of the 11 plaque-producing isolates, *Bgl*II-digested samples of RF DNA extracted (§2.3.1) from these cultures were analysed by agarose gel electrophoresis. For each of the seven isolates whose RF DNA had digested with *Bgl*II, 10 μL culture supernatant was used to infect (§2.5.3) 50 μL *E. coli* K91-Kan, and infected cells were plated on LB-Tet and incubated overnight. For each of the seven isolates, a single colony was used as a source of inoculum for a serial overnight culture (§2.1.1, 37 mL). S.s. DNA extracted (§2.3.2.b) from virions harvested from culture supernatants by one-stage PEG precipitation (§2.6.3) was employed as sequencing template (§2.4) to identify three of the sequenced clones as possessing the target fDRW21 sequence and three others, the fDRW22 sequence.

2.8.3. Amber vector fDRW5

In separate ligations, 10 fmol fUSE5 RF DNA extracted from cultures screened for low pseudorevertant production, digested to apparent completion with *Sfi*I and precipitated with

isopropanol to eliminate "stuffer" fragments (Figure 7-1B), was combined with 30 and 90 fmol of an annealed (§2.2.5.b) oligonucleotide pair constituting a replacement amber-containing "stuffer" fragment (Figure 7-9). These mixtures were ligated (10 U T4 ligase, §2.2.2.b) 22 h at 10°-16°C before transforming (§2.5.1) one-half of the ligation products into *E. coli* K802. Virions produced by transformants were harvested by washing bacterial growth from plates (§2.6.1) and subsequent centrifugation. After heating wash supernatants 30 min. at 70°C (to kill residual host cells), 40 μL of a 10^{-2} dilution of supernatant was used to infect (§2.5.3) 50 μL *E. coli* K37, and dilutions of the infected cells were spread on LB-Tet and incubated overnight. After serial overnight culture (§2.1.1, 45 mL) of selected transformants, virions were recovered from culture supernatants by two-stage PEG precipitation (§2.6.4) and s.s. DNA was extracted (§2.3.2.b) and stored (4°C) for later use. After transforming this s.s. DNA into *E. coli* K802, single colonies (one per original transformant) were used as a source of inocula for serial overnight cultures (§2.1.1, 40 mL). S.s. DNA extracted (§2.3.2.b) from virions harvested from culture supernatants by one-stage PEG precipitation (§2.6.3) was employed as sequencing template (§2.4) to identify the two successfully sequenced clones as possessing the target fDRW5 sequence.

2.8.4. Amber vector fDRW613

In separate ligations, 17 fmol fDRW5 RF DNA digested to apparent completion with *Sfi*I and precipitated with isopropanol to eliminate "stuffer" fragments (Figure 7-9) was combined with 51 and 170 fmol of annealed (§2.2.5.b) oligonucleotide sets L1, M3 and R (Figure 2-2). These mixtures were ligated (0.5 U T4 ligase, §2.2.2.b) 20 h at 15°-18°C before transforming (§2.5.1) one-half of the ligation products into *E. coli* K802. Virions produced by transformants were harvested by washing bacterial growth from plates (§2.6.1) and subsequent centrifugation. Five μL of each wash supernatant was used to infect (§2.5.3) 120 μL *E. coli* host K37, and dilutions of the infected cells were spread on LB-Tet and

Figure shown on following page

Fig. 2-2. fDRW6*nn* amber vector series for constructing *Sau*3AI restriction fragment libraries. Initially, six fDRW6*nn* vectors were planned to accomplish design goals outlined in Chapter 7 (see Figure 7-14); each was to be constructed by inserting a combination of the oligonucleotide sets (**A**) L1, L2, L3, M1, M2, M3 and R into vector fDRW5 (Figure 7-9). Each of M1, M2, M3 contains an amber stop codon. Boxed sequences identify *Bsa*I restriction sites. *Bsa*I recognizes GGTCTC and cleaves to leave 5'-GATC overhangs within the sequences incorporated into the oligonucleotides:

```
5'....-GGTCTCN....-3'      5'-GATCNNN....-3'
3'....-CCAGAGNCTAG-5'      3'-....NNN....-5'
```

Examples of the initially planned vectors are shown in **B**. fDRW612 is comprised of oligonucleotide sets L1, M2 and R, and provides for inserts in reading frame 1 of length $3n + 1$ (where n is an integer); fDRW613 is comprised of L1, M3 and R, and provides for inserts in the same reading frame of length $3n + 2$ (or $3n - 1$); fDRW621, of L2, M1 and R, for inserts in reading frame 2 of length $3n + 1$; fDRW632, of L3, M2 and R, for inserts in reading frame 3 of length $3n + 1$. Only six of the nine possible combinations of oligonucleotide sets were planned; the remaining three allowed for inserts of length $3n$ and accordingly did not meet design objectives (Figure 7-14).

A later design incorporated two amber codons in the "M" sets of oligonucleotides. One of these oligonucleotide sets, MC3, is illustrated in **A**; vector fDRW613C derived from L1 and MC3 is illustrated in **C**. Wild-type sequences of fd-tet are shown in **D**.

The symbol Ⓟ identifies 5' phosphate groups.

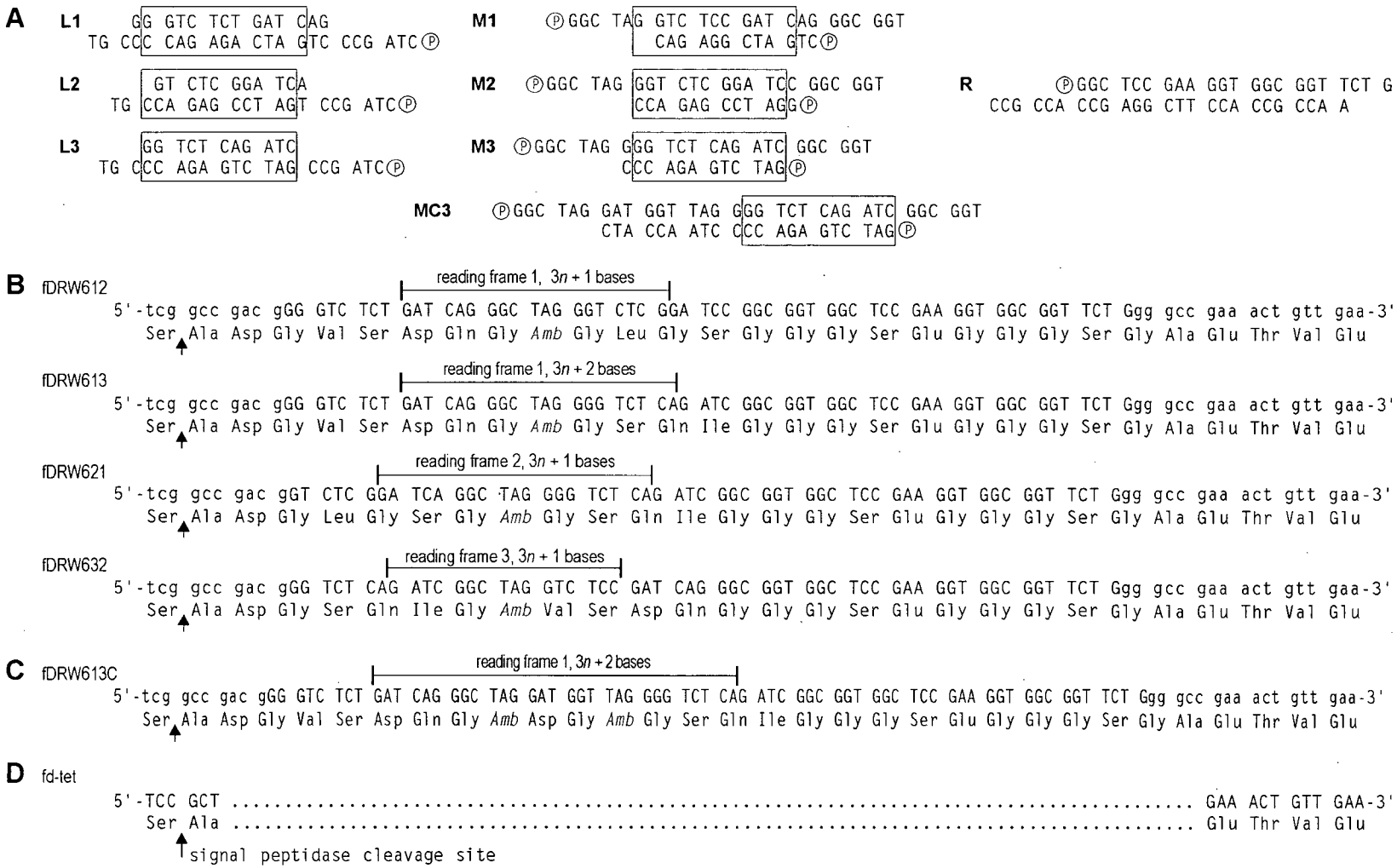

A
L1
GG GTC TCT GAT CAG
TG CCC CAG AGA CTA GTC CCG ATC Ⓟ
L2
GT CTC GGA TCA
TG CCA GAG CCT AGT CCG ATC Ⓟ
L3
GG TCT CAG ATC
TG CCC AGA GTC TAG CCG ATC Ⓟ
M1
Ⓟ GGC TAG GTC TCC GAT CAG GGC GGT
CAG AGG CTA GTC Ⓟ
M2
Ⓟ GGC TAG GGT CTC GGA TCC GGC GGT
CCA GAG CCT AGG Ⓟ
M3
Ⓟ GGC TAG GGG TCT CAG ATC GGC GGT
CCC AGA GTC TAG Ⓟ
MC3
Ⓟ GGC TAG GAT GGT TAG GGG TCT CAG ATC GGC GGT
CTA CCA ATC CCC AGA GTC TAG Ⓟ
R
Ⓟ GGC TCC GAA GGT GGC GGT TCT G
CCG CCA CCG AGG CTT CCA CCG CCA A
B
fDRW612
reading frame 1, 3n + 1 bases
5'-tcg gcc gac gGG GTC TCT GAT CAG GGC TAG GGT CTC GGA TCC GGC GGT GGC TCC GAA GGT GGC GGT TCT Ggg gcc gaa act gtt gaa-3'
Ser Ala Asp Gly Val Ser Asp Gln Gly Amb Gly Leu Gly Ser Gly Gly Gly Ser Glu Gly Gly Gly Ser Gly Ala Glu Thr Val Glu
fDRW613
reading frame 1, 3n + 2 bases
5'-tcg gcc gac gGG GTC TCT GAT CAG GGC TAG GGG TCT CAG ATC GGC GGT GGC TCC GAA GGT GGC GGT TCT Ggg gcc gaa act gtt gaa-3'
Ser Ala Asp Gly Val Ser Asp Gln Gly Amb Gly Ser Gln Ile Gly Gly Gly Ser Glu Gly Gly Gly Ser Gly Ala Glu Thr Val Glu
fDRW621
reading frame 2, 3n + 1 bases
5'-tcg gcc gac gGT CTC GGA TCA GGC TAG GGG TCT CAG ATC GGC GGT GGC TCC GAA GGT GGC GGT TCT Ggg gcc gaa act gtt gaa-3'
Ser Ala Asp Gly Leu Gly Ser Gly Amb Gly Ser Gln Ile Gly Gly Gly Ser Glu Gly Gly Gly Ser Gly Ala Glu Thr Val Glu
fDRW632
reading frame 3, 3n + 1 bases
5'-tcg gcc gac gGG TCT CAG ATC GGC TAG GTC TCC GAT CAG GGC GGT GGC TCC GAA GGT GGC GGT TCT Ggg gcc gaa act gtt gaa-3'
Ser Ala Asp Gly Ser Gln Ile Gly Amb Val Ser Asp Gln Gly Gly Gly Ser Glu Gly Gly Gly Ser Gly Ala Glu Thr Val Glu
C
fDRW613C
reading frame 1, 3n + 2 bases
5'-tcg gcc gac gGG GTC TCT GAT CAG GGC TAG GAT GGT TAG GGG TCT CAG ATC GGC GGT GGC TCC GAA GGT GGC GGT TCT Ggg gcc gaa act gtt gaa-3'
Ser Ala Asp Gly Val Ser Asp Gln Gly Amb Asp Gly Amb Gly Ser Gln Ile Gly Gly Gly Ser Glu Gly Gly Gly Ser Gly Ala Glu Thr Val Glu
D
fd-tet
5'-TCC GCT ..... GAA ACT GTT GAA-3'
Ser Ala ..... Glu Thr Val Glu
signal peptidase cleavage site

Fig. 2-2.

incubated overnight. Eight transductants (four from each ligation) were streaked on LB-Tet and, for each transductant, a single isolate was used as a source of inoculum for a serial overnight culture (§2.1.1, 42 mL). RF DNA extracted (§2.3.1) from these cultures was screened for the presence of a *Bsa*I restriction site by agarose gel electrophoresis of *Bsa*I-digested RF DNA. For each of the six isolates whose RF DNA had digested with *Bsa*I, s.s. DNA extracted (§2.3.2.b) from virions harvested from supernatants of the 42 mL cultures by two-stage PEG precipitation (§2.6.4) was employed as sequencing template (§2.4) to identify four of the six isolates as possessing the target fDRW613 sequence.

2.8.5. Amber vector fDRW613C

In separate ligations, 10 fmol fDRW5 RF DNA digested to apparent completion with *Sfi*I and precipitated with isopropanol to eliminate "stuffer" fragments (see Figure 7-9) was combined with 50 and 250 fmol annealed (§2.2.5.b) oligonucleotide sets L1, MC3 and R (Figure 2-2). These mixtures were ligated (5 U T4 ligase, §2.2.2.b) 18 h at 14.5°-15°C before transforming (§2.5.1) one-half of the ligation products into *E. coli* LE392. Twelve transformants (isolates; six per ligation) were cultured by serial overnight culture (§2.1.1, 40 mL) and *Bsa*I-digested RF DNA extracted (§2.3.1) from each of these cultures was screened for the presence of a *Bsa*I restriction site by agarose gel electrophoresis. Each of the eight isolates whose RF DNA had digested with *Bsa*I was again cultured by serial overnight culture (§2.1.1, ~38 mL) and s.s. DNA extracted (§2.3.2.b) from virions harvested by two-stage PEG precipitation (§2.6.4) of culture supernatants was employed as sequencing template (§2.4) to identify eight of the 12 isolates as possessing the target fDRW613C sequence.

2.8.6. Amber vector fDRW70

In duplicate ligations, 12 fmol fDRW5 RF DNA derived from cultures previously screened for low pseudorevertant production, digested to apparent completion with *Sfi*I and precipitated with isopropanol to eliminate "stuffer" fragments (Figure 7-9) was combined with

36 fmol of an annealed (§2.2.5.b) oligonucleotide pair constituting a replacement amber-containing stuffer fragment (Figure 7-23A). These mixtures were ligated (10 U T4 ligase, §2.2.2.b) overnight at ≥14°C before transforming (§2.5.1) one-half of the ligation products into *E. coli* LE392. RF DNA harvested (§2.3.1) from serial overnight cultures (§2.1.1, 40 mL) of six transformants (isolates) was screened for the presence of *Fsp*I, *Pvu*II and *Xba*I restriction sites and for the absence of an *Sfi*I site by agarose gel electrophoresis of RF DNA digested with these enzymes. For each of the four isolates whose RF DNA possessed the desired restriction patterns, s.s. DNA extracted (§2.3.2.b) from virions harvested from culture supernatants by one-stage PEG precipitation (§2.6.3) was employed as sequencing template (§2.4) to identify isolates possessing the target fDRW70 sequence.

2.8.7. Conventional vector series fDRW8*nn*

In separate ligations, 12 fmol fDRW5 RF DNA, derived from cultures of isolates whose sequences had been confirmed to contain an intact amber codon within the *Sfi*I-excisable "stuffer" fragment (see Figure 7-9), digested (possibly not to completion) with an excess of *Sfi*I (> 100 U μg^{-1} DNA) and precipitated with isopropanol to remove stuffer fragments, was combined with 12, 60, 300 and 1500 fmol semi-degenerate, incompletely complementary oligonucleotides (Figure 7-29B). These mixtures were ligated (5 U T4 ligase, §2.2.2.b) 4 h at 15°-16°C before precipitating the ligation products (with 10 μg tRNA added, per sample, as carrier) with ethanol and transforming (§2.5.1) the resuspended ligation products into *E. coli* LE392. Only 30 transformants were recovered. RF DNA extracted (§2.3.1) from serial overnight cultures (§2.1.1, 40 mL) of 16 of these was screened for the presence of a putative *Srf*I restriction site by agarose gel electrophoresis of *Sma*I-digested RF DNA (Figure 7-29B). For each of the 12 isolates whose RF DNA was digested with *Sma*I, undigested RF DNA was transformed into *E. coli* LE392 and a single transformant was selected and propagated by serial overnight culture (§2.1.1, 42 mL). S.s. DNA extracted (§2.3.2.b) from virions harvested from

culture supernatants by two-stage PEG precipitation (§2.6.4) was employed as sequencing template (§2.4) to identify the six unique fDRW8*nn* sequences shown in Figure 7-29. RF DNA extracted (§2.3.1) from the corresponding 42 mL cultures was employed (§2.12.3.1) to construct the *B. pertussis fhaB* DNase I fragment libraries summarized in Table 7-VI.

2.9. VECTOR STABILITY ISSUES

2.9.1. Initial studies of frameshift vectors

(a) fUSE5 propagated in* E. coli *K91-Kan (Figure 7-2A and B). fUSE5 RF DNA was extracted from three serial overnight cultures (§2.1.1, 500 mL) of *E. coli* K91-Kan that had been infected (§2.5.3) with fUSE5 virions (fUSE5 s.s. DNA packaged by G. P. Smith by means of a pIII-encoding helper plasmid), using a Qiagen Tip 500 DNA extraction/purification. Two series of *Sfi*I digests of this RF DNA were performed. In the first, ~90 μg RF DNA was digested with 100 U *Sfi*I (GIBCO/BRL) in a total volume of 290 μL. After 1 h and 2 h, a 3 μL sample was withdrawn, an additional 100 U *Sfi*I (in 10 μL) added, and digestion continued for an additional hour before analysing samples of undigested and digested DNA by agarose gel electrophoresis (Figure 7-2A). In the second series, ~1 μg samples were digested 2 h with 5, 10 and 20 U *Sfi*I (GIBCO/BRL) before analysis by agarose gel electrophoresis (Figure 7-2B).

(b) fUSEn propagated in* E. coli *K91-Kan (Figures 7-2C and 7-3). After infecting (§2.5.3) 0.5 mL *E. coli* K91-Kan with ~10^7 fUSE1, fUSE2, fUSE3 and fUSE5 virions (except for fUSE2, these had been packaged by G. P. Smith using a pIII-encoding helper plasmid), infected cells were spread on LB-Tet plates and incubated overnight. After washing the resulting bacterial growth from plates with cold PBS (20 mL per fUSE*n* vector), RF DNA was extracted from the wash using a Qiagen Tip 100 DNA extraction/purification kit using volumes of reagents appropriate for 180 mL overnight broth cultures. In a followup to the experiment described earlier (§2.9.1.a), ~1 μg aliquots of fUSE5 RF DNA were digested 2 h with 0, 0.5,

1, 2, 4 and 8 U *Sfi*I (Pharmacia) and analysed by agarose gel electrophoresis (Figure 7-2C). Separately, samples of fUSE1, fUSE2, fUSE3 and fUSE5 RF DNA were digested with *Pvu*II, *Bgl*II, *Xho*I or *Sfi*I as appropriate to each vector's cloning site and, together with undigested samples, were analysed by agarose gel electrophoresis (Figure 7-3).

(c) Initial assessment of pseudorevertant production by fUSE5 that had been passaged in E. coli K91-Kan (Figure 7-4A). After ~1-3 μg RF DNA of each of fUSE2 and fUSE5 propagated in *E. coli* K91-Kan (§2.9.1.b) was electroporated (§2.5.2) into *E. coli* MC1061, isolated colonies of *E. coli* MC1061 transformed with fUSE2 (two transformants) and fUSE5 (20 transformants) were cultured overnight in 2 mL LB-Tet. For each isolate, a single 5 μL sample of a 100-fold dilution of culture supernatant was applied to a lawn of *E. coli* K91-Kan in a manner similar to pfu "spot" assays (§2.7.2.b) and rough estimates of virion production were obtained from the numbers of plaques formed after >24 h incubation (37°C).

(d) Initial assessment of pseudorevertant production by fUSEn vectors provided by Smith (Figure 7-4B and C). After s.s. DNA extracted from Smith-supplied virions (§2.3.2.a) was transformed (§2.5.1) into *E. coli* K802 and transformants spread on plates of LB + 40 μg tetracycline mL^{-1}, selected transformants were assayed twice for virion production. In the first assay (Figure 7-4B) 12 isolates of each of fUSE1, fUSE3 and fUSE5 were cultured overnight in 2 mL LB-Tet. For each isolate a single 5 μL sample of undiluted culture supernatant was applied to a lawn of *E. coli* K91-Kan/pACYC184 in a manner similar to pfu "spot" assays (§2.7.2.b), and rough estimates of virion production obtained from the numbers of plaques formed after overnight incubation (37°C). In the second assay (Figure 7-4C), 18 isolates of fUSE1 and 24 isolates of each of fUSE3 and fUSE5 were cultured overnight in 2 mL LB-Tet. For each isolate, duplicate 5 μL samples of undiluted culture supernatant were applied to lawns of *E. coli* K91-Kan/pACYC184 and virion production assessed after overnight incubation.

(e) Use of chloramphenicol to reduce s.s. DNA species (Figure 7-5). After culturing

(to OD_{600}=1.3-1.5) four isolates of *E. coli* K802/fUSE5 that had been screened for low pseudo-revertant production, cultures were pooled and the pool used as an inoculum (1:50 dilution) for duplicate 110 mL LB-Tet cultures (labelled A and B). These were incubated 4 h (to OD_{600} ~0.7) before adding chloramphenicol to culture B (see Figure 7-5) to a final concentration of 15 μg mL^{-1} and incubating an additional 1 h (culture A, final OD_{600} =1.2) or 1.75 h (culture B, OD_{600}=0.8). RF DNA extracted from these cultures (using a Qiagen Tip 100 DNA extraction/purification kit with reagent P1, P2 and P3 volumes scaled up 1.2-fold for culture A) was analysed by agarose gel electrophoresis (Figure 7-5).

(f) Identification of s.s. bands with S1 nuclease (Figure 7-6). In a followup to the experiment described above (§2.9.1.e), 0.8-0.9 μg samples of DNA derived from culture A were treated: (i) with 7.2 U S1 nuclease (10 min. at room temperature, volume = 14 μL; in 12 mM Tris, 0.2 mM EDTA, 0.9 mM $MgCl_2$, 29 mM sodium acetate, 43 mM NaCl, 0.03 mM $ZnSO_4$, pH ≥4.5) , (ii) with 10 U *Bam*HI (1 h, 37°C) and (iii) with 10 U *Eco*RI (1 h, 37°C). Samples of DNA so treated, together with samples of undigested DNA from cultures A and B, were analysed by agarose gel electrophoresis (0.83 V cm^{-1} for 12 h, Figure 7-6).

(g) Elution of s.s. DNA with 1.35 M NaCl in Qiagen column wash (Figure 7-7). One-mL samples of 12 h cultures of 4 isolates of each of *E. coli* K802/fUSE1 and *E. coli* K802/fUSE5 that had been screened for low pseudorevertant production were diluted 1:50 into fresh medium (LB-Tet) and incubated to OD_{600} ~0.65. After chloramphenicol was added to a final concentration of 15 μg mL^{-1}, cultures were incubated an additional 2.5 h (final OD_{600}= 0.7-0.9) before harvesting DNA using a Qiagen Tip 20 DNA extraction/purification kit generally in accordance with the manufacturer's recommended procedures but with the following modifications: (i) 3 mL volumes of reagents P1, P2 and P3 were used in place of the 0.3 mL volumes recommended for 3-6 mL cultures; (ii) after the recommended wash with Qiagen reagent QC, an additional wash was performed with 3.8 mL QC-1.35 (reagent QC with

NaCl increased from 1.0 M to 1.35 M). Samples of DNA so extracted, together with samples of the QC-1.35 eluate (concentrated using a Centricon 30 microconcentrator) were analysed by agarose gel electrophoresis (Figure 7-7).

2.9.2. Evaluation of *E. coli* host strains for fDRW5 propagation (Table 7-I)

fDRW5 RF DNA was transformed (§2.5.1) into *E. coli* K37, K802 and LE392 and, for each strain, 6 transformants were selected for analysis. For each transformant 150-200 μL of overnight culture was used to inoculate a 50 mL LB-Tet culture. After 24 h incubation, RF DNA was extracted (§2.3.1) from 35 mL culture and samples of undigested RF DNA evaluated by the location and appearance of bands after agarose gel electrophoresis. Virions capable of plaquing on a non-amber-suppressing host (*viz.*, pseudorevertants) were counted by a pfu "spot" assay (§2.7.2.b) using duplicate samples of undiluted culture supernatant applied to *E. coli* K91-Kan/pACYC184.

2.9.3. Comparative analysis of frameshift and amber vector stability (Figure 7-15)

For each of fUSE1, fUSE2, fUSE3 and fUSE5, s.s. DNA extracted from Smith-supplied virions (§2.3.2.a) was transformed (§2.5.1) into *E. coli* K802 and -- in a preliminary analysis of vector stability (not shown) -- selected transformants (isolates) were assayed for virion production. For each vector, two or three isolates that produced few or no pseudorevertants were selected. RF DNA was extracted (§2.3.1) from serial overnight cultures (§2.1.1) of these isolates and stored (4°C) for later use, as follows. For the analyses summarized in Figures 7-15 and 7-16, aliquots of this RF DNA, as well as of s.s. DNA that had been employed as template for sequencing fDRW5, fDRW613 and fDRW613C, were transformed (§2.5.1) into *E. coli* K802 (fUSE1, fUSE2, fUSE3, fUSE5, fDRW5), *E. coli* K37 and K91 (fDRW5), and *E. coli* LE392 (fDRW5, fDRW613 and fDRW613C). For each of these nine host/vector combinations, three transformants were streaked on LB-Tet and three single colony isolates (one per transformant) selected for analysis; these were cultured overnight (37°C, standing) in 200 μL

LB-Tet in microtiter plates. For each isolate (see Figure 2-3 for a flow chart summarizing this and the following procedures) a 20 μL aliquot of the overnight culture was used to inoculate 43-45 mL LB-Tet (path A in Figure 2-3) while 10 μL aliquots were used to inoculate four replicate 200 μL cultures in microtiter plates (path B in Figure 2-3).

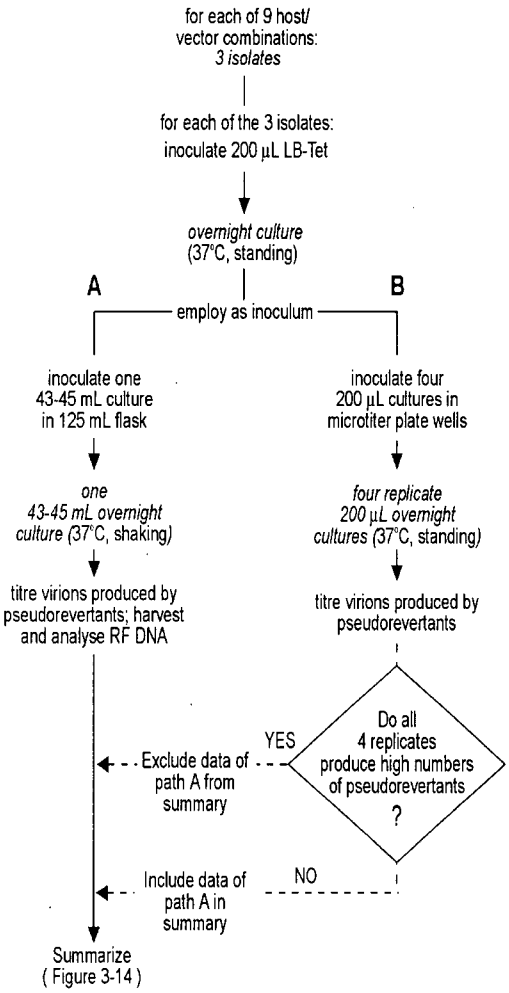


Fig. 2-3. Flow chart for experiment summarized in Figure 7-15, provided as an accompaniment to §2.9.3, "Comparative analysis of frameshift and amber vector stability".

(a) Path A (Figure 2-3). After 21.5 h incubation, virions within supernatant of each 43-45 mL culture were counted by a full-plate plaque assay (§2.7.2.a) using single samples of 10-fold serial dilutions of culture supernatant; *E. coli* K37 and K91-Kan were employed to determine total and pseudorevertant virion production respectively (Figure 7-15). A sample of undigested DNA extracted (§2.3.1) from each culture was analysed by agarose gel electrophoresis to confirm that no isolate was an obvious "deletion" mutant (seen previously

in certain frameshift vector preparations; see Figures 7-5 to 7-7) and to provide crude approximations of the quantities of RF DNA recovered.

(b) Path B (Figure 2-3). After ~30 h incubation (standing), virions within culture supernatant of the four replicate 200 μL cultures were similarly titered by a full-plate plaque assay (§2.7.2.a). These titers were used to assess whether any of the cultures used as inocula for the 43-45 mL cultures (path A) contained a large fraction of pseudorevertants, suggesting that the isolate itself may have been a pseudorevertant. In one case (an isolate of host/vector combination K37/fDRW5), titers for all four replicate 200 μL cultures were high, ~10^9 pfu mL^{-1} culture supernatant. Accordingly, results for this isolate were eliminated from the analysis (Figure 7-15) derived from cultures in path A.

2.9.4. Effects of SupF and wild-type pIII on pseudorevertant production

(a) Assays of pseudorevertant production (Figure 7-16A). S.s. DNA that had been employed as template for sequencing fDRW613, and RF DNA of fDRW5 isolates screened for low pseudorevertant production, were transformed (§2.5.1) into *E. coli* K802 and K802 *recA*. $CaCl_2$-competent cells prepared from these transformants and *E. coli* K91 were transformed with pJB61 (K802 and K802 *recA*), pNK1759 (K802 and K91) and pNK2859 (K802) and plated on LB + 100 μg ampicillin mL^{-1} + 20 μg tetracycline mL^{-1}. In separate assays for strains harboring fDRW5 and fDRW613, four isolates (isolated colonies of original transformants, or isolates derived from streak-purification of transformants) of each host/vector/plasmid combination were cultured (37°C, shaking) ~24 h in 2 mL LB + 20 μg tetracycline mL^{-1} and (for plasmid-harboring strains) 50 μg ampicillin mL^{-1}. A 50 μL aliquot of each culture was used to inoculate 40 mL (fDRW5) or 50 mL (fDRW613) LB containing 20 μg tetracycline mL^{-1} and (plasmid-harboring strains) 50 μg ampicillin mL^{-1}. After ~24 h incubation, virions were enumerated by a full-plate plaque assay (§2.7.2.a) using single samples of 10-fold serial dilutions of culture supernatant. *E. coli* K91/pNK1759 and K91 were

employed as plaquing hosts to measure total virion production and pseudorevertant production, respectively.

(b) Confirmation of gIII expression by pJB61 (Figure 7-16B). Because production of pIII prevents superinfection of host cells (Boeke *et al.* 1982), gIII expression in strains harboring pJB61 can be confirmed by comparing the abilities of virions to plaque on hosts possessing (plaques not expected) or lacking (plaques expected) pJB61. Accordingly, three serial 10-fold dilutions of fUSE2 (plaques on F-piliated host) and fDRW613 (requires amber-suppressing F-piliated host) expected to yield ~10^1-10^3 plaques on a suitable host were plaqued in a manner similar to full-plate plaque assays (§2.7.2.a) on *E. coli* K37 (amber-suppressing), K37/pJB61, K91-Kan (not amber-suppressing) and K91-Kan/pJB61, and plaque formation or lack thereof was determined.

2.9.5. Followup analysis of pseudorevertants and comparison of fUSE5 with fDRW5 (Figures 7-24 to 7-26)

(a) Initial cultures. Aliquots of fUSE5 s.s. DNA extracted (§2.3.2.a) from Smith-supplied virions, s.s. DNA that had been employed for sequencing fDRW5, fDRW613 and fDRW613C, and fDRW70 RF DNA extracted from cultures screened for low pseudorevertant production was transformed (§2.5.1) into *E. coli* LE392. For each vector, four transformants (isolates) were cultured (standing) in 200 μL LB-Tet in microtiter plates. After 24 h each culture was diluted serially by six ten-fold and ten two-fold dilutions (in LB-Tet) in microtiter plates. These were incubated 24 h (standing) before using (for each isolate) 100 μL of the highest two-fold dilution that showed clear evidence of growth (as determined by eye) to inoculate 2 mL LB-Tet. After 24 h incubation (with shaking) of the 2 mL cultures, (i) 0.5 mL culture supernatant was heated to 70°C (10 min.) and stored at 4°C for later use, and (ii) 1.5 mL culture used to inoculate 48 mL LB-Tet; these ~50 mL cultures were incubated 24 h (with shaking).

(b) Assay of pseudorevertant production (Figure 7-24). Virions present in culture supernatant of the 2 mL and 50 mL cultures were enumerated by a full-plate plaque assay (§2.7.2.a) using *E. coli* host strains K91 and K91/pNK1759 and, for each isolate, three samples from a single 10-fold dilution series of culture supernatant. For fUSE5 and fDRW5, each of the four isolates was titered; for the remaining vectors, three of the four isolates were titered. Virions produced during the 24 h incubation of 50 mL cultures were estimated by subtracting, for each isolate, mean titers of 3 samples of the 2 mL cultures from the corresponding titers of the 50 mL cultures (Figure 7-24).

(c) Analysis of fUSE5 and fDRW5 RF DNA (Figures 7-24 and 7-25). After RF DNA of each fUSE5 and fDRW5 isolate was extracted (§2.3.1; 4 mL P1, P2, P3) from a serial overnight culture (§2.1.1, 38 mL), one-half of the harvested DNA was treated with Plasmid-Safe ATP-dependent DNase (Epicentre Technologies; this enzyme degrades s.s. and linear d.s. forms of DNA; per isolate: 10 U, 37°C, 0.5 h, in Boehringer restriction endonuclease buffer M supplemented with 1 mM ATP, volume = 49 μL), and heated to 70°C for 15 min.. After subsequent digestion with *Bam*HI (20 U, 1 h, volume=51 μL) to create linear fragments, each isolate's RF DNA was quantified and examined for "deletion" mutants (seen previously in certain frameshift vector preparations, see Figures 7-5 to 7-7) by agarose gel electrophoresis.

(d) Sequencing of pseudorevertants (Figure 7-26). For each vector, three ~2 mm x ~2 mm agar blocks containing plaques were excised from lawns used to titer pseudorevertants. These were incubated overnight (4°C) in 1 mL PBS before using 50 μL of each resulting eluate to infect (§2.5.3) 50 μL *E. coli* K91-Kan, and subsequently using infected cells to inoculate 50 mL overnight (shaking) cultures. S.s. DNA extracted (§2.3.2.b) from virions harvested from culture supernatants by two-stage PEG precipitation (§2.6.4) was employed as template for sequencing (§2.4) the cloning site and flanking bases of each pseudorevertant.

2.10. CONSTRUCTION OF A LIBRARY OF VARIANT PEPTIDES DERIVED FROM THE *P. FALCIPARUM* CIRCUMSPOROZOITE PROTEIN (CSP-LIBRARY)

2.10.1. Construction of a clone displaying PNANPNANPNA

In each of two ligations, 10 fmol fUSE5 RF DNA derived from a culture screened for low pseudorevertant production, digested to apparent completion with *Sfi*I and precipitated with isopropanol to remove "stuffer" fragments, was combined with 30 fmol of an annealed (§2.2.5.b) oligonucleotide pair encoding the circumsporozoite protein (CSP) sequence PNANPNANPNA (Figure 7-8A). These mixtures were ligated (0.5 U T4 ligase, §2.2.2.b) 18 h at 13.5°-17.5°C before electroporating (§2.5.2) one-eighth of the ligation products into *E. coli* MC1061. After overnight culture (standing) of 96 transformants (48 per ligation) in wells of microtiter plates (200 μL culture^{-1}) virion-producing clones were identified by applying 5 μL culture supernatant to lawns of *E. coli* K91-Kan/pACYC184 in a manner similar to pfu "spot" assays (§2.7.2.b). Five of the 50 transformants that produced relatively higher numbers of plaques were selected for further analysis. RF DNA extracted (§2.3.1) from serial overnight cultures (§2.1.1, 50 mL) of these was digested with *Alw*NI (diagnostic for the target recombinant) to confirm that all possessed the desired insert. Sequencing template was prepared by (i) extracting s.s. DNA in the manner described earlier (§2.3.2.a) from one-stage PEG-precipitated virions (§2.6.3) derived from serial overnight cultures (§2.1.1, 100 mL) of these clones, and (ii) subsequently purifying DNA with a Qiagen Tip 20 DNA extraction/purification kit. For four transformants, this s.s. DNA was employed as sequencing template (§2.4) to confirm that all possessed the desired insert.

2.10.2. Construction of a first CSP-library

(a) fDRW5 propagation in E. coli K802. After transforming *E. coli* K802 with an aliquot of fDRW5 s.s. DNA that had been employed in sequencing this vector, transformants

were cultured (as a pool) by serial overnight culture (§2.1.1, 450 mL). Virions were harvested from culture supernatant by two-stage PEG precipitation (§2.6.4) and stored (4°C) for later use (§2.10.3.1). RF DNA was extracted from cells using, for each 210 mL culture volume, a Qiagen Tip 100 DNA extraction/purification kit generally in accordance with the manufacturer's recommendations, but with an additional column wash with 1.35 M NaCl (§2.9.1.g). After digesting ~20 μg fDRW5 RF DNA with 200 U *Sfi*I (≥1 h; NEB, volume = 200 μL), digested DNA was purified using a Qiagen Tip 20 DNA extraction/purification kit.

(b) Ligation and electroporation. In three ligations, 13 fmol fDRW5 RF DNA digested with *Sfi*I (possibly incompletely: see Figure 7-10A) and precipitated with isopropanol to remove "stuffer" fragments was combined with 13, 130 and 1,300 fmol annealed (§2.2.5.b) semidegenerate oligonucleotides encoding variants of the CSP immunodominant repeats (Figure 7-8B). These mixtures were ligated (5 U T4 ligase, §2.2.2.b) 17 h (15°-19°C) before electroporating (§2.5.2) one-eighth of the ligation products into *E. coli* MC1061.

(c) Isolation of virion-producing clones. For each ligation, virions comprising the resulting library were harvested by washing transformants (6×10^4 to 2.5×10^5 cfu) from plates (§2.6.1) and subsequent centrifugation. Virions were recovered by one-stage PEG precipitation (§2.6.3) of wash supernatants. For each library, an aliquot of harvested virions was used to infect (§2.5.3) *E. coli* K91-Kan and dilutions of infected cells (10^0 to 10^{-4}) were spread on LB-Tet plates and incubated overnight. For each library, 52 isolates selected from these plates were transferred as short streaks to master "stock" plates and incubated overnight before storing (4°C) for later use as a source of candidate clones for sequencing.

(d) Assessment of recombinant fraction. RF DNA extracted (§2.3.1) from serial overnight cultures (§2.1.1) of 12 candidate clones (4 per library) derived as described above was digested with *Hae*II to determine that, based on loss of a 3,965 base pair (bp) fragment and concomitant gain of ~2,890 and 1,075 bp fragments, all 12 clones were recombinants.

(e) Identification of unique clones by sequencing (Figure 7-11). Twenty-six of 32 possible (Figure 7-8) unique clones and one product of "illegitimate" recombination were identified by sequencing (§2.4) the gIII inserts of 63 candidate clones. Usually this was done by sequencing s.s. DNA extracted (§2.3.2.b) from one- or two-stage PEG-precipitated virions (§2.6.3, §2.6.4) produced by (1°) transductants (selected from master "stock" plates) in serial overnight cultures (§2.1.1, 30-50 mL). In other cases, such as when sequences were ambiguous, *E. coli* K802 was transformed with s.s. or RF DNA derived from these 1° transductants or *E. coli* K91-Kan was infected (§2.5.3) with their virion progeny, and sequencing template was prepared from virions produced by these 2° transformants or transductants in serial overnight cultures (§2.1.1, 30-50 mL).

2.10.3. Unsuccessful construction of a second, proline-biased CSP-library

(a) fDRW5 propagation in E. coli K37. After infection (§2.5.3) of 2 mL *E. coli* K37 with an aliquot of PEG-precipitated fDRW5 virions recovered (§2.10.2.a) from cultures used for propagation of fDRW5 RF DNA that had been employed in successful construction of the first CSP-library, infected cells were transferred to 40 mL fresh medium and incubated 6.5 h before using 20 mL of the resulting cultures to inoculate each of two 900 mL cultures. These were incubated 19 h before harvesting RF DNA using a Qiagen Tip 500 DNA extraction/purification kit. RF DNA was further purified by CsCl density gradient centrifugation by centrifuging recovered DNA (in 3.6 mL TE pH 8.0 with 50% w/w CsCl; 20°C) 19-20 h at 45,000 rpm (194,000 x g_{max}) in a Beckman VTi65 rotor, extracting the bottom most pronounced bands three times with NaCl-saturated butanol, and precipitating extracted DNA with ethanol. That the *Sfi*I restriction sites and amber codon within the *Sfi*I-excisable "stuffer" fragment were intact was confirmed by sequencing (§2.4) with 1.0 μg of CsCl-purified RF DNA.

(b) Ligation and electroporation. In separate ligations, 17 fmol CsCl-purified fDRW5

RF DNA, digested to apparent completion with *Sfi*I and precipitated with isopropanol to remove "stuffer" fragments (see Figure 7-10B), was combined with 51 and 170 fmol of an annealed (§2.2.5.b) semidegenerate oligonucleotide pair (encoding variants of the CSP immunodominant repeats, as before) with a bias in the bases (C or G) encoding proline *versus* alanine (see Figure 7-8) favouring proline over alanine ~4:1. These mixtures were ligated (5 U T4 ligase, §2.2.2.b) 20 h at ≥15°C before electroporating (§2.5.2) one-twentieth of the ligation products and, as a control, ~0.5 fmol undigested fDRW5 RF DNA into *E. coli* MC1061.

(c) Isolation and sequencing of virion-producing clones. For each electroporation, virions were harvested from bacterial growth on plates (1.5×10^3 to 2.3×10^3 cfu for ligation products, 4×10^5 cfu for undigested fDRW5 RF DNA) and virion-producing clones were selected and sequenced in a manner similar to that described for the first CSP-library (§2.10.2.c, §2.10.2.e).

2.10.4. Successful construction of a third, proline-biased CSP-library (Table 7-II)

(a) Ligation and electroporation. In three separate ligations, 10 fmol fDRW5 RF DNA propagated in *E. coli* LE392, digested to apparent completion with *Sfi*I and precipitated with isopropanol to eliminate "stuffer" fragments was combined with 0 (control), 30 and 90 fmol (*i.e.,* molar insert:vector ratios of 0:1, 3:1 and 9:1) of an annealed (§2.2.5.b) semidegenerate oligonucleotide pair prepared with a bias in the bases (C or G) encoding proline *versus* alanine (§2.10.3.b). These vector/insert mixtures were ligated (5 U T4 ligase, §2.2.2.b) 18 h at 15°-19°C before electroporating (§2.5.2) the following into *E. coli* MC1061: (i), (ii) and (iii), one-tenth of the products of ligations with 0:1, 3:1 and 9:1 molar insert:vector ratios respectively; (iv) one-twentieth of the products of the ligation with a 9:1 molar insert:vector ratio combined with 0.13 fmol undigested fDRW5 RF DNA; and (v) 0.13 fmol undigested fDRW5 RF DNA.

(b) Isolation and sequencing of virion-producing unique clones. For each electroporation, virions comprising the resulting library were harvested by washing

transformants from plates and PEG precipitating wash supernatant essentially as described for the first CSP-library (§2.10.2.c). Virion yields were estimated by a pfu "spot" assay (§2.7.2.b) using duplicate ~5 µL samples of 100-fold serial dilutions of wash supernatant applied to lawns of *E. coli* K91-Kan. Virion-producing recombinants were isolated essentially as described earlier (§2.10.2.c). RF DNA extracted (§2.3.1) from serial overnight cultures (§2.1.1) of six candidate clones was digested with *Hae*II to determine that, based on loss of a 3,965 bp fragment and concomitant gain of ~2,890 and 1,075 bp fragments, all six candidate clones were recombinants. Four previously unidentified recombinant clones were identified by sequencing the inserts into gIII of 19 candidate clones.

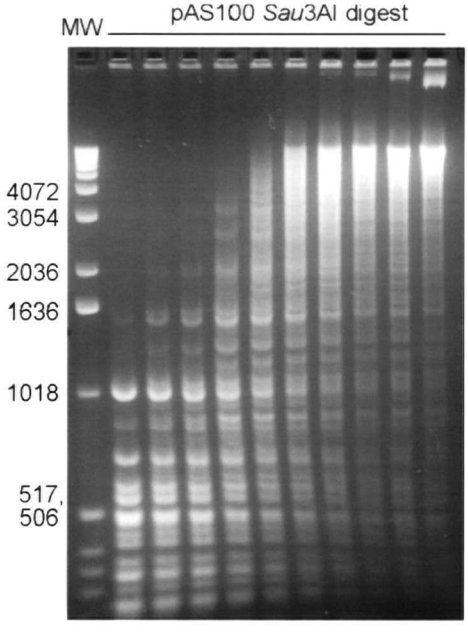


Fig. 2-4. Partial digest of pAS100 with *Sau*3AI. Agarose gel shows samples from one of two sets of partial digests: ~5.5 µg samples of pAS100 were incubated (0.5 h, 60 µL volumes) with (left to right) 10, 7, 5, 3.4, 2.4, 1.7, 1.2, 0.8, 0.6 and 0.4 U *Sau*3AI.

2.11. *B. PERTUSSIS FHAB* RESTRICTION FRAGMENT LIBRARIES

2.11.1. Construction and analysis of *B. pertussis fhaB Sau*3AI fragment (FHA-S) libraries

***(a) pAS100 propagation and partial digestion with* Sau*3AI*.** pAS100 (Table 2-II) DNA was extracted from *E. coli* DH5α/pAS100 (a gift from A. Siebers) cultured to OD_{600}=0.5 in 500 mL LB containing 100 µg ampicillin mL^{-1} using a Qiagen Tip 500 DNA

extraction/purification kit. Two sets of partial digests of pAS100 were performed. In the first, ten 5.5 μg samples (60 μL volumes) of pAS100 were digested 0.5 h with two-fold serial dilutions of *Sau*3AI (10 to 0.2 U). In the second set, ten 5.5 μg samples (in 60 μL volumes) were digested 0.5 h with a second set of 10 serial dilutions of *Sau*3AI (10, 7, 4.9, ..., 0.4 U). After analysis (Figure 2-4) by 1.5% agarose gel electrophoresis, digests were pooled and fragments separated on a preparative scale by 1.5% agarose gel electrophoresis (2 V cm^{-1}, 4-4.5 h). Gel bands containing fragments of ~1300-2300 bp were excised and placed in sections of dialysis tubing. Fragments were electroeluted (Sambrook *et al.* 1989) onto dialysis membrane by placing the closed tubing under TAE buffer in an electrophoresis chamber and applying 2.5 V cm^{-1} for ~2 h. After the current was reversed for 2 min., the fluid contents of the tubing together with 5 mL TAE washes were filtered through 0.2 μM Acrodisc filters and extracted repeatedly with *sec*-butanol until the sample volume was reduced to ~3.2 mL. Samples were subsequently ethanol precipitated and resuspended in TE pH 8 before being additionally purified by Select-D G-50 (5 Prime -> 3 Prime) spin columns and concentrated by ethanol precipitation.

(b) Ligation and electroporation. In four separate ligations (I, II, III and IV), 16 fmol of an equimolar mixture of fDRW20, fDRW21 and fDRW22 (Figure 7-12) RF DNA digested to apparent completion with *Bgl*II was combined with 21, 105, 560 and 2560 ng (roughly corresponding to 1:1, 5:1, 25:1 and 125:1 molar insert:vector ratios) of size-fractioned (1,300-2,300 bp) *Sau*3AI fragments of pAS100 (§2.11.1.a). These mixtures were ligated (10 U T4 ligase, §2.2.2.b) overnight at ≥14°C before electroporating (§2.5.2) one-eighth of the ligation products into *E. coli* MC1061.

(c) Initial characterization of virion production and virion-producing clones. For each resulting library (I, II, III and IV), 96 of the 1.4 x 10^5 to 1.8 x 10^5 transformants (Table 7-III) were cultured overnight in 200 μL LB-Tet in microtiter plate wells. For each of the 384

transformants so cultured, 5 μL of culture supernatant was applied to lawns of *E. coli* K91-Kan/pACYC184 in a manner similar to pfu "spot" assays (§2.7.2.b) to identify and characterize virion-producing clones. Twenty-four transformants (isolates) that produced a relatively diminished number of virions (<2 x 10^4 pfu mL^{-1} culture supernatant, explained in Table 7-III footnotes), together with a transformant that produced ≥2 x 10^4 pfu mL^{-1} and two that produced no detectable plaques (<500 pfu mL^{-1}), were selected for further analysis. After streaking on LB-Tet plates, a single-colony isolate of each was used as a source of inoculum for serial overnight culture (§2.1.1, 50 mL). RF DNA extracted (§2.3.1) from these cultures was digested with *Eco*RI, *Eco*RV and *Bam*HI and analysed by 0.6% agarose gel electrophoresis (i) to show that, except for 6 of the clones that produced a diminished number of virions, all of the 26 clones selected for analysis could be identified as recombinants, and (ii) to quantify RF DNA for use as sequencing template.

(d) Initial assessment of f1 and library virion migration in agarose gels (Figure 7-13). To assess virion migration (Griess *et al.* 1990) in agarose gels (Figure 7-13A), 32, 6.4, 1.3 and 0.3 μg samples of CsCl-purified (§2.6.5) f1 virions were electrophoresed and stained with Coomassie Blue (§2.7.3). In a followup experiment (Figure 7-13B), 6 and 0.6 μg samples of PEG-precipitated virions of each of the four FHA-S libraries were similarly electrophoresed and stained.

(e) Assessment by agarose gel electrophoresis of virion production by selected clones. Single colony isolates of six clones that produced fewer than 2 x 10^4 pfu mL^{-1} culture supernatant (a "diminished number of virions", §2.11.1.c) and one clone that produced ≥2 x 10^4 pfu mL^{-1} were cultured by serial overnight culture (§2.1.1, 50 mL). Separately, "stock" cultures of seven CSP-library clones (§2.10) were similarly used as inocula for serial overnight culture (§2.1.1, 40 mL). Virions of the 7 FHA-S library clones were harvested from culture supernatants by one-stage PEG precipitation (§2.6.3) and resuspended in 0.01 volume TBS.

The relative quantities of virions (as physical particles) within 13 μL samples of these putative 100-fold concentrated preparations and within 13 μL unconcentrated culture supernatant of two CSP-library clones were compared by examining Coomassie-stained agarose gels of electrophoresed virions (§2.7.3).

(f) Attempted enrichment for virion-producing clones by agarose gel electrophoresis. Three samples (6, 6 and 40 μg) of PEG-precipitated virions of library I (§2.11.1.e, Table 7-III) in sample loading buffer (§2.7.3) were loaded into the two sample and single preparative wells, respectively, of a 0.7% low melting point (LMP) agarose gel cast with a preparative comb, and electrophoresed ~4 h at 2 V cm-1 before staining the sample lanes with Coomassie Blue (§2.7.3). A 0.9 g agarose section of the unstained preparative lane, corresponding to the region between the trailing edge of the fast-running monophage band (Figure 7-13) and the leading edge of the fastest-running putative polyphage band, was excised and heated at 70°C for 15 min. (to melt LMP agarose; virions are relatively insensitive to heating: see Salivar *et al.* 1964) in 2 mL LB before cooling at 37°C for 10 min. (LMP agarose remains fluid at this temperature) and subsequently adding 2 mL *E. coli* K91-Kan (OD_{600} = 0.75). After 20 min. (37°C) the agarose/virion/cell mixture was spread on LB-Tet and incubated overnight (37°C). Virions were harvested by washing transductants from plates (§2.6.1) and subsequent two-stage PEG precipitation (§2.6.4) of wash supernatant. The effect of this enrichment method was examined by agarose gel electrophoresis and subsequent Coomassie staining (§2.7.3) of 6 and 0.6 μg samples of these "size-enriched" library I virions, together with 6 and 0.6 μg samples of unenriched library I virions.

2.11.2. Construction and analysis of *B. pertussis fhaB HhaI/HinP1I/HpaII* restriction fragment (FHA-H/H/H) libraries (Figures 7-17 to 7-22)

Methods of library construction described below were intended to allow the construction of a set of clones producing virions displaying an assortment (in both amino acid content and

peptide length) of peptides derived from the *B. pertussis fhaB* gene, *i.e.*, a "library" of diverse peptides. Certain methods reflect a goal of eliminating non-recombinants without regard to the effect of these methods on library content; *viz.*, not all possible peptides needed to be represented.

(a) fUSE1 propagation. After *E. coli* K802 was transformed with fUSE1 RF DNA extracted (§2.3.1) from a culture of an isolate previously screened for low pseudorevertant production, serial overnight cultures (§2.1.1, 40 mL) were prepared for three transformants. Fifty-μL samples (one per transformant) of ten-fold serial dilutions ($100\text{-}10^{-4}$) of culture supernatants were assayed for production of pseudorevertant, using a full-plate plaque assay (§2.7.2.a); no plaques were detected (limit of detection = 20 pfu mL^{-1} culture supernatant) for the three transformants.

For each transformant, ~1 μg RF DNA extracted (§2.3.1) from the 40 mL culture was incubated with 5 U Mung Bean Nuclease (NEB; in 10 mM Tris, 10 mM $MgCl_2$, 100 mM NaCl, 1 mM DTT, 1 mM $ZnCl_2$; pH 7.9; 70 μL volume) for 30 min. at 30°C to digest s.s. species in the preparation and, after recovering DNA by phenol/chloroform extraction and ethanol precipitation, each sample was digested with 10 U *Pvu*II (2 h). In spite of an excess of *Pvu*II (10 U per μg DNA), each preparation showed evidence of an incomplete digest as judged by agarose gel electrophoresis. Accordingly, samples were pooled and the ~3 μg RF DNA digested again with 100 U *Pvu*II (2 h, 300 μL volume) before recovering DNA by phenol/chloroform extraction and ethanol precipitation.

(b) pAS100 propagation and generation of blunt-end fragments. Plasmid pAS100, carrying the 10 kbp *Eco*RI fragment of *B. pertussis fhaB* (see Table 2-II), was transformed (§2.5.1) into *E. coli* K802, transformants were plated on LB with 100 μg ampicillin mL^{-1}, and plasmid DNA extracted from serial overnight cultures (§2.1.1, 125 mL LB + 50 μg ampicillin mL^{-1}) of each of two transformants. For each of *Hha*I, *Hin*P1I and *Hpa*II, ~10 μg

of pooled pAS100 DNA was digested 1 h with ~30 U restriction endonuclease before recovering DNA by phenol/chloroform extraction and ethanol precipitation. For each digest, ~5-10 μg DNA was incubated with 10 U Mung Bean Nuclease (NEB) (30°C, 1 h, in 50 mM sodium acetate, 30 mM NaCl, 1 mM $ZnSO_4$, pH 5; volume = 200 μL) to remove 3' and 5' overhangs. As well, ~1 μg of each of the *Hin*P1I and *Hpa*II digests was incubated with 5 U *E. coli* DNA Polymerase I large (Klenow) fragment (NEB) (15 min., 20°-25°C, in 10 mM Tris, 5 mM $MgCl_2$, 7.5 mM DTT, 0.2 mM dNTPs, pH 7.5, volume = 60 μL), to fill in 3' recessed ends. DNA in each sample of Mung bean nuclease (MBN)- and Klenow-treated fragments was recovered by phenol/chloroform extraction and ethanol precipitation.

***(c) Primary libraries: ligation, re-digestion with* PvuII *and transformation of* E. coli K802.** Five libraries were constructed: libraries I, II and III, with MBN-treated *Hha*I, *Hin*P1I and *Hpa*II fragments, respectively; and libraries IV and V, with Klenow-treated *Hin*P1I and *Hpa*II fragments. For each library, ligations (§2.2.2.b) with 3:1 and 9:1 molar insert:vector ratios (per ligation: ~75 or ~225 fmol restriction fragments, 25 fmol fUSE1, 1.8 U T4 ligase) were carried out for 21 h at 13.5°-17.5°C. Ligation products were heated 10 min. at 70°C to inactivate T4 ligase and digested with 10 U *Pvu*II (volume = 200 μL) to diminish non-recombinants. For each library, one-fourth of the ligation products were pooled and transformed (§2.5.1) into *E. coli* K802. For each transformation, 5 x 10^2 to 2 x 10^3 transformants were recovered on LB-Tet plates.

(d) Secondary libraries: re-digestion of DNA harvested from transformants with PvuII and subsequent transformation of E. coli K802. For each primary library, transformants were washed from plates (§2.6.1) and RF DNA harvested (§2.3.1) from these washings was digested with an apparent excess of *Pvu*II (26 U, 2 h, volume = 30 μL) to diminish non-recombinants. Products of this digestion were transformed (§2.5.1) into *E. coli* K802. For each transformation, 4 x 10^4 to 4 x 10^5 transformants were recovered on LB-Tet plates.

(e) Screening for virion production by a TU assay (Figure 7-17). Twenty-four transformants from each secondary library were cultured overnight in 1 mL LB-Tet and the numbers of virions produced estimated by a TU "spot" assay (§2.7.2.c) using a single sample for each 10-fold serial dilution (10^{-1} to 10^{-4}) of culture supernatant. Each of the 45 clones that produced a detectable number of virions (limit of detection = 10^4 TU mL^{-1} culture supernatant) was streaked on LB-Tet and incubated overnight before transferring individual colonies as short streaks to a "master" plate (LB-Tet). After overnight incubation, this master plate was stored at 4°C and used as a source of inocula for subsequent experiments.

(f) Extraction of RF DNA from, characterization of virion production by (Figure 7-18), and preparation of PEG-precipitated virion stocks of virion-producing clones. Each of 45 virion-producing clones identified in Figure 7-17 was cultured 36 h (Experiment 1) or 24 h (Experiment 2) in 1 mL LB-Tet. One mL (Experiment 1) or 0.1 mL (Experiment 2) of the resulting cultures was diluted into 40 mL LB-Tet and incubated 20 h. RF DNA extracted (§2.3.1; 3 mL P1, P2, P3) from cultures for Experiment 1 was stored at 4°C for later analysis (§2.11.2.7.a, §2.11.3.1). The numbers of virions produced by the 45 clones (Experiments 1 and 2) were estimated by a TU "spot" assay (§2.7.2.c) using a single sample for each 5-fold serial dilution (5^0 to 5^{-11}) of culture supernatant. For Experiment 1, virions were harvested from ~30 mL culture supernatant by two-stage PEG precipitation (§2.6.4) and stored at 4°C for later use.

(g) Determination of insert size by electrophoresis of HaeIII fragments. *Hae*III digestion of fUSE1 produces 8 fragments (352 to 2528 bp) including an 850 bp fragment which contains the *Pvu*II cloning site. With the goal of determining the sizes of inserts on the basis of changes in size of the 850 bp fragment, RF DNA of virion-producing clones (§2.11.2.f) was digested with *Hae*III (3-5.25 h) and restriction fragments were analysed by 1.7% agarose gel electrophoresis. Only 17 clones produced restriction patterns that could

reasonably be analysed; for the remaining clones, DNA preparations were of poor quality (a "smear" of high- to low-molecular weight species was pronounced, unexpected species were present) or of low yield, or no new fragments could be discerned within the range of molecular weight markers employed and under the electrophoresis conditions employed. For each of these 17 clones, sizes of fragments containing inserts were estimated using the relative mobility (R_f) of these fragments compared to the R_f of known fragments, by means of a least squares regression of a fourth order polynomial (see legend to Figure 2-5). Standard curves based on this regression analysis are shown in Figure 2-5.

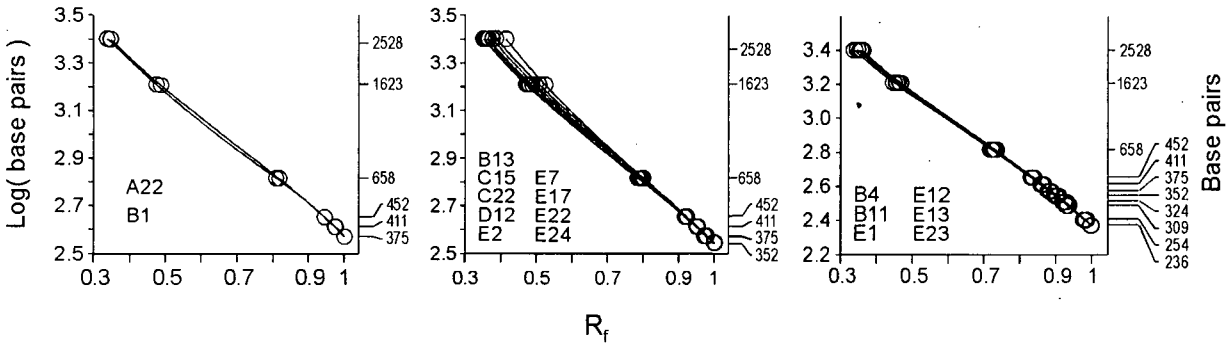


Fig. 2-5. Relative mobility (R_f) of *Hae*III fragments of FHA-H/H/H library clones. Figure is derived from agarose gel electrophoresis of *Hae*III restriction fragments of 17 virion-producing clones (§2.11.2.g). Clones are shown in groups; in each group a common range of known fragment sizes was used to derive a
fourth order polynomial employed in estimating sizes of pAS100-derived inserts:

$$\log(L) = a_0+a_1R_f+a_2R_f^2+a_3R_f^3+a_4R_f^4$$

where L = length of fragment in base pairs;
R_f = relative mobility; a_0, a1, a_2, a_3 and a_4 are constants.

(h) Determination of insert size by sequencing. Nine virion-producing clones were selected for sequencing. For each, PEG-precipitated virions (§2.11.2.f) were amplified (§2.5.3) in *E. coli* K91-Kan, and s.s. DNA was extracted (§2.3.2.b) from virions harvested from the resulting culture supernatant by two-stage PEG precipitation (§2.6.4). Of the nine samples, only five produced s.s. DNA of sufficient quantity (this correlated roughly with levels of virion production) or quality to be successfully employed as sequencing template (§2.4).

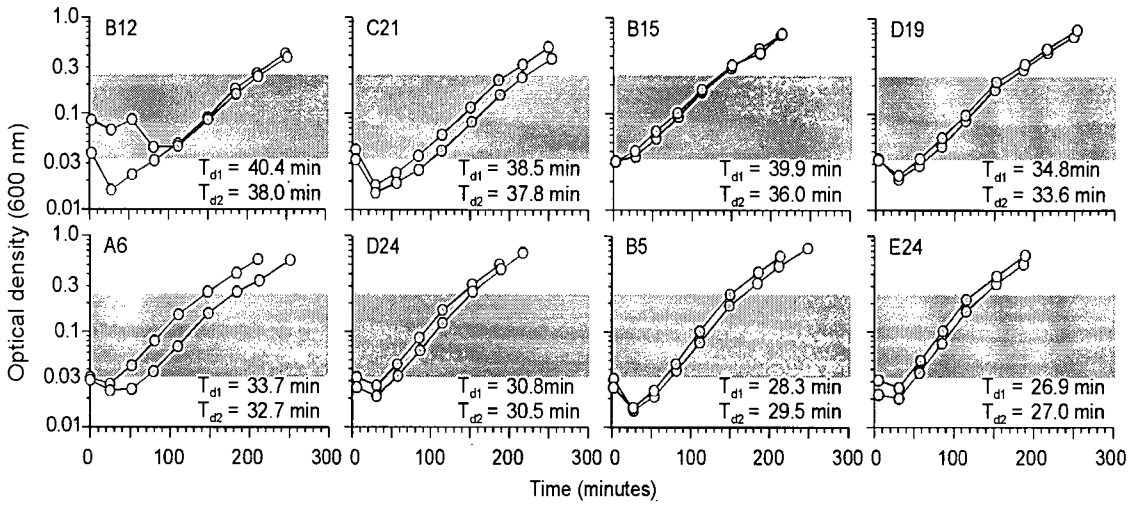


Fig. 2-6. Host cell doubling times (T_d) for FHA-H/H/H library clones. Culture conditions and other methodology are described in §2.11.2.i. Shaded bands highlight a common range of OD_{600} values employed in calculating T_d values for each of nine clones. T_{d1} and T_{d2} are estimates for duplicate cultures.

(i) Relationship between virion production and host cell doubling time (Figure 7-20).

For 12 clones, samples of duplicate 20 h cultures (1 mL LB-Tet) inoculated from a "master" plate (§2.11.2.e) were diluted into 40 mL fresh medium to yield a starting $OD_{600}\approx0.006$; these cultures were incubated with shaking (175 rpm) and OD_{600} measured at ~30 min. intervals using 1 mL samples (discarded) in 1 cm cuvettes. Doubling time (T_d) for each duplicate was calculated on the basis of a least squares linear regression of $\log(OD_{600})$ *versus* time for three data points with OD_{600} values between 0.04 and 0.3 and with $\log(OD_{600})$ *versus* time roughly forming a straight line (Figure 2-6). In a separate assay, duplicate small volume overnight cultures of the 12 clones were diluted 1:400 into 40 mL fresh medium (LB-Tet) and incubated overnight. The numbers of virions produced by these clones were estimated by a TU "spot" assay (§2.7.2.c) using a single sample for each 5-fold serial dilution (5^0 to 5^{-11}) of supernatant of each duplicate culture.

2.11.3. Followup study of FHA-H/H/H library clones (Figures 7-21 and 7-22)

Figure 2-7 provides an overview of the methods described in the following sub-sections.

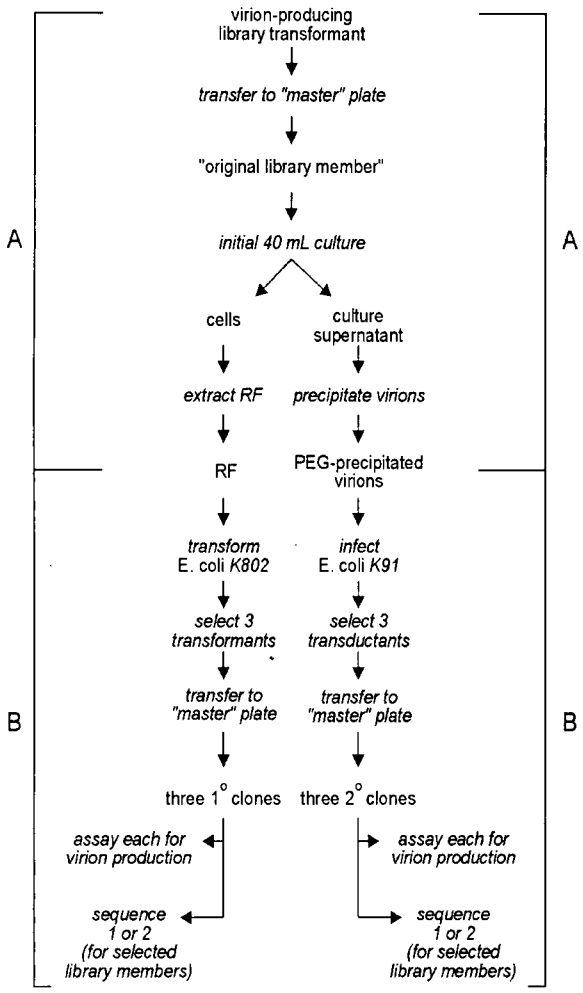


Fig. 2-7. Flow chart for followup study of FHA-H/H/H library clones, provided as an accompaniment to §2.11.3, the text of which is summarized in **B**. Origin of the material used to recover the "original" library members and derive sub-clones from their virion progeny is described in §2.11.2.f ("Extraction of RF DNA from ....") and is outlined in **A**.

(a) Recovery of "original" library members and derivation of sub-clones from their virion progeny. Forty "original" library members were recovered by transforming (§2.5.1) *E. coli* K802 with RF DNA extracted from the initial cultures of virion-producing FHA-H/H/H library clones (§2.11.2.f); for each of the 40 transformations, three transformants ("1° clones") were transferred as short streaks to "master" plates (LB-Tet), incubated overnight and stored (4°C) for later use. Sub-clones of these "original" library members were derived by infecting (§2.5.3) 100 μL *E. coli* K91-Kan with small aliquots of PEG-precipitated virions harvested from the same initial cultures (§2.11.2.f), and spreading serial 10-fold dilutions of infected cells on LB-Tet plates. After overnight incubation, three transductants ("2° clones") of each library member were transferred to "master" plates (LB-Tet), incubated overnight and stored (4°C) for

later use. As required, master plates were sub-cultured onto fresh LB-Tet plates and again stored (4°C) after overnight incubation.

(b) Comparison of virion production by 1° and 2° clones (Figure 7-21). For each of three 1° clones (in *E. coli* K802) and three 2° clones (in *E. coli* K91) of the 40 library members, sterile toothpicks were touched to "master" plates and used to inoculate 2 mL LB-Tet cultures. After 18 hours incubation, virions produced in these cultures were estimated by a TU "spot" assay (§2.7.2.c) using a single sample for each 10-fold serial dilution (10^{-1} to 10^{-6}) of culture supernatant.

(c) Sequencing of selected 1° and 2° clones (Figure 7-22). Nine library members were selected for sequencing. For four of these, virion titers of the 1° clones were substantially lower than those of the corresponding 2° clones. For the remaining five, titers of the 1° and 2° clones were similar. For the former group, attempts were made to sequence two 1° and two 2° clones; for the latter group, a single 1° and single 2° clone. Two methods were employed to extract RF DNA from these clones for use as sequencing template. In the first of these, employed for 26 preparations, RF DNA was extracted from 18 mL of serial overnight cultures (§2.1.1, 35-40 mL) inoculated from master plates, using Nucleobond AX-20 columns according to the manufacturer's recommended protocol for a three-fold scale-up of reagents. Extracted DNA was treated with ATP-dependant DNase (see §2.9.5.c; per preparation: 10 U, 37°C, 1 h, in manufacturer-supplied buffer supplemented with 1 mM ATP, volume=405 μL), incubated 20 min. at 70°C, and again purified (to remove enzyme and degradation products) with Nucleobond AX-20 columns. In the second method, employed for 25 preparations, RF DNA was extracted using Qiagen reagents (§2.3.1; 4 mL P1, P2, P3). After treating extracted DNA with ATP-dependant DNase (see §2.9.5.c; 5 U, 37°C, 4.75 h, volume=12.5 μL), DNA was recovered from reaction mixtures by phenol/chloroform extraction and ethanol precipitation. Sequencing (§2.4, using primer A for most samples, primer B for selected samples) results

were generally poor and repeated efforts to sequence certain clones were unsuccessful. Of 51 sequencing samples, only 16 (30%) provided sequencing chromatograms that were marginally acceptable or better. Limited sequencing data that included sequences for 1° clones were obtained for only five of the nine library members selected for analysis.

2.12. CONSTRUCTION OF *B. PERTUSSIS FHAB* DNASE I FRAGMENT LIBRARIES

2.12.1. Construction and analysis of FHA series 70 libraries (Table 7-V)

(a) fDRW70 propagation. Four samples of RF DNA extracted from cultures that had been used to prepare sequencing template during fDRW70 vector construction were transformed (§2.5.1) into *E. coli* LE392 and for each sample a single transformant was used as a source of inoculum for a serial overnight culture (§2.1.1, 30 mL). Samples of RF DNA extracted (§2.3.1) from each of these cultures were digested ~2 h with *Fsp*I, *Pvu*II and *Xba*I (separately) and analysed by agarose gel electrophoresis to confirm that the *Fsp*I and *Pvu*II sites flanking the excisable stuffer fragment and amber codon (*Xba*I used for diagnosis) within the stuffer fragment (Figure 7-23A) were intact. Larger-scale samples (~10 μg) were doubly digested 3.25 h with *Fsp*I and *Pvu*II (50 U each, 100 μL total volume), phenol/chloroform extracted and isopropanol precipitated (to remove "stuffer" fragments); these preparations were pooled and used in library construction, as below.

(b) Ligation, electroporation and harvesting of virions (Table 7-V). Four libraries were constructed with DNase I-generated fragments (Anderson 1981; Smith 1992) of the 10 kbp *B. pertussis fhaB Eco*RI restriction fragment (see Table 2-II, entry for pAS100). These fragments (a gift from A. Siebers) had been size-fractioned (extracted from polyacrylamide gels after electrophoresis, by A. Siebers) into groups of 30-75 bp (used for library 70-A), 75-150 bp (70-B), 150-300 bp (70-C) and 300-600 bp (library 70-D). For each library, two ligations were performed, each with 30 fmol fDRW70 RF DNA and ~90 fmol of *B. pertussis fhaB* fragments (thus roughly 3:1 molar insert:vector ratios). Two control ligations (70-X and 70-Y)

were also carried out, each consisting of 30 fmol fDRW70 RF DNA alone (*i.e.*, no *fhaB* fragments were included). After ligating (2 U T4 ligase, §2.2.2.b) each vector/insert mixture or control 33 h at ≥13°C, one-tenth (first ligation for each library) or one-twentieth (second ligation) of the ligation products were electroporated (§2.5.2) into *E. coli*. For each ligation, the number of transformants was estimated by counting colonies on a single LB-Tet plate spread with a 10 μL sample (Table 7-V). After washing transformants from plates (§2.6.1) and subsequent centrifugation of the washings, virion yields were estimated by full-plate plaque assays (§2.7.2.a) using single samples of 100-fold serial dilutions of wash supernatant (Table 7-V). Virions were subsequently harvested by two-stage PEG precipitation (§2.6.4) of wash supernatant.

(c) Library evaluation (Figures 7-27 and 7-28). To identify virion-producing transformants by a tetracycline-resistance transduction assay, toothpicks were used to transfer growth from colonies of 24 transformants derived from each ligation for libraries 70-A, 70-B, 70-C, 70-D and 70-X to 20 μL *E. coli* K91-Kan (OD_{600} ~ 1.0) in LB + 0.2 μg tetracycline mL^{-1} + 100 μg kanamycin mL^{-1} (to inhibit growth of *E. coli* MC1061) in wells of 96-well microtiter plates. After 0.5 h incubation (standing), 180 μL LB + 20 μg tetracycline mL^{-1} + 100 μg kanamycin mL^{-1} was added to each well and plates were incubated 14 h (standing) before transferring 10 μL of the subsequent culture to 190 μL fresh medium in microtiter plates. After an additional 25 h incubation (standing), the optical density at 595 nm (OD_{595}) was determined for each culture using a Biorad model 3550 microplate reader (Figure 7-27). Twenty four selected clones (typically, those with the greatest OD_{595} values; see Figure 7-27) were cultured by serial overnight culture (§2.1.1, 30 mL). S.s. DNA extracted (§2.3.2.b) from virions harvested by two-stage PEG precipitation (§2.6.4) of culture supernatants was used as sequencing template (§2.4) to determine that all of the selected clones were (unexpectedly) of a novel class of pseudorevertant (Figure 7-28).

2.12.2. Construction of FHA series 80 libraries (Table 7-VI)

(a) fDRW8nn propagation. fDRW836, fDRW861, fDRW863, fDRW864 and fDRW867 (Figure 7-29) extracted from cultures used to prepare sequencing template for these vectors (§2.8.7) was digested to apparent completion with *Sma*I and treated with Mung Bean Nuclease (NEB; 1 U per μg DNA, 30°C, 30 min.) to eliminate s.s. DNA, before recovering DNA by phenol/chloroform extraction and ethanol precipitation.

(b) Two-stage ligations. Four libraries (80-A to -D) were constructed, each with a different size-fractionated set of DNase I-generated fragments of the 10 kbp *B. pertussis fhaB* *Eco*RI restriction fragment (§2.12.1.b). For each library, three ligations were performed, each with 21 fmol of an equimolar mixture of five fDRW8*nn* vectors (§2.12.2.a) and 63 fmol, 210 fmol or 1.1 pmol (roughly 3:1, 10:1 and 50:1 molar insert:vector ratios) of *B. pertussis fhaB* fragments. Two control ligations (80-X and 80-Y) were also performed, each with 21 fmol of fDRW8*nn* RF DNA alone (*i.e.*, no *fhaB* fragments were included). After ligating (3 U T4 ligase, §2.2.2.b) each vector/insert mixture or control 12 h at 16.5°-19°C, and with the goal of diluting each mixture to favour circularization of ligation products, an additional 3 U T4 ligase was added and the volume adjusted to 150 μL. After an additional 14 h incubation, ligation products for the three insert:vector ratios were pooled, yielding a single pool for each of libraries 80-A, -B, -C and -D.

(c) SrfI digestion of ligation products. To reduce the numbers of non-recombinants, ligation products for each library (80-A, ..., 80-Y) were incubated 10 min. at 65°C to inactivate T4 ligase, precipitated with isopropanol (with 5 μg tRNA as carrier) and digested (except library 80-X) 1.5 h with 4 U *Srf*I before incubating 20 min. at 65°C to inactivate *Srf*I, ethanol precipitating and resuspending in 7 μL water.

(d) Electroporation and harvesting of virions. Two electroporations were carried out for each library (80-A, ..., 80-Y). In each, one-half (3.5 μL) of the ligation products were

electroporated (§2.5.2) into *E. coli* MC1061. For each ligation, the number of transformants was estimated by counting colonies on a single LB-Tet plate spread with a single 10 μL sample; values are reported in Table 7-VI. To confirm that these estimates were reasonable, 16-32 cm^2 areas of plates (245 mm x 245 mm NUNC culture dishes) spread with ~2 mL (90%) of transformants were also counted. For each library, virions were harvested by washing transformants from plates (§2.6.1) and subsequent two-stage PEG precipitation (§2.6.4) of wash supernatant; yields (as physical particles) were estimated by UV spectroscopy (§2.7.1, Table 7-VI).

2.13. ANTIBODIES AND IMMUNOLOGICAL REAGENTS

2.13.1. Rabbit pAbs produced against wild-type phage f1

Rabbit α-f1 polyclonal antibodies were obtained from New Zealand White rabbits. Each of two rabbits was immunized with 0.1 mg (as determined by a BCA protein assay, similar to that described later, §2.16.1.2) CsCl-purified (§2.6.5) f1 virions delivered with RIBI adjuvant via intraperitoneal, intramuscular, intradermal and subcutaneous routes, following the adjuvant manufacturer's recommended protocol; this immunization was repeated four weeks later. After an additional two weeks, rabbits were sacrificed and antibodies purified as described below.

After centrifuging (3000 x g_{max}, 4°C, 30 min.) 11 mL serum, selected serum proteins were precipitated from the resulting supernatant by slowly adding with stirring 0.5 volume saturated $(NH_4)_2SO_4$ (76.1% w/w in H_2O) and storing the resulting mixture >24 h at 4°C. After centrifugation as before to remove the precipitate, antibodies within supernatant were precipitated by adding 1.0 volume saturated $(NH_4)_2SO_4$ and storing the resulting mixture overnight at 4°C. After centrifugation as before, the antibody-containing precipitate was resuspended in 4 mL PBS; this was dialysed against PBS (1 L, 4°C) three times over a 24 hour period. In each of two preparations, half of the dialysed sample was purified by Protein A affinity chromatography using a prepacked column (Pierce Immunochemicals, binding capacity

= 34 mg IgG) following the manufacturer's recommended procedures for antibody purification and column regeneration. Recovered antibodies were concentrated by precipitation with saturated $(NH_4)_2SO_4$. After centrifugation, the antibody-containing precipitate was resuspended in 4.2 mL PBS; this was dialysed three times against PBS (1.5 L, 4°C) before adding NaN_3 to a final concentration of 0.02% and storing (as 50 µL and 250 µL aliquots) at -20°C. This final preparation (~2.2-fold less volume than the starting serum) contained (estimated by A280; Harlow & Lane 1988) 4.7 mg IgG mL^{-1}.

2.13.2. Mouse mAbs produced against *P. falciparum* CSP

Pf4C11.6 (IgG2a), Pf1G3.4 (IgG1), Pf5C1.1 (IgG1), Pf2A10 (IgG2a), Pf5G5.3 (IgG1), Pf1B2.2 (IgG1), P54A4.1 (IgG3), Pf2F1.1 (IgM), PvNSV3 (IgG) and Pb4B10 (IgG), made and screened against *P. falciparum* (Pf), *P. vivax* (Pv) and *P. berghei* (Pb) sporozoites, recognize CS protein (Wirtz *et al.* 1987; Burkot *et al.* 1991; Wirtz *et al.* 1991; R. A. Wirtz, unpublished data) and were provided by R. A. Wirtz, Walter Reed Army Institute of Research (WRAIR). Pf2F1.1 was used as unpurified ascitic fluid; other mAbs had been purified by Protein A affinity chromatography. Pf2A10 and Pf1B2.2 were biotinylated by Kirkegaard & Perry Laboratories, Inc., Gaithersburg, MD.

(a) Confirmation and determination of mAb isotypes. For each of affinity-purified Pf2A10, Pf1B2.2, Pf4C11.6, Pf5G5.3, Pf1G3.4, Pf5C1.1 and Pf5A4.1 and for Pf2F1.1 ascitic fluid, ten wells of an Immulon-1 microtiter plate were coated with 100 ng affinity-purified mAb or 0.5 µL Pf2F1.1 ascitic fluid (in PBS, 100 µL $well^{-1}$). After 1 h incubation at 37°C, plates were washed three times with wash buffer (PBS + 0.05% Tween 20). Each α-CSP antibody was probed with each of six goat α-mouse isotype mAbs (α-IgG_1, α-IgG_{2A}, α-IgG_{2B}, α-IgG_3, α-IgM and α-IgA; Sigma isotyping kit ISO-2) by adding 100 µL of a 1:3,000 dilution (in wash buffer) of each goat α-mouse mAb to duplicate wells. After 0.5 h incubation at room temperature, plates were washed three times before addition of peroxidase-conjugated mouse

α-goat IgG 2° antibodies (1:5,000 dilution in wash buffer, 100 μL $well^{-1}$, §2.13.4). After 15 min., plates were washed four times before developing with substrate (OPD, §2.13.5) and reading colored reaction products as outlined later (§2.16.2.a). Wells with the greatest A_{490} values were used to confirm (see references cited in preceding section) or determine that isotypes are: Pf2A10, IgG_{2A}; Pf1B2.2, IgG_3; Pf2F1.1, IgM; Pf4C11.6, IgG_{2A}; Pf5G5.3, IgG_1; Pf1G3.4, IgG_1; Pf5C1.1, IgG_1; Pf5A4.1, IgG_3.

(b) Quantification of Pf2F1.1. The concentration of IgM in Pf2F1.1 ascitic fluid was estimated by ELISA using purified IgM (Sigma M-5170) as a standard: 5-fold serial dilutions of Pf2F1.1 and the IgM standard were coated onto wells of Immulon 3 plates and probed with peroxidase-conjugated goat α-mouse IgM mAbs (Sigma) using OPD (§2.13.5) as substrate in a manner similar to that described later (§2.16.2.1).

2.13.3. Rabbit pAbs produced against *B. pertussis* FHA

Three rabbit α-FHA sera were a gift from A. Siebers. These had been prepared in New Zealand White rabbits using two preparations of FHA purified by heparin-sepharose affinity chromatography (Menozzi *et al.* 1991). One such preparation, of so-called "native" FHA, had been eluted from the heparin-sepharose column using a NaCl gradient according to the published (Menozzi *et al.* 1991) procedure. A second preparation, of so-called "SDS-denatured" FHA, could be removed from the heparin-sepharose column only by use of sodium dodecyl sulfate (A. S. and B. Finlay, unpublished). Sera from two rabbits immunized with "native" FHA are identified as "FN1/4" and "FN2/4". Serum from a third rabbit immunized with "SDS-denatured" FHA is identified as "FS1/4". Following a recommendation by A.S. (based on the sera's recognition of *E. coli* expressing *B. pertussis fha*B-derived fragments using a *Pseudomonas aeruginosa* OprF expression system), sera FN2/4 and FS1/4 were employed for most work described here.

(a) α-FHA antibody purification and biotinylation for biopanning and plaque lifts. Antibodies employed for biopanning FHA-70 and -80 libraries and subsequent plaque lifts were purified as follows. For each of FN2/4 and FS1/4, ~950 μL serum was centrifuged 30 min. at 3000 x g_{max} in an Eppendorf 5415C microfuge; selected serum proteins were precipitated by diluting the resulting supernatant to 3.8 mL in PBS + 0.02% NaN_3 and slowly adding, with stirring, 1.9 mL saturated $(NH_4)_2SO_4$ and storing the resulting mixture overnight at 4°C. After centrifugation to remove the precipitate, antibodies were precipitated from supernatant by adding an additional 1.9 mL saturated $(NH_4)_2SO_4$ and storing the resulting mixture overnight at 4°C. After centrifugation, the antibody-containing precipitate was resuspended in 750 μL PBS and dialysed against PBS (400 mL, 4°C) three times over a ~24 h period.

Each dialysed sample was absorbed against *E. coli* antigens using a Pierce Immunochemicals immobilized *E. coli* lysate kit in accordance with the manufacturer's recommended procedures. One-half of each *E. coli*-absorbed serum was purified by Protein A affinity chromatography using a prepacked column (Pierce Immunochemicals, binding capacity = 34 mg IgG) following the manufacturer's recommended procedures. The remaining volume of each *E. coli*-absorbed serum and the entire volume of each Protein A-purified pAbs were separately precipitated with 1.0 volume $(NH_4)_2SO_4$ and resuspended in 0.5 mL PBS before washing each sample by centrifuging three times at 3,000-4,350 x g_{max} (4°C) in a Centricon 50 microconcentrator, with 1 mL and 2 mL volumes of PBS being applied after the first and second centrifugations respectively. For each Centricon 50-treated Protein A-purified serum, ~2 mg IgG (as determined by A280; Harlow & Lane 1988) was biotinylated using a Pierce Immunochemicals Sulfo-NHS-biotinylation kit, as follows. After ~2 mg of antibody in 1 mL PBS was combined with 20 μL (0.4 mg) Sulfo-NHS-biotin and incubated 2 h on ice, excess Sulfo-NHS-biotin was removed by repeated washings in a Centricon 50 microconcentrator. Avidin and HABA (2-4'-hydroxyazobenzene benzoic acid) were used in accordance with the

manufacturer's recommended procedures to determine that the molar ratios of "surface-exposed" or "available" (*viz.*, to streptavidin) biotin to antibody were 2.3:1 and 1.8:1 for FN2/4 and FS1/4 respectively. Each of the (i) *E. coli*-absorbed sera, (ii) protein A-purified and (iii) biotinylated IgG preparations were employed in later experiments. Volume changes have been considered in reporting the dilutions employed for each serum or antibody; *i.e.*, each dilution is reported in terms of an effective dilution of crude serum. The effects of the various stages of purification on the abilities of FN2/4 and FS1/4 to recognize FHA were determined (Figure 2-8) by applying serial dilutions (3.1 to 100 ng in 2 μL PBS) of heparin-sepharose affinity-purified FHA and, as a control, recombinant CSP protein PfR32tet32, to nitrocellulose and probing these samples in a manner similar to that described later (§2.18, Immunoblots), with 1:2,000 and 1:10,000 effective dilutions of sera or antibodies from each level of purification.

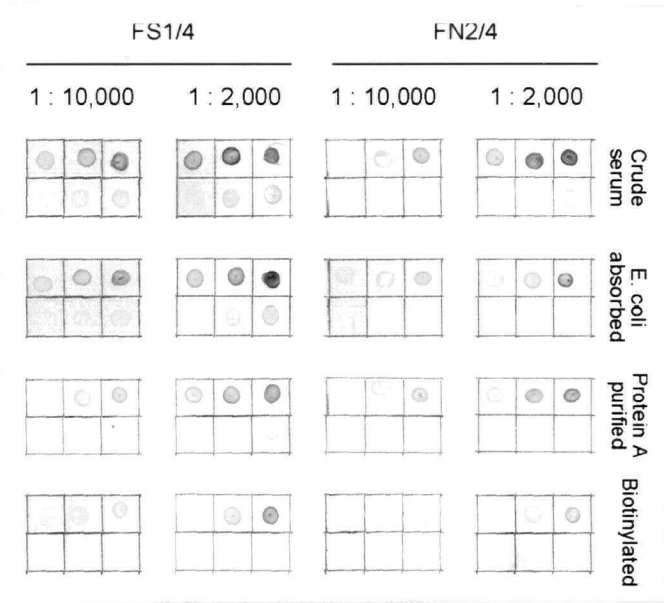


Fig. 2-8. Effect of α-FHA sera purification and biotinylation on recognition of FHA. Samples (~2 μL) of heparin-sepharose affinity-purified FHA (and recombinant CSP as controls; not shown) were applied to gridded (5 mm x 5 mm) nitrocellulose in the pattern

FHA (ng)	
100	12.5
50	6.3
25	3.1

and probed with FN2/4 and FS1/4 α-FHA crude and purified sera and antibodies at the indicated dilutions, using a method similar to that described for FHA controls in §2.15.2.b (see also Figure 2-11).

(b) α-FHA antibody purification for ELISA and immunoblots. FN2/4 and FS1/4 (crude) sera were absorbed against *E. coli* antigens (by means of a Pierce Immunochemicals immobilized *E. coli* lysate kit used as before) for use in ELISA and immunoblots. Volume changes have been considered in reporting the dilutions employed for these sera; *i.e.*, each

dilution is reported in terms of an effective dilution of crude serum. FS2/4 serum was not pre-absorbed against *E. coli* antigens.

2.13.4. Secondary antibodies

Peroxidase-conjugated goat α-mouse IgG or goat α-rabbit IgG polyclonal antibodies (Jackson ImmunoResearch or GIBCO/BRL) were employed as 2° antibodies for ELISAs in which virions coated onto wells of Immulon plates were probed with mouse mAbs or rabbit pAbs. For other ELISAs in which cross-reactivity of the 2° antibody with antigen (components of mouse ascitic fluid, §2.13.2.a) or capture antibody (mouse mAbs, §2.17.1, §2.17.2) was of concern, peroxidase-conjugated mouse α-goat IgG (§2.13.2.a) or α-rabbit IgG (§2.17.1, §2.17.2) pAbs that had been absorbed against serum proteins of mouse and other species; (Jackson ImmunoResearch) were employed; these 2° antibodies were also employed as a matter of convenience in certain conventional ELISAs. Alkaline phosphatase-conjugated goat α-rabbit IgG pAbs (GIBCO/BRL) were employed in plaque lifts and immunoblots.

2.13.5. Immunochemicals, recombinant CS protein and related material

OPD (Sigma, *o*-phenylenediamine, 1 mg mL^{-1} in 0.1 M citrate pH 4.5 with 0.012% H_2O_2) was employed as substrate for peroxidase-conjugated 2° antibodies (ELISA). BCIP/NBT (5-bromo-4-chloro-3-indolyl phosphate *p*-toluidine salt, 50 μg mL^{-1}; nitroblue tetrazolium chloride, 100 μg mL^{-1}, in substrate buffer: 100 mM Tris pH 9.6, 40 mM $MgCl_2$) was employed as substrate for alkaline phosphatase-conjugated 2° antibodies (plaque lifts and immunoblots). Nitrocellulose discs employed in plaque lifts and immunoblots were purchased from Schleicher and Schleicher.

PfR32tet32, Pbtet32 and PvNS1v20, recombinant CS proteins of *P. falciparum*, *P. berghei* and *P. vivax* respectively, provided by R. A. Wirtz (WRAIR), have been previously described (Young *et al.* 1985; Egan *et al.* 1987; Gordon *et al.* 1990). Bovine serum albumin (BSA) was purchased from Sigma (as fraction V) or from Intergen (as "bovuminar standard powder").

2.14. BIOPANNING (AFFINITY SELECTION OF TARGET CLONES)

2.14.1. Biopanning random peptide libraries with α-CSP mAbs

Sixteen sets of biopans (affinity purifications of target clones from libraries) were carried out (Chapter 5, Figure 5-1). Thus, each of two random peptide libraries (§2.1) was biopanned with four quantities (1.2 μg ≈ 8 pmol, 120 ng ≈ 0.8 pmol, 12 ng ≈ 80 fmol and 1.2 ng≈ 8 fmol) of each of two α-CSP mAbs, Pf2A10 and Pf1B2.2.

(a) First round of biopanning. For each biopan, the indicated quantity of biotinylated mAb was combined with 2.6 x 10^{10} virions (physical particles, "biopan input") of the 6-mer library or 2.5 x 10^{11} virions (physical particles) of the 15-mer library in a final volume of 80 μL PBS with 1% extensively dialysed BSA (d-BSA = 5 g BSA mL^{-1} in 25 mL PBS dialysed three times over two days against 4 L PBS). Samples (in 500 μL Eppendorf tubes) were incubated ~20-24 h at 4°C on a slowly rotating culture tube carrier. After blocking (1 h, 37°C, 0.1% d-BSA in PBS, 200 μL $well^{-1}$) microtiter plate wells to which streptavidin (SA) had been covalently bound (Pierce Reacti-bind SA-coated polystyrene microtiter plates, 400 ng ≈ 7 pmol SA per well, 28 pmol biotin binding capacity per well) and subsequently washing each well 3 times with wash buffer (PBS + 0.5% Tween 20, 200 μL $well^{-1}$), virion/antibody mixtures were added to wells (one sample per well) and incubated 15 min. at room temperature with gentle motion on a platform shaker. For each biopan sample, 50 μL of mixture ("unbound virions") was removed and stored for later use as a control. After wells were washed 10 times with wash buffer (200 μL $well^{-1}$), 200 μL 0.1 N HCl (adjusted to pH 2.2 with glycine) was added to each well to elute bound virions. After 15 min. elution with gentle motion on a platform shaker, 12 μL 2 M Tris base (unadjusted pH) was added to each well. Each sample of eluted and pH-neutralized virions was recovered immediately and combined with 1.5 mL E. coli K91-Kan (OD_{600} = 0.3). After incubating 1 h (37°C, standing, in LB + 0.2 μg tetracycline mL^{-1}), each virion/cell mixture was transferred to 30 mL LB + 20 μg

tetracycline mL^{-1} and incubated 23 hours (shaking) before virions were harvested (as "amplified biopan output") from culture supernatant by two-stage PEG precipitation (§2.6.4). For use as controls, four 10 μL aliquots of "unbound virions" for each antibody/library combination were pooled and used to infect (§2.5.3) ~2 mL *E. coli* K91-Kan. After infected cells were transferred to 100 mL LB + 20 μg tetracycline mL^{-1} and incubated overnight (shaking), virions were harvested from culture supernatants by two-stage PEG precipitation (§2.6.4).

(b) Subsequent rounds. The second and third rounds of biopanning were carried out in the same manner as the first round, except that 2.5 x 10^{11} virions (physical particles, as quantified by UV spectroscopy, §2.7.1) of the 1 x 10^{14} - 3 x 10^{14} virions harvested from the previous round as "amplified biopan output" were employed as "biopan input".

(c) Selection and propagation of selected clones. For each of the four biopans that employed 120 ng mAb, approximately 1-2 x 10^9 virions (physical particles; volume = 1 μL) of amplified output from the third biopan were used to infect (§2.5.3) 1 mL *E. coli* K91-Kan, and infected cells were spread on LB-Tet plates and incubated overnight. After streaking on LB-Tet to obtain well-isolated colonies, 24 clones were cultured by serial overnight culture (§2.1.1, 30-50 mL) and virions were harvested from culture supernatants by two-stage PEG precipitation (§2.6.4). Virion yields were estimated by UV spectroscopy (§2.7.1).

2.14.2. Biopanning FHA-70 and -80 libraries (Figure 7-30)

For each of eleven FHA peptide and "control" libraries (Table 2-V), four biopans were carried out with 5 x 10^8 to 7 x 10^{11} virions and 3.6 μg (≈24 pmol), 360 ng (≈2.4 pmol), 36 ng (≈240 fmol) and 3.6 ng (≈24 fmol) of an equimolar pool of biotinylated protein-A purified rabbit α-FHA sera FN2/4 and FS1/4 (§2.13.3.a). For each biopan, the indicated quantities of virions and antibodies were combined in a final volume of 100 μL PBS with 1% extensively dialysed BSA (§2.14.1.a) and incubated ~16 h (4°C, in 500 μL Eppendorf tubes) on a slowly rotating tube carrier. After blocking (2 h, 37°C, 1.0% d-BSA in PBS, 200 μL $well^{-1}$)

streptavidin-coated microtiter plate wells (28 pmol biotin binding capacity per well, §2.14.1.a), blocking solution was discarded and virion/antibody mixtures were added to wells (one sample per well) and incubated 20 min. at room temperature with gentle motion on a platform shaker. After wells were washed 10 times with wash buffer (PBS + 0.5% Tween 20, 200 μL well, ~5 min agitation on a platform shaker for each wash), 150 μL 0.1 N HCl (adjusted to pH 2.2 with glycine) was added to each well to elute bound virions. After 20 min. elution on a platform shaker, 9 μL 2 M Tris base (unadjusted pH) was added to each well and each sample of eluted and pH-neutralized virions recovered immediately and stored (4°C) in microtiter plates that had been previously blocked with 1% BSA in PBS (200 μL well^{-1}). The numbers of virions recovered for each biopan (Figure 7-30) were estimated by a TU "spot" assay (§2.7.2.c) using single 5 μL samples of 10-fold serial dilutions of eluted virions.

Table 2-V. Virions employed in biopanning FHA-70 and -80 libraries.

Library	Insert size (base pairs)	Virions employed, log (physical particles)
70-X	no insert	8.8
70-A	30-75	8.8
70-B	75-150	8.8
70-C	150-300	8.7
70-D	300-600	9.2
80-X	no insert	11.6
80-Y	no insert	10.9
80-A	30-75	11.4
80-B	75-150	11.4
80-C	150-300	11.1
80-D	300-600	11.8

2.15. PLAQUE LIFTS AND RELATED METHODS EMPLOYED TO IDENTIFY ANTIBODY-REACTIVE CLONES IN FHA-70 AND -80 LIBRARIES

2.15.1. Methodology issues

Two methods of using plaque lifts to identify antibody-reactive clones within biopan eluates of FHA-70 and -80 libraries were explored using wild-type phage f1 and α-f1 pAbs. For each method, two samples of f1 virions sufficient to yield 200-400 plaques were plaqued on *E. coli* K91-Kan in a manner similar to full-plate plaque assays (§2.7.2.a). The two resulting lawns (A and B) were treated differently with respect to application of nitrocellulose discs (§2.13.5). In method A (Figure 2-9A) a nitrocellulose disc was laid down on lawn A after 1 h incubation (37°C; as in a protocol for plaque lifts of λgt11 expression libraries) and the lawn with applied nitrocellulose was incubated overnight (37°C). In method B (Figure 2-9B), lawn B was incubated overnight (37°C) before application of a nitrocellulose disc with slight pressure applied by a glass spreading rod.

Following overnight incubation, both lawns with applied nitrocellulose were incubated 10 min. at 4°C (to allow nitrocellulose to be removed without lifting the [less fluid] top agar) before removing and placing each nitrocellulose disc in a Petri dish. After being washed three times with gentle shaking in 10 mL blocking buffer (5% skim milk powder, 0.1% Tween 20 in PBS; 30 min wash^{-1}), each disc was cut in two parts and each part incubated 1 h at room temperature with gentle shaking in 10 mL of a 1:5,000 or 1:50,000 dilution of Protein A-purified α-f1 (1°) pAbs (§2.13.1) in blocking buffer. After washing three times with blocking buffer (10 mL wash^{-1}, 10 min. wash^{-1}), each piece was incubated 1 h at room temperature with gentle shaking in 10 mL alkaline phosphatase-conjugated 2° antibodies (1:3,000 dilution in blocking buffer; §2.13.4). After three washes (10 mL wash^{-1}, 10 min. wash^{-1}) in blocking buffer and one in substrate buffer (§2.13.5), 8 mL BCIP/NBT substrate (§2.13.5) was added

to each disc, and reactions were allowed to proceed 2.5-8.5 min. before being stopped by rinsing the nitrocellulose pieces in water. Based on the puzzling and unsatisfactory results for method A and the conversely satisfactory results for method B (Figure 2-9), method B was chosen for subsequent plaque lifts.

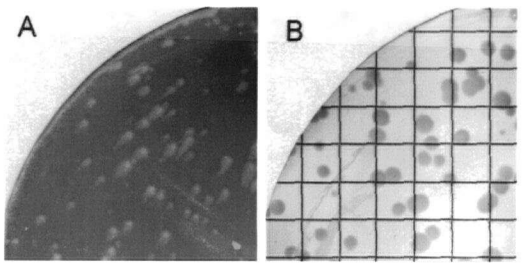


Fig. 2-9. Alternative methods of applying nitrocellulose to lawns. Plaque lifts of f1 virions probed with α-f1 pAbs (§2.15.1). In method A, a nitrocellulose disc was applied after 1 h incubation of the phage/bacteria lawn and the disc/lawn subsequently incubated overnight. In method B, the disc was applied after overnight incubation of the lawn. The resulting plaque lifts were probed with 1:5,000 (illustrated here) and 1:50,000 (not shown; results were similar) dilutions of α-f1 pAbs.

Alternative blocking and washing conditions were explored using CSP-library clones (§2.10) and α-CSP mAb Pf2A10. Two-μL samples (five-fold serial dilutions from 1 μg to 8 ng protein) of CSP-library clones displaying the peptides NANPNAN (recognized by Pf2A10) and NANANAN (not recognized), NPNANPN (recognized) and recombinant CSP proteins R32tet32 (recognized) and Pbtet32 (not recognized) were applied to nitrocellulose and processed in a manner similar to that described in the preceding section (§2.15.1), except that four combinations of blocking and washing methods were explored. Method A employed 5% skim milk powder + 0.1% Tween 20 in PBS for all washes and antibody dilutions. Method B employed 3% BSA, 1% skim milk powder for the first wash and antibody dilutions, and 0.1% Tween 20 in PBS for remaining washes. Methods C and D were similar to method B, except skim milk powder (Method C) or Tween 20 (Method D) were omitted.

Based on minimal background staining of nitrocellulose and appropriate signals for CSP-library clones and recombinant CSP, method B was chosen for subsequent plaque lifts. Results of a followup experiment using method B to probe plaque lifts of CSP-library clones are shown in Figure 2-10.

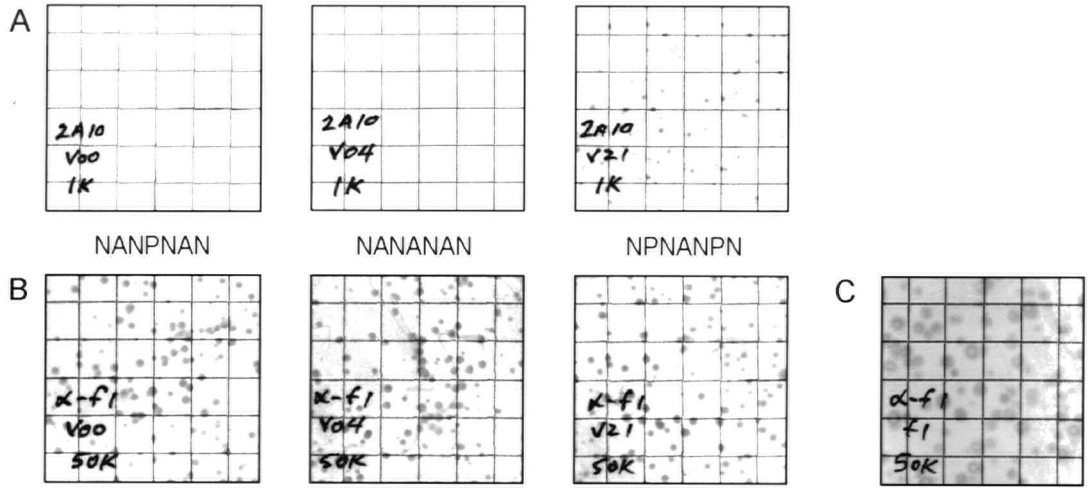


Fig. 2-10. Plaque lifts of CSP-library and wild-type f1 virions. **A**. Plaque lifts -- performed using squares of gridded (5 mm x 5 mm) nitrocellulose -- of CSP-library clones probed with α-CSP mAb Pf2A10. Plaque lifts of NANPNAN and NANANAN were expected to be negative; the lift of NPNANPN was expected to be positive. **B**. Concurrent plaque lifts, from the same lawns as in **A**, probed with α-f1 pAbs to demonstrate that plaques existed and were successfully transferred to nitrocellulose. **C**. Wild-type f1 virions probed with α-f1 pAbs illustrates the small size of plaques of recombinant fd-tet derivatives (**B**) compared to those of wild-type Ff phage (**C**).

2.15.2. Identification of antibody-reactive clones in FHA-70 libraries and assessment of the effect of library amplification on the fraction of antibody-reactive clones recovered

(a) Library amplification. For each of libraries 70-A, -B, -C and -D that had been biopanned with 3.6 μg and 360 ng of pooled FN2/4 and FS1/4 α-FHA pAbs (§2.14.2), 40 μL biopan eluate ($<10^5$ TU) was combined with 1 mL *E. coli* K91-Kan (OD_{600}=0.2, thus, ~2 x 10^8 cells) and the resulting mixture incubated 0.5 h at 37°C before plating on LB-Tet. After 18 h incubation (37°C), virions were harvested by washing bacterial growth (2 x 10^2 to 3 x 10^3 cfu per eluate) from plates (§2.6.1) and subsequent centrifugation.

(b) Plaque lifts of unamplified and amplified FHA-70 eluates (Figure 7-32). For each of libraries 70-A, -B, -C and -D, quantities of unamplified eluates (from biopanning with 3.6 μg and 360 ng pooled FN2/4 and FS1/4 α-FHA pAbs, §2.14.2) and amplified eluates (immediately preceding section) sufficient to yield ≤ 2 x 10^3 plaques on each of 4 lawns were

plaqued on *E. coli* K91-Kan in a manner similar to full-plate plaque assays (§2.7.2.a). After overnight incubation to allow plaque development, plates were equilibriated to room temperature before applying a nitrocellulose disc (§2.13.5) to each lawn with light pressure applied by a glass spreading rod. After 40 min. incubation at 4°C (in some later experiments

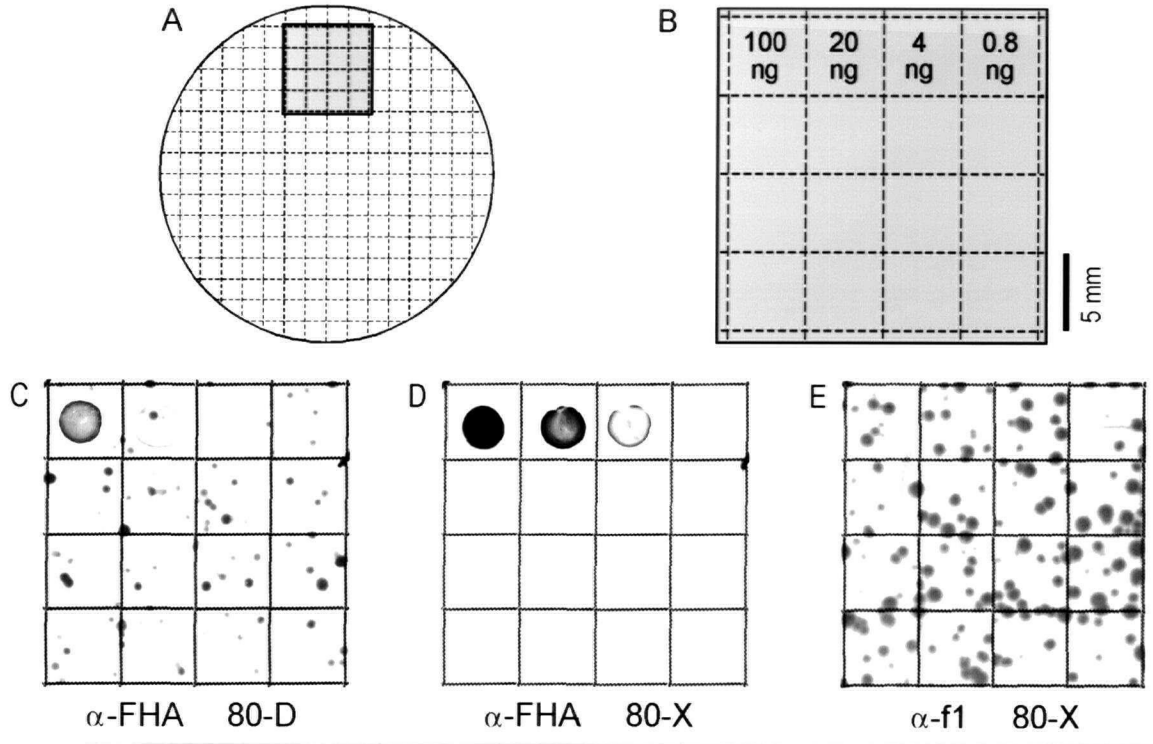


Fig. 2-11. Plaque lifts of biopanned FHA-80 library clones. **A.** Disc of gridded (5 mm x 5 mm) nitrocellulose identifying the location of the four-by-four matrices shown in B to D. **B.** Initial plaque lifts employed to identify antibody-reactive clones included 100, 20, 4 and (optionally) 0.8 ng samples of heparin-sepharose affinity-purified FHA as positive controls. **C** and **D.** Plaque lifts from early trials in which relatively low dilutions (1:4,000) of each of FN2/4 and FS1/4 (as a pool) were employed as 1° antibodies; later trials included higher dilutions (1:8,000 each) of pooled antibodies and 1:8,000-1:32,000 dilutions of FN2/4 alone. **C** and **D** are derived, respectively, from the "large fragment" (300-600 bp inserts) library 80-D and the control (no insert) library 80-X. The relatively greater intensity of staining of FHA controls in **D** compared to **C** reflects the greater color development time (BCIP/NBT substrate with alkaline phosphatase-conjugated 2° antibodies) of the 80-X *versus* 80-D plaque lifts. **E.** Plaque lift of control library 80-X probed with rabbit α-f1 pAbs. This concurrent control for the corresponding plaque lift shown in D, intended to show that plaques existed and were transferred to nitrocellulose, was included only in early experiments.

this time was arbitrarily increased to up to 1.5 h), nitrocellulose discs were removed and dried (room temperature, 30 min.) before applying 2 μL samples (100, 20, 4 ng) of heparin-sepharose affinity-purified FHA as positive controls (Figure 2-11) and allowing these to dry

(room temperature, 20 minutes). Each disc was incubated 1 h at room temperature with gentle agitation in 10 mL blocking buffer (1% skim milk powder; 3% BSA, reduced to 2% BSA in some later experiments) before storing overnight (4°C) in blocking buffer (overnight storage was for convenience and not always repeated in later experiments). After equilibriation (20 min., with gentle agitation) to room temperature, each disc was washed three times (per wash: 10 mL PBS + 0.05% Tween 20, 15 min. with gentle agitation) before addition of (i) 1:8,000 dilutions (in blocking buffer, 8 mL per disc) of each of *E. coli*-absorbed protein A-purified FN2/4 and FS1/4 pAbs (§2.13.3.a), and (ii) 1:8,000, 1:32,000 and 1:128,000 dilutions of FN2/4 alone. After 1 h incubation (room temperature) with gentle agitation, each disc was washed three times before addition of 10 mL alkaline phosphatase-conjugated 2° antibodies (1:3,000 dilution in blocking buffer; §2.13.4). After 1 h incubation with gentle agitation, each disc was washed three times with wash buffer and once (15 min. with gentle agitation) with 10 mL substrate buffer (§2.13.5) before addition of 8 mL BCIP/NBT substrate (§2.13.5). Reactions were allowed to proceed until (after 22-96 min.) adequate signal had developed for an apparently significant fraction of plaques; reactions were stopped by rinsing the nitrocellulose discs in water. Each *set* of amplified and unamplified samples was developed for the same period of time.

Plaques visible (even faintly so) on nitrocellulose were counted as "antibody-reactive"; plaques visible on the corresponding bacterial lawn were counted (or estimated in some cases, by counting plaques on one-quarter of the lawn) as "total" plaques (Figure 7-32). Average color +intensity of antibody-reactive plaques varied considerably from library to library, such that plaques of the "larger-fragment" libraries 70-C and 70-D developed more quickly and to more intense color than did those of the "smaller-fragment" libraries 70-A and 70-B; indeed, plaques of the latter libraries were commonly faint and difficult to discern, even with a hand-held magnifying lens. Example plaque lifts from early trials with library 80-D are shown in

Figure 2-11.

(c) Selection and plaque purification of antibody-reactive FHA-70 clones. As summarized in Table 6-II (Chapter 6), 56 antibody-reactive clones were selected from FHA-70 libraries for further analysis. These were chosen from those that reacted most strongly with antibody (as judged by plaque lift color intensity at a given antibody dilution, or by any visible reactivity at a high antibody dilution).

For each selected clone, a ~2 mm x 2 mm plaque-containing agar plug (identified by the position of the target plaque on gridded nitrocellulose and corresponding grid marks left on the lawn by applying nitrocellulose with light pressure) was excised and transferred to 1 mL PBS. After overnight incubation (4°C) to allow elution of virions, 5-fold serial dilutions of each virion eluate were plaqued on *E. coli* K91-Kan in a manner similar to full-plate plaque assays (§2.7.2.a). After overnight incubation (37°C) to allow plaque development, plaque lifts of sections of the resulting lawns were performed essentially as described above (§2.15.2.b), except (i) all plaque lifts were probed with a 1:8,000 dilution of Protein A-purified FN2/4 pAbs, and (ii) in many cases, plaque lifts were performed with one or several 2 to 2.5 cm square pieces of nitrocellulose (*versus* a complete disc). This procedure was repeated once or twice in essentially the same manner (exceptions are noted below) until an apparently clonal population was isolated for each of 51 of the originally selected 56 clones, *viz.*, until the isolation of a single plaque that, after excision and elution, gave rise to plaque lifts in which all plaques were antibody-reactive. During one round of plaque purification, plaque lifts for seven weakly-reactive clones were performed using a 1:4,000 dilution of FN2/4, and those for ten other weakly-reactive clones were performed using an equimolar pool of 1:4,000 dilutions of each of FN2/4 and FS1/4. At some time during the repeated rounds of plaque purification, reactivity with antibody was "lost" for five of the originally selected 56 clones. Conceivably, this may reflect errors in selecting plaques or recombination within *B. pertussis fha*B-derived

inserts; these five were not considered in subsequent analyses. As a final clonal isolation step, 20 μL samples of 5-fold serial dilutions (5^{-1} to 5^{-6}) of each apparently monoclonal virion eluate were used to infect (§2.5.3) 80 μL *E. coli* K91-Kan, and infected cells were spread on LB-Tet plates. After overnight incubation, a single colony derived from each clone was transferred as a short streak to an LB-Tet plate. After overnight incubation these "FHA-70 master stock plates" were stored at 4°C and used as inocula for subsequent virion propagation.

(d) Virion propagation. For each of the 51 clones isolated as described above, colonies on "FHA-70 master stock plates" were used as a source of inocula for serial overnight cultures (§2.1.1, 50 mL). Virions were harvested from ~32 mL of each resulting culture by two-stage PEG precipitation (§2.6.4) and quantified by UV spectroscopy (§2.7.1). These virions were used (i) as a source of template (§2.3.2.b) for sequencing (§2.4) *fha*B-derived inserts, (ii) in ELISA (§2.16.4.2) and (iii) as a source of virions used to infect *E. coli* K91-Kan in later large-scale preparations of selected clones (§2.15.4).

2.15.3. Identification of antibody-reactive clones in FHA library 80-A

(a) Identification and selection of clones. Quantities of FHA library 80-A biopan eluates (derived from biopanning this library with 3.6 μg and 360 ng pooled FN1/4 and FS2/4 α-FHA pAbs, §2.14.2) sufficient to yield ~200 plaques on each of six lawns were plaqued on *E. coli* K91-Kan in a manner similar to that described for full-plate plaque assays (§2.7.2.a). After overnight incubation (37°C) plaque lifts of these lawns were performed essentially as described earlier (§2.15.2.b) except that for each eluate, duplicate plaque lifts were probed with 1:8,000, 1:32,000 and 1:128,000 dilutions of protein A-purified FN2/4 pAbs (§2.13.3.a). Positive plaques and total plaques (Chapter 6, Table 6-III) were counted as described earlier.

(b) Clonal isolation by limiting dilution transduction. For each of 58 clones selected from among those identified with a 1:32,000 dilution of FN2/4 (Chapter 6, Table 6-III), a ~2 mm x 2 mm plaque-containing agar plug was excised, transferred to 0.5 mL LB and stored

overnight (4°C) to allow elution of virions. For each clone, 10 µL samples of 10-fold serial dilutions (10^{-1} to 10^{-12}) of these virion eluates were transferred to 90 µL *E. coli* K91-Kan (OD_{600}=0.2) in microtiter plate wells. After 1 h incubation (37°C, in LB + 0.2 µg tetracycline mL^{-1}), 10 µL of each well was transferred to 90 µL LB-Tet in microtiter plate wells and, after overnight incubation (standing), the OD_{595} of each culture was read to determine the highest dilution of each clone showing obvious host cell growth. A single sample of each such dilution was streaked on LB-Tet and, after overnight incubation, a single colony was selected and employed in virion propagation, as below.

(c) Virion propagation. Colonies selected as described immediately above were used as a source of inocula for serial overnight cultures (§2.1.1, 40 mL). Virions harvested from ~32 mL of each culture supernatant by two-stage PEG precipitation (§2.6.4) and quantified by UV spectroscopy (§2.7.1) were used (i) as a source of s.s. template (§2.3.2.b) for sequencing (§2.4) *B. pertussis fha*B-derived inserts, (ii) in ELISA (§2.16.4.2)and (iii) as a source of virions used to infect *E. coli* K91-Kan in later large-scale preparations of selected clones (§2.15.4).

2.15.4. Large-scale propagation of selected FHA-70 and -80 clones

Because of concerns regarding batch-to-batch variability (Chapter 6, §6.3.4) among previous virion preparations (§2.15.2.d, §2.15.3.c) of antibody-reactive clones (particularly with respect to varying quantities of non-virion contaminants that may have interfered with virion binding to Immulon plates), 44 selected *B. pertussis fha*B-derived clones, together with two fDRW70 pseudorevertants (to be employed as controls) were propagated in large volume and harvested from culture supernatant in such a way as to minimize variation among the five batches of such preparations.

(a) Culture of phage-infected* E. coli *K91-Kan. For each clone, one µL PEG-precipitated virions from a preparation that had served as a source of sequencing template for identification of unique clones was used to infect (§2.5.3) 1 mL *E. coli* K91-Kan, and infected

cells were transferred to 200 mL 2xYT + 20 μg tetracycline mL^{-1} in 1 L Erlenmeyer flasks. After 20.5-22.5 h incubation (OD_{600} = 1.5-1.8 for 23 arbitrarily selected cultures), virions were harvested from culture supernatant and purified as follows.

(b) PEG precipitation and syringe filtration. After centrifuging 195.0 ± 0.5 mL of each culture in a 250 mL centrifuge bottle (15 min, 5,500 x g_{max}, 4°C), 190.0 ± 0.5 mL culture supernatant and 28.5 mL PEG formulation "B" (§2.6.3) were combined and mixed by 100 inversions. After overnight incubation at 4°C, each sample was centrifuged (1 h, 15,300 x g_{max}, 4°C), supernatant was discarded and each bottle drained (≥0.5 h) on paper towel. After addition of 12.5 mL TBS to each pellet, each sample was rocked gently (room temperature) for 0.5 h before transferring resuspended virions into 50 mL centrifuge bottles. After centrifugation (15 min., 4,400 x g_{max}, 4°C), 12.5 ± 0.5 mL cleared supernatant was transferred to a 50 mL centrifuge bottle and precipitated again, as follows. After addition of 1.9 mL PEG formulation B to each bottle and subsequent mixing (by 100 inversions), each sample was incubated overnight at 4°C. After centrifugation (0.5 h, 35,000 x g_{max}, 4°C), supernatant was discarded and each bottle drained (≥0.5 h) on paper towel. After addition of 3.0 mL TBS, each bottle was incubated overnight at 4°C before transferring resuspended virions as two 1.5 mL aliquots into two 2-mL Eppendorf tubes. After centrifuging each tube 3 min. at maximum speed in an Eppendorf 5415C microfuge, 1.4 mL of cleared supernatant was transferred to a fresh Eppendorf tube and precipitated again, as follows. After addition of 210 μL PEG formulation B to each tube and subsequent mixing (by 100 inversions) each sample was incubated 4 h on ice before centrifugation (30 min., maximum speed in an Eppendorf 5415C microfuge, 4°C). After careful removal of supernatant by pipettor, tubes were centrifuged briefly and residual supernatant again removed by pipettor. After addition of 525 μL PBS, each sample was incubated overnight at 4°C. Virions within each sample were resuspended by gentle pipetting, incubation for 1 h at 37°C, and repeated pipetting. After centrifugation

(10 min., maximum speed in a 5415C, 4°C), 500 µL cleared supernatant of each of the two samples for each clone were pooled into a 1.5 mL Eppendorf tube, and centrifuged as before. Centrifugation (15 min. at room temperature) and subsequent transfer of cleared supernatant to fresh tubes was repeated twice. Finally, each sample was filtered with a 0.2 µm syringe filter (VWR). Recovered virions were quantified by UV spectroscopy (§2.7.1).

2.15.5. Assessment of biopanning enrichment (Figure 7-31)

For each of libraries 70-A and 70-B, quantities of unenriched library virions (*viz.*, not biopanned) and biopan eluates (§2.14.2) sufficient to yield $\leq 5 \times 10^3$ plaques on each of two lawns were plaqued on *E. coli* K91-Kan in a manner similar to full-plate plaque assays (§2.7.2.a). After overnight incubation (37°C), plaque lifts of these lawns were performed essentially as described earlier (§2.15.2.b), except that plaque lifts were probed with a 1:8,000 dilution of Protein A-purified FN2/4 pAbs (§2.13.3.a) and color development was stopped after 90 minutes. Positive plaques and total plaques were counted as described earlier (§2.15.2.b). Additionally, concurrent plaque lifts of selected library 70-A and 70-B virions were probed with Protein A-purified α-f1 pAbs (1:10,000 dilution, §2.13.1) to demonstrate adequacy of plaque transfer (illustrated in Figure 2-11E).

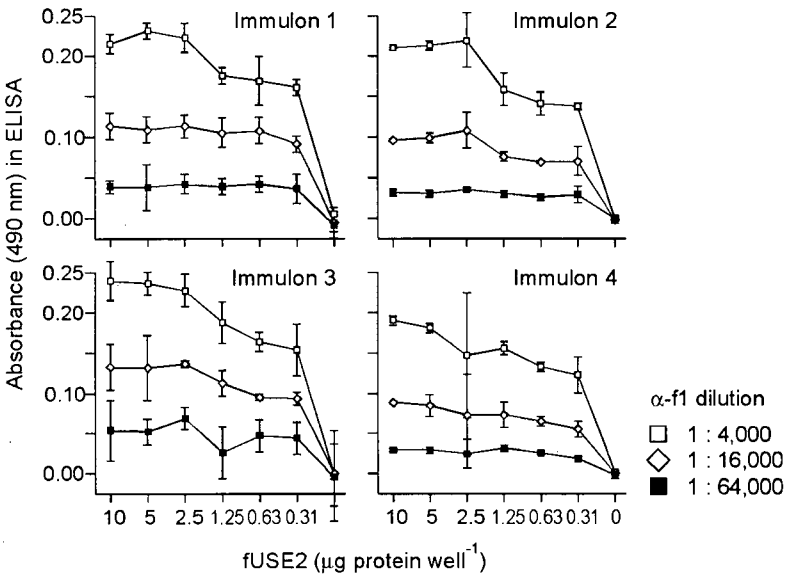


Fig. 2-12. Evaluation of Immulon plates for ELISA. Wells of Immulon-1, -2, -3 and -4 microtiter plates were coated with the indicated quantities of CsCl-purified fUSE2 virions and probed with Protein A-purified rabbit α-f1 (1°) pAbs and peroxidase-conjugated 2° antibodies (§2.15.5.a). Values shown are means of triplicate wells ± two standard deviations.

2.16. ELISA METHOD "A": IMMOBILIZED VIRIONS.

2.16.1. Methodology

(a) Evaluation of Immulon plates and determination of optimal virion quantity for coating plates (Figure 2-12). Triplicate wells of Immulon-1, -2, -3 and -4 plates were coated with 0-10 μg CsCl-purified (§2.6.5) fUSE2 virions in 100 μL PBS and incubated ~20 h at 4°C before washing three times (PBS + 0.05% Tween 20). After addition of blocking buffer (PBS + 1% BSA, 200 μL $well^{-1}$), plates were incubated 1.1 h at 37°C before washing three times. After addition of four-fold dilutions (1:4,000-1:64,000) of protein A-purified rabbit α-f1 pAbs (§2.13.1), plates were incubated 1.25 h at 37°C before washing three times. After addition of peroxidase-conjugated 2° antibodies (1:3,000 dilution in blocking buffer, 100 μL $well^{-1}$; §2.13.4), plates were incubated 1 h at 37°C before washing six times, developing with OPD (§2.13.5) and reading colored reaction products as described later (§2.16.2.1). Although mean absorbance values varied little among plates (Figure 2-12), coefficients of variation varied considerably from plate to plate and with the different quantities of virions used to coat wells. Considering this, Immulon-2 plates coated with 2 μg phage protein $well^{-1}$ were used in subsequent ELISAs.

(b) BCA assay for quantifying virions bound to Immulon plates (Figure 2-13). To determine the feasibility of using the colorimetric bicinchoninic acid (BCA) assay (Smith *et al.* 1985) to quantify virions (as total protein) bound to Immulon-2 plates, triplicate wells were coated 36 h at 4°C with 0-2.5 μg CsCl-purified (§2.6.5) fUSE2 virions (in PBS, 100 μL $well^{-1}$) and washed in the manner described for ELISA (§2.16.1.a). After adding a mixture of 10 μL H_2O and 200 μL BCA reagent (Sigma bicinchoninic acid kit for protein determination) to each well, plates were incubated at 37°C and sample absorbances (A_{570}) measured at one- to eight-hour intervals (Figure 2-13A) using a Biorad 3550 microplate reader. Concurrently, 0-5 μg of (i) fUSE2 virions and (ii) BSA (in PBS, 10 μL $well^{-1}$) were loaded into triplicate wells of

Immulon-2 plates and, after adding 200 µL BCA reagent to each well (without wells having been washed as for ELISA), plates were incubated at 37°C and A_{570} determined as described above.

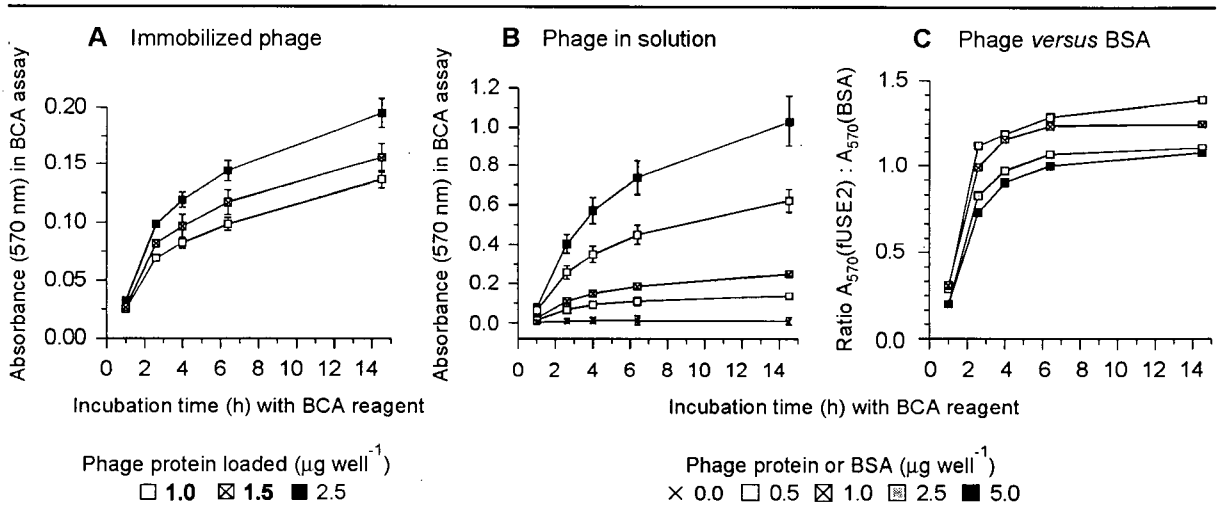


Fig. 2-13. BCA assay for quantification of virions bound to Immulon plate wells: effect of incubation time. **A**. CsCl-purified fUSE2 virions coated onto wells of Immulon-2 microtiter plates and assayed (using a colorimetric BCA assay) for total protein remaining after washing plates in the manner employed for ELISA. **B**. fUSE2 virions and BSA assayed in the same manner as in **A** except that plates were not washed. Values in **A** and **B** are means of triplicate wells ± two standard errors.

Although A_{570} values for virions in solution (Figure 2-13, plot B) were higher than those for virions bound to wells (plot A; this may reflect a low fraction of virions that bind to wells), in both cases (A and B) A_{570} was proportional to virion concentration over a wide range of incubation times, and the rate of change in A_{570} had begun to diminish by 4 h. As well, the ratio of A_{570} values for virions in solution *versus* BSA (BSA chosen to be illustrative of a conventional application of a BCA assay) had reached a near plateau by 4 h (Figure 2-13C). These findings indicated that virions bound to Immulon-2 plate wells could be quantified using a BCA assay with ≥4 incubation at 37°C.

2.16.2. ELISA of CSP-library clones (Chapter 4)

(a) Standard ELISA (Figures 4-3 to 4-5). Triplicate wells of Immulon-2 plates were coated with (i) 2 µg CsCl-purified (§2.6.5) virions of CSP-library clones or fUSE2 (negative

control) in 100 μL PBS, or (ii) 100 ng recombinant CSP protein PfR32tet32 or PvNS1V20 or a 1:500 dilution of culture supernatant containing PBtet32, in 100 μL PBS, or (iii) as a control, PBS only. After >16 h incubation at 4°C, plates were washed three times (PBS + 0.05% Tween 20) before addition of blocking buffer (PBS + 1% BSA + 0.5% bovine milk casein, 200 μL $well^{-1}$), incubation for 0.5-1.5 h at 37°C, and washing as before. After addition of (i) three dilutions (see §2.16.2.2) of α-CSP (1°) mAbs (in blocking buffer, 100 μL $well^{-1}$) and (ii) as a control, no 1° mAbs, plates were incubated ~1 h at 37°C before washing three times. After addition of peroxidase-conjugated 2° antibodies (1:3,000 dilution in blocking buffer, 100 μL $well^{-1}$; §2.13.4), plates were incubated ~1 h at room temperature before washing 6 times. After addition of OPD(§2.13.5), reactions were allowed to proceed ≤35 min. before being stopped by addition of 3N H_2SO_4 (100 μL $well^{-1}$). Colored reaction products were measured at 490 nm using a Biorad model 3550 microplate reader.

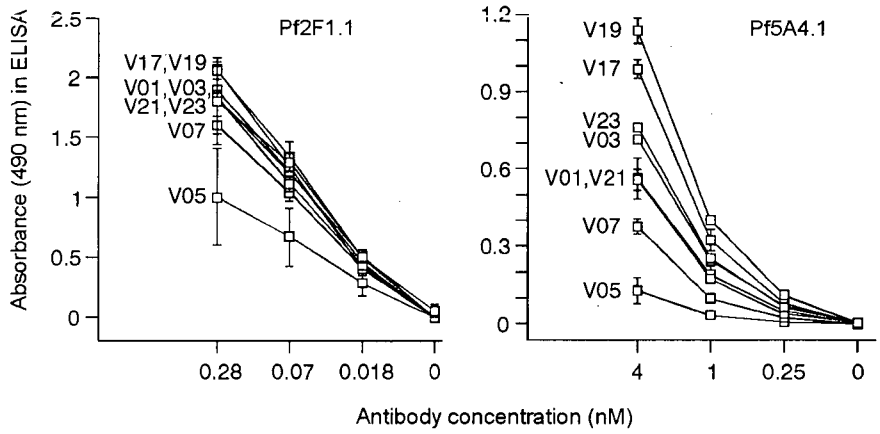


Fig. 2-14. Titration of recognition of CSP-library clones by α-CSP mAbs. Immulon-2 plates were coated with CsCl-purified CSP-library clones (2 μg $well^{-1}$) and probed with the indicated quantities of α-CSP mAbs and peroxidase-conjugated 2° antibodies (§2.16.2.a). Values shown are means ± two standard errors. Most means are of triplicate wells (n=3) coated with virions from a single clone or a pool of two independently isolated and processed clones displaying the same peptide. Other means are of triplicate wells of each of two independently isolated, processed and assayed clones displaying the same peptide (n=6). Encoded peptides are: V01, NANPNPN; V03, NANPDPN; V05, NANANPN; V07, NANADPN; V17, NPNPNPN; V19, NPNPDPN; V21, NPNANPN; V23, NPNADPN. Data shown here derive from the experiment summarized in Figures 4-5B and C (Chapter 4).

(b) Titration curves for CSP-library clones probed with α-CSP mAbs (Figure 2-14). Antibody concentrations employed in the ELISAs described above were chosen to emphasize clone-to-clone differences in peptide-MAb interaction. In most cases, the ELISA signal was titerable in a near linear fashion around the second of the three chosen mAb concentrations; *viz.*, increasing or decreasing mAb concentration yielded a corresponding increase or decrease in ELISA signal. This is illustrated in Figure 2-14 for CSP-library clones that reacted strongly (Chapter 4, Figure 4-5) with two mAbs employed in this study, Pf2F1.1 and Pf5A4.1.

(c) Quantification of CSP-library clones bound to Immulon plates (Figure 4-8). Immulon-2 plates were coated and washed as described for ELISA (§2.16.2.a). After a mixture of 10 μL H_2O and 200 μL BCA reagent (§2.16.1.b) was added to each well, plates were incubated 4 h at 37°C before measuring relative quantities of bound virion protein as described earlier (§2.16.1.b).

(d) Effect on ELISA of varying virion quantities, selected CSP-library clones (Figure 4-9).

These ELISAs were performed essentially as described earlier (§2.16.2.a), except that wells were coated with 4, 2, 1, 0.5, 0.25 and 0 μg CsCl-purified (§2.6.5) virions.

(e) Competition ELISA, selected CSP-library clones (Figure 4-10). These ELISAs were performed generally as described earlier (§2.16.2.a) except that each dilution of each 1° mAb was preincubated (2 h at room temperature) with varying quantities (0 to 100 μM in blocking buffer) of the synthetic peptide NPNANPNANPNA (Sigma). As well, virions used in these assays were purified only by two-stage PEG precipitation (§2.6.4), not by CsCl density gradient centrifugation.

2.16.3. Preliminary ELISA of random peptide library clones with α-CSP mAbs

(a) ELISA "A" (Figures 5-3 and 5-4, graphs labelled "A"). For each of the four random peptide library biopannings that employed 120 ng mAb (§2.14.1), individual wells of

Immulon-2 plates were coated with (i) 2 μg two-stage PEG precipitated virions (§2.6.4) of 24 clones selected from the third round of biopanning and (ii) as a control, coating buffer only (PBS). After overnight incubation (4°C), plates were washed three times (PBS + 0.05% Tween 20) before adding 200 μL well^{-1} blocking buffer (PBS + 1% BSA + 0.5% bovine milk casein) and incubating 1 h at 37°C. After washing three times, a relatively high concentration of 1° antibodies (1:333 dilution in blocking buffer of Pf2A10, Pf1B2.2 or Pb4B10, as appropriate) was added to each well (100 μL well^{-1}) and plates were incubated 0.75 h at 37°C before washing three times. After addition of peroxidase-conjugated 2° antibodies (1:3000 dilution in blocking buffer, 100 μL well^{-1}; §2.13.4), plates were incubated 0.75 h at 37°C before washing, developing with OPD (§2.13.5) and reading reaction products as described earlier (§2.16.2.a).

(b) ELISA "B" (Figures 5-3 and 5-4, graphs labelled "B"). These ELISAs were performed as described above (§2.16.3.a), except (i) 1° antibodies were used at a 1:500 dilution; (ii) additional controls were employed in which 1° antibodies Pf2A10 and Pf1B2.2 were pre-incubated with the recombinant CSP protein R32tet32 (10 ng and 1 μg mL^{-1}) ~2 h at room temperature; and (iii) plates were incubated only 0.5 h with 1° and 2° antibodies.

2.16.4. Initial ELISA and BCA assays of FHA-70 and -80 clones (Figures 6-3 and 6-4)

(a) Pooling of siblings and amplification of selected clones. Two-stage PEG-precipitated virions amplified from initial antibody-reactive plaques selected from plaque lifts of FHA-70 and -80 libraries (§2.15.2.d, §2.15.3.c), and which had been employed as a source of sequencing template to identify unique clones, were pooled or amplified as follows. For selected clones for which multiple siblings existed, virion preparations from up to 22 siblings were pooled. Other clones, particularly those possessing unique sequences (*i.e.*, no siblings existed), were amplified by infecting (§2.5.3) *E. coli* K91-Kan with a small aliquot of virions and culturing infected cells overnight in LB-Tet (35-50 mL) before harvesting virions by two-

stage PEG precipitation (§2.6.4).

(b) ELISA and concurrent BCA assay (Figure 6-3). For each unique clone, triplicate wells of Immulon-2 microtiter plates were coated with (i) 1 μg PEG-precipitated virions (in PBS, 50 μL well-1) of antibody-reactive clones (a single pool, as described in the preceding section, or one or more individual siblings) and, as controls, (ii) virions of fDRW70 pseudo-revertants and (iii) PBS only. After overnight incubation at 4°C, plates were washed three times (PBS + 0.05% Tween 20) before adding blocking buffer (PBS + 1% BSA + 1% skim milk powder, 150 μL $well^{-1}$) and incubating 2 h at 37°C. After plates were washed three times, 1° antibodies (1:2,000 and 1:8,000 dilutions in blocking buffer, *E. coli*-absorbed FN2/4 and FS1/4 α-FHA pAbs, §2.13.3.a; 100 μL $well^{-1}$) were added and plates were incubated 1 h at 37°C. After plates were washed 3 times, peroxidase-conjugated 2° antibodies were added (1:3,000 dilution in blocking buffer, 100 μL $well^{-1}$; §2.13.4) and plates were incubated 1 h at 37°C before washing, developing with OPD (§2.13.5) and reading reaction products (Figure 6-3B) as outlined earlier (§2.16.2.a). Concurrently, a second set of Immulon-2 plates was coated and washed as for ELISA, and quantities of bound virions were determined by a BCA protein assay performed essentially as described earlier (§2.16.2.c). In a followup ELISA (Figure 6-3D), selected clones were assayed in the same manner as before, except that virions were probed with (i) FN2/4 pAbs (at a 1:2,000 dilution) and (ii) as a control to demonstrate lack of recognition of virions by 2° antibodies, with no 1° antibodies.

(c) Clone-to-clone variability of binding to Immulon plates. Selected clones (specifically, three that bound poorly to Immulon-2 plates, four that bound well) assayed as described in the preceding section were again assayed, except that varying quantities of virions were used to coat Immulon-2 plate wells. Duplicate plates (with duplicate wells for each clone on each plate) were coated with 5 times (5x), 1x and 0.2x the quantity (1 μg in 50 μL PBS) of virions employed previously. An ELISA with a 1:2,000 dilution of *E. coli*-absorbed protein

A-purified FN2/4 (one plate) and a BCA assay (the 2nd plate) were performed essentially as described in the preceding section.

2.16.5. Competition ELISA and related BCA assays of FHA-70 and -80 clones

(a) Clone-to-clone variability of binding to Immulon plates (Figure 6-7). For each of 44 antibody-reactive clones and for two pseudorevertants of fDRW70 (A and B), duplicate wells of Immulon-2 plates were coated with two-fold serial dilutions (250 ng to 8 μg $well^{-1}$, 100 μL $well^{-1}$) of PEG-precipitated and syringe-filtered virions (§2.15.4). After overnight incubation at 4°C, plates were washed and a BCA protein assay performed essentially as described earlier (§2.16.2.c).

(b) Competition ELISA. α-FHA antibodies for this "competition" ELISA were prepared by preincubating 1:5,000 dilutions of each of *E. coli*-absorbed FN2/4 and FS1/4 (§2.13.3.a; in blocking buffer with the concentration of NaCl adjusted to 233 mM) with heparin sepharose-affinity purified FHA at final concentrations of 30 μg, 5 μg, 833 ng and 0 ng mL^{-1} for 2.75 h at 37°C before diluting the antibody preparations 2-fold immediately prior to their use in ELISA. For each antibody-reactive clone or control (fDRW70 pseudorevertant), duplicate wells of Immulon-2 plates were coated with 1 μg PEG-precipitated and filtered (§2.15.4) virions (100 μL $well^{-1}$ in PBS). After overnight incubation at 4°C, plates were washed 3 times (PBS + 0.05% Tween 20) before adding blocking buffer (PBS + 1% BSA + 1% skim milk powder, 200 μL $well^{-1}$) and incubating 1 h at 37°C. After plates were washed 3 times, α-FHA antibodies prepared as described above were added (100 μL $well^{-1}$) and plates were incubated 0.75 h at 37°C. After plates were washed 3 times, peroxidase-conjugated 2° antibodies (1:3,000 dilution in blocking buffer, 100 μL $well^{-1}$, §2.13.4) were added and plates were incubated 1 h at 37°C before washing, developing with OPD (§2.13.5) and reading reaction products (Figure 6-8B) as described earlier (§2.16.2.a). Concurrently, an additional set of Immulon-2 plates were coated with virions and washed as for ELISA before quantifying the

relative quantities of virions remaining bound to wells (Figure 6-8C) by a BCA protein assay as described earlier (§2.16.2.c). Concurrent controls for this BCA assay, generally illustrative of those periodically employed for other BCA assays of bound virion protein, are shown in Figure 2-15.

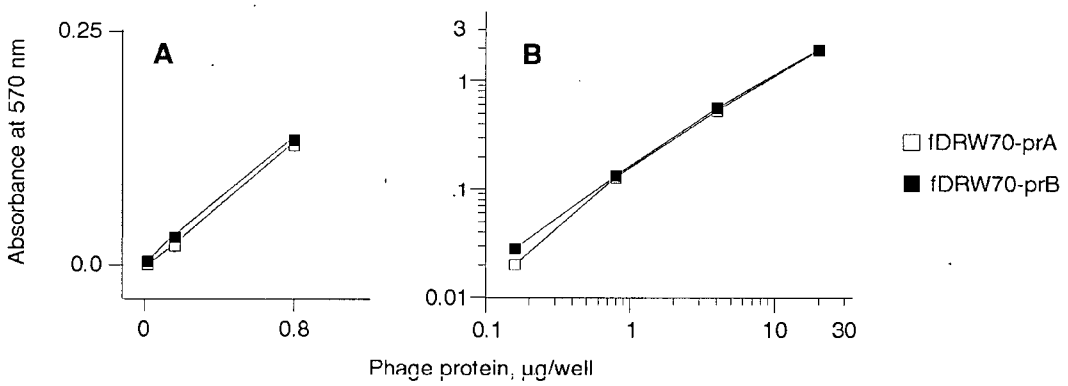


Fig. 2-15. Control for BCA assay. In an assay performed concurrently with those of §2.16.5.b, the indicated quantities of virions of fDRW70 pseudorevertants "prA" and "prB" were loaded into triplicate wells of Immulon-2 plates. Total virion protein was assayed (without washing wells as in the corresponding assay of bound virions) using a colorimetric BCA protein assay. Data in **A** and **B** illustrate overlapping sets of data points derived from the same assay; note that axes in A are linear-linear, while those in B are log-log. Values shown are means of triplicate wells. Although error bars representing two standard errors were plotted, these are not evident because of their small size: they are effectively "hidden" by the data point symbols

2.17. ELISA METHOD "B": IMMOBILIZED ANTIBODIES

2.17.1. Capture ELISA of selected CSP-library clones (Figure 4-7)

After addition of biotinylated mAbs (Pf2A10, 2.5 ng ≈ 17 fmol well^{-1}; Pf1B2.2, 250 ng ≈ 1.7 pmol well^{-1}) in wash buffer (100 µL well^{-1}; 25 mM Tris pH 7.6, 150 mM NaCl, 0.1% BSA, 0.05% Tween 20) to wells of streptavidin-coated 96-well microtiter plates (28 pmol biotin-binding capacity per well, §2.14.1.a), plates were incubated 2 h at room temperature before washing three times (200 µL well^{-1}). After addition of CsCl-purified (§2.6.5) virions (0.08-2 µg well^{-1} in 100 µL wash buffer), plates were incubated 30 min. at room temperature before washing three times. After addition of Protein A-purified rabbit α-f1 pAbs (1:5,000 dilution in wash buffer; 100 µL well^{-1}; §2.13.1), plates were incubated 30 min. at room

temperature before washing three times. After addition of peroxidase-conjugated 2° antibodies (1:5,000 dilution in wash buffer, 100 μL well^{-1}, §2.13.4), plates were incubated 30 min. at room temperature before washing three times, developing with OPD (§2.13.5) and reading reaction products generally as described earlier (§2.16.2.a).

2.17.2. Capture ELISA of random peptide library biopan eluates (Figure 5-2)

After addition of biotinylated mAbs (Pf2A10 or Pf1B2.2, 120 ng ≈ 0.8 pmol well^{-1}) in wash buffer (100 μL well^{-1}, as in preceding section) to wells of streptavidin-coated microtiter plates (28 pmol biotin binding capacity per well, §2.14.1.a), plates were incubated and washed as described above (§2.17.1). After addition of 3.6 μg PEG-precipitated virions (amplified output from third round of biopanning or unbound fraction from first round, §2.14.1), plates were incubated and washed as described above (§2.17.1). After addition of rabbit α-f1 1° Abs (as in §2.17.1, but additionally purified to remove antibodies that may have reacted with *E. coli* contaminants in PEG-precipitated virion preparations, using a Pierce Immunochemicals immobilized *E. coli* lysate kit in accordance with the manufacturer's instructions), plates were incubated 0.75 h at room temperature before washing, adding 2° antibodies and performing other procedures as described above (§2.17.1).

2.18. IMMUNOBLOTS

2.18.1. Assessment of binding to nitrocellulose of FHA-70 and -80 clones (Figure 6-4) and immunoblots of antibody-reactive clones (Figure 6-6, assay (i))

For FHA library clones and for two fDRW70 pseudorevertants (A and B), triplicate ~2 μL samples (800 ng virion protein) of PEG-precipitated and filtered virions (§2.15.4) were applied to nitrocellulose discs (§2.13.5) in the pattern shown in Figure 6-5 (Chapter 6). Triplicate ~2 μL samples of two-fold serial dilutions (per 2 μL sample: 6 ng to 3.2 μg) of PEG-precipitated and filtered virions (§2.15.4) of fDRW70 pseudorevertants were similarly applied to nitrocellulose discs, but in a grid fashion (Chapter 6, Figure 6-4). After samples had

dried, each nitrocellulose disc (or "blot", contained in a standard petri dish) was incubated 1 h at room temperature with gentle rocking in 10 mL blocking buffer (PBS + 3% BSA + 1% skim milk powder). After three 10 min. washes at room temperature in 10 mL wash buffer (PBS + 0.05% Tween 20), each blot was incubated 1 h at room temperature with gentle rocking in 10 mL primary antibody (per blot, one of (i) 1:8,000, (ii) 1:32,000 or (iii) 1:128,000 dilution in blocking buffer of protein A-purified α-f1 pAbs (§2.13.1); additionally, for library clones, 1:2,000 dilution of *E. coli*-absorbed (§2.13.3.a) (iv) FN2/4 or (v) FS1/4 pAbs or (vi) crude FN1/4 α-FHA serum, or (vii) as a control, no antibody). After three 10 min. washes, each blot was incubated 1 h at room temperature with gentle rocking in 10 mL alkaline phosphatase-conjugated 2° antibodies (1:3,000 dilution in blocking buffer; §2.13.4). After three 10 min. washes in wash buffer followed by a single 10 min. wash in 10 mL substrate buffer (§2.13.5), 10 mL BCIP/NBT substrate (§2.13.5) was added to each blot. Reactions were allowed to proceed 5 min. (blots probed with 1:8,000 and 1:32,000 dilutions of α-f1 pAbs), 9 min. (1:128,000 of α-f1 pAbs) or 14 min. (α-FHA pAbs or serum) before being stopped by rinsing blots in water.

2.18.2. Immunoblots of antibody-reactive FHA-70 and -80 clones (Figure 6-6, (ii))

For each of the antibody-reactive clones shown in Figure 6-5 (Chapter 6), and for two fDRW70 pseudorevertants (A and B), triplicate ~2 μL samples (800 ng phage protein) of PEG-precipitated and filtered virions (§2.15.4) were applied to nitrocellulose discs in the pattern indicated in Figure 6-5. After drying, blots were probed with 1:8,000, 1:32,000 or 1:128,000 dilutions in blocking buffer of (i) protein A-purified α-f1 pAbs (§2.13.1) or *E. coli*-absorbed (§2.13.3.a) (ii) FN2/4 or (iii) FS1/4 α-FHA pAbs or (iv) crude FN1/4 α-FHA serum. Blots were washed, blocked, probed with 2° antibodies as described above (§2.18.1). Blots probed with α-f1 pAbs were stopped after 6.5 min., while those probed with α-FHA pAbs or serum were stopped after 13.5 min.

2.19. ANALYSIS OF ASX-PRO TURNS

Methods employed in analysis of Asx-Pro turns are described in Chapter 3.

2.20. PREDICTED SIGNAL PEPTIDE CLEAVAGE

Each of the peptides encoded by the antibody-reactive FHA-70 and -80 library clones identified in Figure 6-1 (Chapter 6) were examined for signal peptide cleavage expected from their display in three alternative vectors: (i) fDRW70, which separates displayed peptides from signal peptides with the linker sequence ADGAGA; (ii) an fDRW8nn vector, which separates displayed peptides with ADGP; and (iii) a vector which uses no linker, but rather fuses the displayed peptide directly to the signal peptide. Each of the peptide sequences shown in Table 2-VI reflects an FHA library clone as it would be expressed in one of these three vectors; these sequences were analysed by the neural networks developed by Nielsen *et al.* (1997) and available via the WWW server of the Center for Biological Sequence Analysis (http://www.cbs.dtu.dk/). C-scores (cleavage-site scores, one of three scores provided by the networks) and most likely cleavage positions are tabulated in Table 2-VI and summarized in Figure 6-9 (Chapter 6).

Table 2-VI. Predicted[a] signal peptidase cleavage of antibody-reactive FHA-70 and -80 clones.

Clone ID[b]	Last residue of native preprotein + 30 residues of N-terminal mature region[c]	C-score[d]
I-a	S.ADGAGALSVSAGGNLRAKQLVSSAQLEVRG	0.521
	S.ADGPLSVSAGGNLRAKQLVSSAQLEVRGQR	0.427
	LSVSA.GGNLRAKQLVSSAQLEVRGQREVAL	0.274
I-b	S.ADGAGAIAPRGWRATTISQVDLHDLSAARG	0.521
	S.ADSPIAPRGWRATTISQVDLHDLSAARGAD	0.427
	SIA.PRGWRATTISQVDLHDLSAARGADISGG	0.156
I-c	S.ADGAGADLSAARGADISGEGRGAGSGAETV	0.521
	S.ADGPDLSAARGADISGEGRGAGSGAETVES	0.459
	S.DLSAARGADISGEGRGAGSGAETVESCLAK	0.188
I-d	S.ADGAGALSAARGADISGGVGTGAETVESCL	0.521
	S.ADGPLSAARGADISGGVGTGAETVESCLAK	0.427
	SLSA.ARGADISGGVGTGAETVESCLAKPHTE	0.615
I-e	S.ADGAGASAARGADISGGAGSGAETVESCLA	0.521
	S.ADGLSAARGADISGGAGSGAETVESCLAKP	0.427
	SSAA.RGADISGGAGSGAETVESCLAKPHTEN	0.518
I-f	S.ADGAGALSAARGADISGEGRGVGSGAETVE	0.521
	S.ADGPLSAARGADISGEGRGVGSGAETVESC	0.427
	SLSA.ARGADISGEGRGVGSGAETVESCLAKP	0.615
I-g	S.ADGAGALSAARGADISGEGRGAGSGAETVE	0.521
	S.ADGPLSAARGADISGEGRGAGSGAETVESC	0.427
	SLSA.ARGADISGEGRGAGSGAETVESCLAKP	0.615
I-h	S.ADGAGASAARGADISGEGGAGTGAETVESC	0.521
	S.ADGPSAARGADISGEGGAGTGAETVESCLA	0.427
	SSAA.RGADISGEGGAGTGAETVESCLAKPHT	0.518
II	S.ADGAGAALSIDSMTALGAGAGAETVESCLA	0.521
	S.ADGPSAARGADISGEGGAGTGAETVESCLA	0.427
	SSAA.RGADISGEGGAGTGAETVESCLAKPHT	0.518
III-a	S.ADGAGAMTVRDVAAAADLALQAGDASGAGA	0.521
	S.ADGPMTVRDVAAAADLALQAGDASGAGAET	0.514
	S.MTVRDVAAAADLALQAGDASGAGAETVESC	0.514
III-b	S.ADGAGAMTVRDVAAAADLALQAGDALPGAG	0.521
	S.ADGPMTVRDVAAAADLALQAGDALPGAGAE	0.427
	S.MTVRDVAAAADLALQAGDALPGAGAETVES	0.236
III-c	S.ADGAGARDVAAAADLALQGVGTGAETVESC	0.521
	S.ADGPRDVAAAADLALQGVGTGAETVESCLA	0.427
	S.RDVAAAADLALQGVGTGAETVESCLAKPHT	0.370
IV	S.ADGAGAIVVEAGELVSHAGGAGSGAETVES	0.521
	S.ADGPIVVEAGELVSHAGGAGSGAETVESCL	0.427
	SIVVEA.GELVSHAGGAGSGAETVESCLAKPH	0.316

V	S.ADGAGATKGEMQIAGKGGGSPGAGSGAETV	0.521
	S.ADGPTKGEMQIAGKGGGSPGAGSGAETVES	0.427
	STKGE.MQIAGKGGGSPGAGSGAETVESCLAK	0.217
VI-a	SADG.AGADYTVSADAIALAGVGTGAETVESC	0.521
	S.ADGPDYTVSADAIALAGVGTGAETVESCLA	0.459
	S.DYTVSADAIALAGVGTGAETVESCLAKPHT	0.208
VI-b	S.ADGAGATVSADAIALAAQGAGTGAETVESC	0.521
	S.ADGPTVSADAIALAAQGAGTGAETVESCLA	0.427
	STVS.ADAIALAAQGAGTGAETVESCLAKPHT	0.487
VII	S.ADGAGANKIRLMGPLQGVGSGAETVESCLA	0.521
	S.ADGPNKIRLMGPLQGVGSGAETVESCLAKP	0.427
	S.NKIRLMGPLQGVGSGAETVESCLAKPHTEN	0.211
VIII	S.ADGAGAITVTSRGGFDNEGKMESNKAGAGA	0.521
	S.ADGPITVTSRGGFDNEGKMESNKAGAGAET	0.427
	S.ITVTSRGGFDNEGKMESNKAGAGAETVESC	0.179
IX-a	S.ADGAGAHLRNTGQVVAGQGVGTGAETVESC	0.521
	S.ADGPHLRNTGQVVAGQGVGTGAETVESCLA	0.427
	SHLR.NTGQVVAGQGVGTGAETVESCLAKPHT	0.246
IX-b	SADG.AGAVVAGHDIHIMGVGTGAETVESCLA	0.521
	S.ADGPVVAGHDIHIMGVGTGAETVESCLAKP	0.427
	SVVA.GHDIHIMGVGTGAETVESCLAKPHTEN	0.618
X-a	SADG.AGANKLGRIRAGEDMQGAGAETVESCL	0.521
	S.ADGPNKLGRIRAGEDMQGAGAETVESCLAK	0.427
	S.NKLGRIRAGEDMQGAGAETVESCLAKPHTE	0.206
X-b	SADG.AGANKLGRIRAGEDMHLDAPRIEGVGT	0.521
	S.ADGPNKLGRIRAGEDMHLDAPRIEGVGTGA	0.427
	S.NKLGRIRAGEDMHLDAPRIEGVGTGAETVE	0.145
XI-a	S.ADGAGALDQNRYEYIWGLYPGAGAGAETVE	0.521
	S.ADGPLDQNRYEYIWGLYPGAGAGAETVESC	0.427
	S.LDQNRYEYIWGLYPGAGAGAETVESCLAKP	0.249
XI-b	SADG.AGADQNRYEYIWGLYQGVGSGAETVES	0.521
	S.ADGPDQNRYEYIWGLYQGVGSGAETVESCL	0.459
	SDQN.RYEYIWGLYQGVGSGAETVESCLAKPH	0.311
XI-c	S.ADGAGANRAGLSPATWNFQSTYELLDYLLD	0.521
	S.ADGPNRAGLSPATWNFQSTYELLDYLLDQN	0.427
	SNRA.GLSPATWNFQSTYELLDYLLDQNRYEY	0.488
XI-d	S.ADGAGAEYIWGLYQTYTEWSVNTLKNLDLG	0.521
	S.ADGPEYIWGLYQTYTEWSVNTLKNLDLGYQ	0.479
	S.EYIWGLYQTYTEWSVNTLKNLDLGYQAKPA	0.202
XI-e	S.ADGAGATEWSVNTLKNLDLGYQAKPAPTAP	0.521
	S.ADGPTEWSVNTLKNLDLGYQAKPAPTAPPM	0.427
	S.TEWSVNTLKNLDLGYQAKPAPTAPPMPKAP	0.208
XII	S.ADGAGAVHDQLGQRYGKALGGMDAETKEVD	0.521
	S.ADGPVHDQLGQRYGKALGGMDAETKEVDGI	0.427
	SVHD.QLGQRYGKALGGMDAETKEVDGIIQEF	0.514
XIII	S.ADGAGAGRPEGLKIGAHSATSVSGSFDALP	0.521
	S.ADGPGRPEGLKIGAHSATSVSGSFDALPGA	0.427
	S.GRPEGLKIGAHSATSVSGSFDALPGAGAGA	0.201

XIV	S.ADGAGALAAVLVNPHIFTRIGAAQTSLADG	0.521
	S.ADGPLAAVLVNPHIFTRIGAAQTSLADGAA	0.427
	SLAA.VLVNPHIFTRIGAAQTSLADGAAGPAG	0.327

[a]Predicted signal peptide cleavage, based on the neural networks of Nielsen *et al.* (1997).

[b]As shown in Figure 6-1 (Chapter 6).

[c]Each table entry shows three sequences, each including up to 30 residues of the N-terminal mature region of a displayed peptide. One of the sequences is that of an antibody-reactive FHA-library clone. The other two reflect the FHA-derived peptide as it would be expressed in an alternative vector, as described in §2.20. The first line of each entry represents the peptide in fDRW70, which separates displayed peptides from signal peptides with a ADGAGA linker peptide. The second line represents the peptide in an fDRW8*nn* vector, in which a ADGP linker is employed, while the third line represents the peptide in a vector in which no linker is employed. The symbol "." indicates the position yielding the highest C-score and the putative cleavage site for C-scores $\geq$0.49. In wild-type Ff phage, cleavage is believed to occur on the C-terminal side of the first Ser residue shown. Each sequence was preceded by the preprotein sequence VKKLLFAIPLVVPFYSH when submitted to the neural networks.

[d]C-score, or cleavage-site score, as described in Nielsen *et al.* (1997) and in §2.20.

Chapter 3

The "Asx-Pro turn" as a local structural motif stabilized by alternative patterns of hydrogen bonds, and a consensus-derived model of the sequence Asn-Pro-Asn

3.1. ABSTRACT

Analyses of databases derived from the Brookhaven Protein Data Bank have identified a set of related turn structures formed by the sequence Asx-Pro-Xxx_n. In a variety of flanking structural contexts, more than 60 per cent of Asx-Pro sequences adopt a turn conformation stabilized by a set of alternative hydrogen bonds among the sidechain O^{δ} and backbone C=O carbonyl oxygens of Asx (residue *i*) and the backbone NH of residues *i*+2, *i*+3 and in some cases *i*+4. In contrast, the structures adopted by Ser-Pro, His-Pro and other Xxx-Pro sequences reflect more heterogenous hydrogen-bonding patterns. As expected, structures formed by Asx-Pro-Asx are similar to those formed by Asx-Pro-Xxx_n but in some cases additional hydrogen bonds are formed between the Asx sidechains. Hydrogen bond patterns within Asx-Pro and Asn-Pro-Asn turns are consistent with published NMR studies of helical $(Asn\text{-}Pro\text{-}Asn\text{-}Ala)_n$ peptides, indicating that a consensus structure reflecting these hydrogen bonds can serve as a partial model of the Asn-Pro-Asn-Ala tetrapeptide repeats of *Plasmodium falciparum* circumsporozoite protein.

3.2. INTRODUCTION

Much of our understanding of protein structure has derived from analyses of databases of known three-dimensional structures, particularly the Brookhaven Protein Data Bank (PDB; Bernstein *et al.* 1977), such as those of backbone dihedral angles (Ramachandran *et al.* 1963; Ramachandran & Sasisekharan 1968), amino acid sidechain conformations (McGregor *et al.* 1987;

Schrauber *et al.* 1993; Dunbrack & Karplus 1994), and backbone and sidechain hydrogen bonds (Kabsch & Sander 1983; Ippolito *et al.* 1990; Stickle *et al.* 1992; McDonald & Thornton 1994). From these and other studies, it is clear that some amino acids influence local conformation more than others, proline being an obvious example: its pyrrolidine ring limits both its own ϕ,ψ conformational space and that of the preceding residue (MacArthur & Thornton 1991). Amino acids with short polar sidechains such as aspartate, asparagine and serine are other examples. These, together with proline, occur frequently in ß-turns (Chou & Fasman 1978, 1979; Wilmot & Thornton 1988) and at the N-termini of α- and 3_{10}-helices, and are strongly associated with certain clusters of backbone conformations (Presta & Rose 1988; Richardson & Richardson 1988; Karpen *et al.* 1992; Dasgupta & Bell 1993). As well, certain structural motifs have been shown to be defined by local sequence motifs. Examples include (a) the "tyrosine corner" (Hemmingsen *et al.* 1994), in which a tyrosine (in a sequence which includes proline) hydrogen bonds its sidechain OH to the backbone NH or CO of a preceding residue; (b) type VI turns formed by aromatic amino acids flanking proline (Yao *et al.* 1994a, 1994b); (c) ring motifs formed by hydrogen bonding of Asn or Gln sidechains to backbone atoms (Le Questel *et al.* 1993); and (d) the SXXE "capping box" in which reciprocal sidechain-mainchain H-bonds (Harper & Rose 1993), together with hydrophobic interactions (Seale *et al.* 1994), stabilize an α-helix cap.

The current study describes another local sequence and structural motif, identified in a narrowly-focussed study of Asx-Pro and other Xxx-Pro sequences. Interest in these specific sequences arose from investigations (Chapter 4; see also Wilson *et al.* 1997) of Asx-Pro-containing peptides derived from the circumsporozoite protein of *Plasmodium falciparum*. The current study was prompted (a) by the intriguing findings of Richardson and Richardson (1988) and related studies (Presta & Rose 1988; Karpen *et al.* 1992; Lyu *et al.* 1992; Chakrabartty *et al.* 1993; Dasgupta & Bell 1993; Forood *et al.* 1993) that there is a striking propensity for asparagine,

aspartate and serine at an α-helix N-cap and for proline in the first helix position, (b) by a similar positional distribution of these residues in ß-turns (Chou & Fasman 1978, 1979; Wilmot & Thornton 1988) and (c) by the Presta and Rose hypothesis (1988) that an α-helix N-cap can be stabilized by hydrogen bonding between the sidechain of a polar residue flanking the helix and the initial backbone NH group.

Given that characteristics of asparagine, aspartate, serine and proline must independently lend themselves to formation of H-bonds required for a turn or helix cap, expected conformations of these residues in combination (Asx-Pro or Ser-Pro) should be few and predictable. Because previous studies have not focussed on these amino acids in combination and, moreover, have generally focussed on a single structure class (α-helix, 3_{10}-helix or ß-turn), several PDB-derived databases representative of all protein structure classes were analysed to specifically examine H-bonds formed by these residues in combination.

3.3. ANALYTICAL METHODS

3.3.1. Molecular databases and datasets

Three datasets were employed, each consisting of Brookhaven PDB coordinates for a number of proteins, their corresponding secondary structure profiles (DSSP profiles) defined by the program DSSP (Kabsch & Sander 1983), and additional H-bond information as described later. PDB coordinates were obtained from the Internet archive of the Brookhaven National Laboratory. Most DSSP profiles were obtained from the Internet archive of the European Molecular Biology Laboratory (EMBL) at Heidelberg, Germany; others were generated in-house.

Each of datasets A and B was chosen to be representative of all protein structure classes. Dataset A consists of 80 non-homologous proteins, 77 of which had been refined to a resolution $\leq$3.0 Å; their Brookhaven codes are as listed in Table 1 of Adzhubei and Sternberg (1993), except that entries 1HMQ and 4FD1 were replaced by more current entries 2HMQ and

5FD1, respectively. Dataset B consists of 102 non-homologous proteins refined to 3.0 Å resolution or better, as described in Boberg *et al.* (1992). Brookhaven codes for dataset B (chain designations in parentheses) are: 1ACX, 1BP2, 1CCR, 1CLA, 1CSE(E,I), 1CTF, 1CTX, 1ECA, 1ETU, 1FC2(C,D), 1GCR, 1GD1(O), 1GOX, 1GP1(A), 1HIP, 1HOE, 1I1B, 1LDM, 1LRD(4), 1LZ1, 1NXB, 1PCY, 1PHH, 1PRC(C,L,M,H), 1PYP, 1R69, 1RHD, 1SGT, 1SN3, 1TIM(A), 1TNF(A), 1UBQ, 1UTG, 1WSY(A,B), 256B(A), 2AAT, 2AZA(A), 2CAB, 2CCY(A), 2CDV, 2CI2, 2CPP, 2CTS, 2CYP, 2FB4(H,L), 2FXB, 2GBP, 2GN5, 2HHB(A,B), 2HLA, 2HMG(A,B), 2LH2, 2LTN(A,B), 2MHR, 2OVO, 2PAB(A), 2PAZ, 2PFK(A), 2RNT, 2RSP(A), 2SGA, 2SNS, 2SOD(O), 2SSI, 2TAA, 2TS1, 2WRP, 3ADK, 3APP, 3B5C, 3BCL, 3FXC, 3FXN, 3GAP(A), 3GRS, 3INS(A,B), 3LZM, 3PGK, 3PGM, 3RN3, 3TLN, 451C, 4CHA(A), 4HVP(A,I), 4PEP, 4XIA(A), 5CPA, 5CPV, 5FD1, 5PTI, 5RXN, 6ACN, 6AT1(A,B), 7API(A,B), 8ADH, 6AT1(A,B), 8CAT(A), 8DFR, 9PAP and 9WGA(A). Most analyses were performed twice, once for each dataset, to assess bias in choice of dataset. Although there is considerable overlap between these datasets (they have 55 proteins in common, and an undetermined number of proteins share at least limited local homology), the different methods (Boberg *et al.* 1992; Adzhubei & Sternberg 1993) used to select the proteins included in each dataset and differences in content were expected to preclude a biased analysis.

Dataset C consists of 39 non-homologous protein fragments containing Asp-Pro-Asn (15 fragments), Asn-Pro-Asn (9 fragments) and Asn-Pro-Asp (15 fragments). These were identified by searching Release 14 of the NRL3D (Pattabiraman *et al.* 1990) database (from the Internet archive of the National Center for Biotechnology Information) for all occurrences of these tripeptides in proteins refined to better than 3.0 Å resolution. Highly homologous or identical fragments were eliminated by choosing the entry with the highest resolution or, for multichain proteins, the lowest chain code. Only structures determined by X-ray crystallography were included. Structures with only C^{α} coordinates were eliminated. A fragment in entry 4ENL

was eliminated because of high temperature factors.

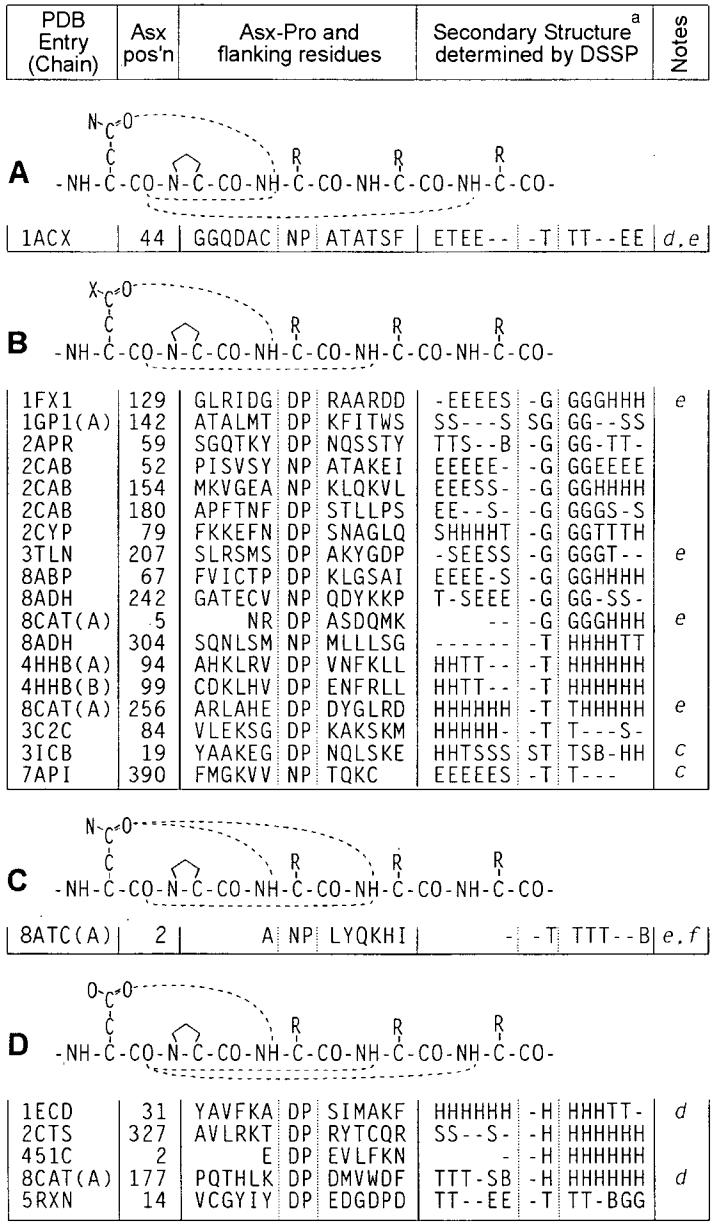
3.3.2. Computer software

The program DSSP was provided by U. Hobohm (EMBL, Germany) under academic licence. The program BABEL version 1.05, provided by P. Walters and M. Stahl (Dolata Research Group, University of Arizona, U.S.A.), was used to calculate interatomic distances from PDB coordinates. The program WHATIF, provided by G. Vriend (EMBL, Germany) under academic licence, was used to identify H-bonds. The program RASMOL version 2.3, provided by R. Sayle (1992) (University of Edinburgh, Scotland), was used to view and manipulate three-dimensional models of PDB coordinates, and to produce the molecular models shown in Figures 3-6, 3-9, 3-11 and 3-12.

3.3.3. Hydrogen bonds

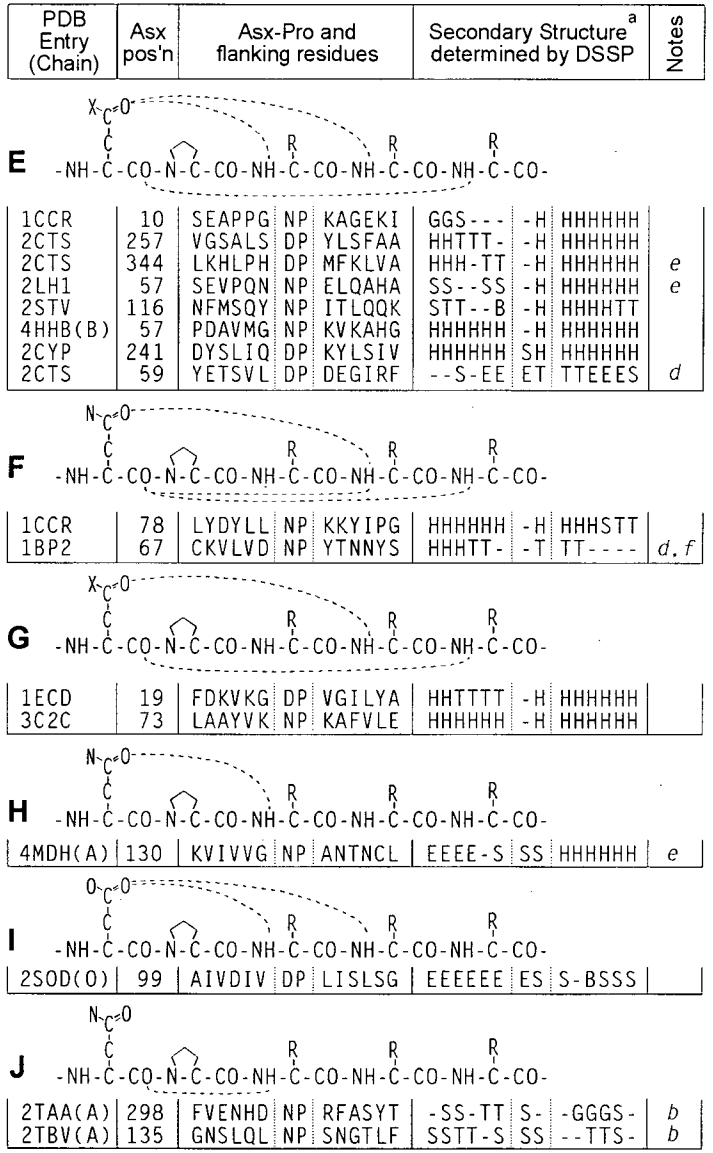
The program DSSP identifies backbone H-bonds using a simple energy model that considers both the distance between oxygen and nitrogen atoms in backbone C=O and NH groups and the deviation from an optimal (for H-bond formation) angle between these groups. Using a relatively low cutoff value of -0.5 kcal mol^{-1} for binding energy (a good H-bond is about -3 kcal mol^{-1}), DSSP allows C=O and NH groups to be misaligned up to 63° at the ideal (2.9 Å) O-N distance, and allows O-N distances of up to 5.2 Å for perfect alignment. H-bonds identified by the program DSSP were used in the analysis presented in Table 3-I and Figure 3-3, and for identification of backbone-backbone H-bonds in analyses that also included sidechain-backbone H-bonds (Figures 3-1, 3-2, 3-4, 3-5, 3-10 and 3-11). Because DSSP identifies only backbone H-bonds, a separate criterium was employed to identify sidechain-backbone H-bonds: a sidechain-backbone H-bond was considered to exist if the interatomic distance between a sidechain donor/acceptor heavy atom (O or N) and a backbone acceptor/donor heavy atom was $\leq$3.5 Å.

This mixture of criteria appears to have been adequate and seems warranted by five



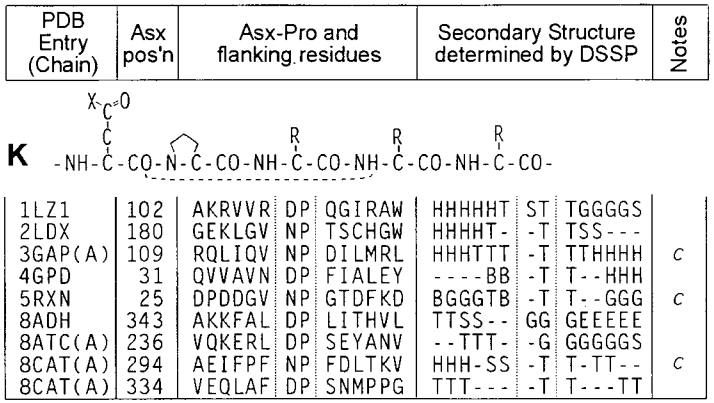
PDB Entry (Chain)	Asx pos'n	Asx-Pro and flanking residues	Secondary Structure[a] determined by DSSP	Notes
A				
1ACX	44	GGQDAC NP ATATSF	ETEE-- -T TT--EE	*d,e*
B				
1FX1	129	GLRIDG DP RAARDD	-EEEES -G GGGHHH	*e*
1GP1(A)	142	ATALMT DP KFITWS	SS---S SG GG--SS	
2APR	59	SGQTKY DP NQSSTY	TTS--B -G GG-TT-	
2CAB	52	PISVSY NP ATAKEI	EEEEE- -G GGEEEE	
2CAB	154	MKVGEA NP KLQKVL	EEESS- -G GGHHHH	
2CAB	180	APFTNF DP STLLPS	EE--S- -G GGGS-S	
2CYP	79	FKKEFN DP SNAGLQ	SHHHHT -G GGTTTH	
3TLN	207	SLRSMS DP AKYGDP	-SEESS -G GGGT--	*e*
8ABP	67	FVICTP DP KLGSAI	EEEE-S -G GGHHHH	
8ADH	242	GATECV NP QDYKKP	T-SEEE -G GG-SS-	
8CAT(A)	5	NR DP ASDQMK	-- -G GGGHHH	*e*
8ADH	304	SQNLSM NP MLLLSG	------ -T HHHHTT	
4HHB(A)	94	AHKLRV DP VNFKLL	HHTT-- -T HHHHHH	
4HHB(B)	99	CDKLHV DP ENFRLL	HHTT-- -T HHHHHH	
8CAT(A)	256	ARLAHE DP DYGLRD	HHHHHH -T THHHHH	*e*
3C2C	84	VLEKSG DP KAKSKM	HHHHH- -T T---S-	
3ICB	19	YAAKEG DP NQLSKE	HHTSSS ST TSB-HH	*c*
7API	390	FMGKVV NP TQKC	EEEEES -T T---	*c*
C				
8ATC(A)	2	A NP LYQKHI	- -T TTT--B	*e,f*
D				
1ECD	31	YAVFKA DP SIMAKF	HHHHHH -H HHHTT-	*d*
2CTS	327	AVLRKT DP RYTCQR	SS--S- -H HHHHHH	
451C	2	E DP EVLFKN	- -H HHHHHH	
8CAT(A)	177	PQTHLK DP DMVWDF	TTT-SB -H HHHHHH	*d*
5RXN	14	VCGYIY DP EDGDPD	TT--EE -T TT-BGG	

Fig. 3-1. continued on following page



PDB Entry (Chain)	Asx pos'n	Asx-Pro and flanking residues	Secondary Structure[a] determined by DSSP	Notes
E				
1CCR	10	SEAPPG NP KAGEKI	GGS--- -H HHHHHH	
2CTS	257	VGSALS DP YLSFAA	HHTTT- -H HHHHHH	
2CTS	344	LKHLPH DP MFKLVA	HHH-TT -H HHHHHH	*e*
2LH1	57	SEVPQN NP ELQAHA	SS--SS -H HHHHHH	*e*
2STV	116	NFMSQY NP ITLQQK	STT--B -H HHHHTT	
4HHB(B)	57	PDAVMG NP KVKAHG	HHHHHH -H HHHHHH	
2CYP	241	DYSLIQ DP KYLSIV	HHHHHH SH HHHHHH	
2CTS	59	YETSVL DP DEGIRF	--S-EE ET TTEEES	*d*
F				
1CCR	78	LYDYLL NP KKYIPG	HHHHHH -H HHHSTT	
1BP2	67	CKVLVD NP YTNNYS	HHHTT- -T TT----	*d,f*
G				
1ECD	19	FDKVKG DP VGILYA	HHTTTT -H HHHHHH	
3C2C	73	LAAYVK NP KAFVLE	HHHHHH -H HHHHHH	
H				
4MDH(A)	130	KVIVVG NP ANTNCL	EEEE-S SS HHHHHH	*e*
I				
2SOD(O)	99	AIVDIV DP LISLSG	EEEEEE ES S-BSSS	
J				
2TAA(A)	298	FVENHD NP RFASYT	-SS-TT S- -GGGS-	*b*
2TBV(A)	135	GNSLQL NP SNGTLF	SSTT-S SS --TTS-	*b*

Fig. 3-1 continued from previous page.



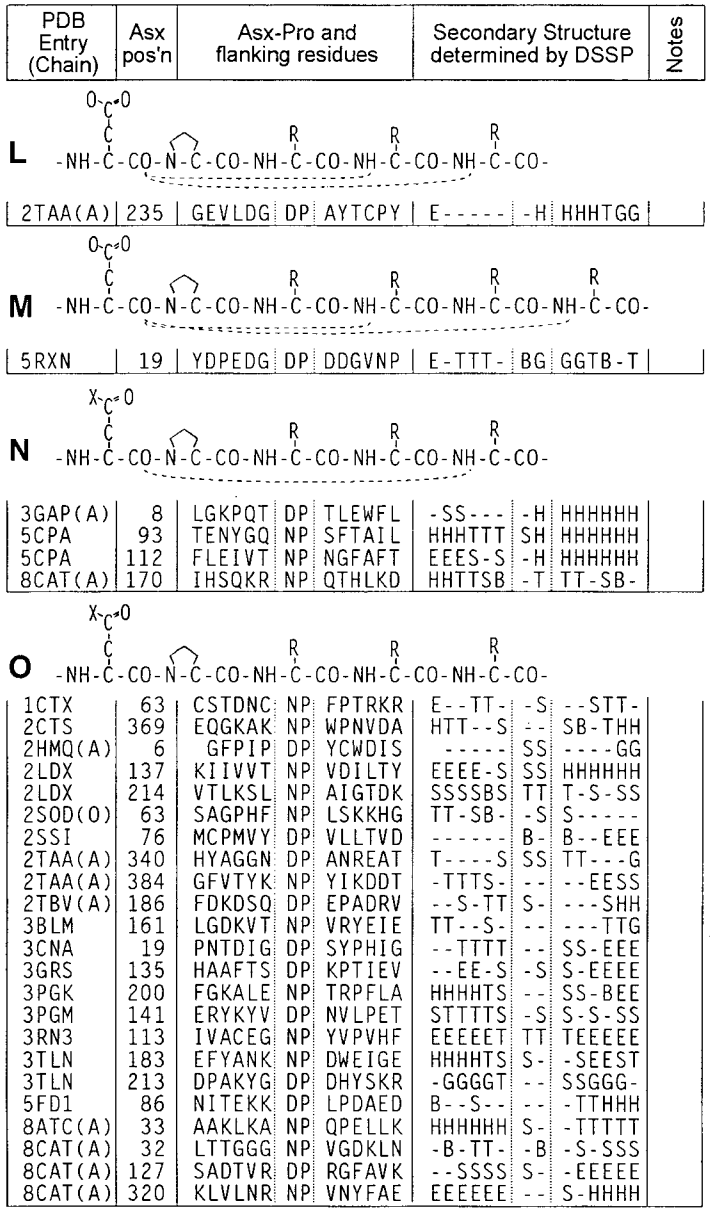
PDB Entry (Chain)	Asx pos'n	Asx-Pro and flanking residues	Secondary Structure determined by DSSP	Notes
1LZ1	102	AKRVVR DP QGIRAW	HHHHHT ST TGGGGS	
2LDX	180	GEKLGV NP TSCHGW	HHHHT- -T TSS---	
3GAP(A)	109	RQLIQV NP DILMRL	HHHTTT -T TTHHHH	c
4GPD	31	QVVAVN DP FIALEY	----BB -T T--HHH	
5RXN	25	DPDDGV NP GTDFKD	BGGGTB -T T--GGG	c
8ADH	343	AKKFAL DP LITHVL	TTSS-- GG GEEEEE	
8ATC(A)	236	VQKERL DP SEYANV	--TTT- -G GGGGGS	
8CAT(A)	294	AEIFPF NP FDLTKV	HHH-SS -T T-TT--	c
8CAT(A)	334	VEQLAF DP SNMPPG	TTT--- -T T---TT	

Fig. 3-1. Local H-bonds in Asx-Pro sequences, Dataset A. Classification of local sidechain-backbone and backbone-backbone H-bonds formed by 79 Asx-Pro sequences identified in the 13,900 residues of dataset A. These Asx-Pro sequences correspond to the bars shown in Figure 3-3(b) for Asp-Pro (42 sequences) and Asn-Pro (37 sequences), as well as to Asx-Pro data presented in Figures 3-4 to 3-6. Specifically, groups A to G correspond to the 47 per cent of (37 of 79) Asx-Pro sequences that form both sidechain-backbone and backbone-backbone H-bonds, summarized in Figure 3-4; groups H and I, to the 3 per cent (2 of 79) that form only sidechain-backbone H-bonds; groups J to N, to the 22 per cent (17 of 79) that form only backbone-backbone H-bonds; and group O, to the 29 per cent (23 of 79) that form no local H-bonds. Also, groups B to G correspond to the 36 Asx-Pro residues that form $O^{\delta}(i)\leftarrow C{=}O(i+n_{n=2\text{ or }3})$ sidechain-backbone and $C{=}O(i)\leftarrow NH(i+n_{n=3\text{ or }4})$ backbone-backbone H-bonds shown in Figures 3-5 (part B, dataset A) and 3-6.

[a]Structure codes assigned by DSSP, as follows: G, 3_{10}-helix; h, α-helix; S, bend; T, reverse turn; E, ß-strand; B, residue hydrogen bonding to a ß-strand; -, no secondary structure.

[b,c,d]The indicated $i\leftarrow i+2$ (note *b*), $i\leftarrow i+3$ (note *c*) or $i\leftarrow i+4$ (note *d*) backbone-backbone H-bond does not meet all criteria of WHATIF, but meets criteria of DSSP.

[e,f]The indicated $i\leftarrow i+2$ (note *e*) or $i\leftarrow i+3$ (note *f*) sidechain-backbone H-bond does not meet all criteria of WHATIF, but meets the O-N distance criterium.



PDB Entry (Chain)	Asx pos'n	Asx-Pro and flanking residues	Secondary Structure determined by DSSP	Notes
2TAA(A)	235	GEVLDG DP AYTCPY	E----- -H HHHTGG	
5RXN	19	YDPEDG DP DDGVNP	E-TTT- BG GGTB-T	
3GAP(A)	8	LGKPQT DP TLEWFL	-SS--- -H HHHHHH	
5CPA	93	TENYGQ NP SFTAIL	HHHTTT SH HHHHHH	
5CPA	112	FLEIVT NP NGFAFT	EEES-S -H HHHHHH	
8CAT(A)	170	IHSQKR NP QTHLKD	HHTTSB -T TT-SB-	
1CTX	63	CSTDNC NP FPTRKR	E--TT- -S --STT-	
2CTS	369	EQGKAK NP WPNVDA	HTT--S -- SB-THH	
2HMQ(A)	6	GFPIP DP YCWDIS	----- SS ----GG	
2LDX	137	KIIVVT NP VDILTY	EEEE-S SS HHHHHH	
2LDX	214	VTLKSL NP AIGTDK	SSSSBS TT T-S-SS	
2SOD(O)	63	SAGPHF NP LSKKHG	TT-SB- -S S-----	
2SSI	76	MCPMVY DP VLLTVD	------ B- B--EEE	
2TAA(A)	340	HYAGGN DP ANREAT	T----S SS TT---G	
2TAA(A)	384	GFVTYK NP YIKDDT	-TTTS- -- --EESS	
2TBV(A)	186	FDKDSQ DP EPADRV	--S-TT S- ---SHH	
3BLM	161	LGDKVT NP VRYEIE	TT--S- -- ---TTG	
3CNA	19	PNTDIG DP SYPHIG	--TTTT -- SS-EEE	
3GRS	135	HAAFTS DP KPTIEV	--EE-S -S S-EEEE	
3PGK	200	FGKALE NP TRPFLA	HHHHTS -- SS-BEE	
3PGM	141	ERYKYV DP NVLPET	STTTTS -S S-S-SS	
3RN3	113	IVACEG NP YVPVHF	EEEEET TT TEEEEE	
3TLN	183	EFYANK NP DWEIGE	HHHHTS S- -SEEST	
3TLN	213	DPAKYG DP DHYSKR	-GGGGT -- SSGGG-	
5FD1	86	NITEKK DP LPDAED	B--S-- -- -TTHHH	
8ATC(A)	33	AAKLKA NP QPELLK	HHHHHH S- -TTTTT	
8CAT(A)	32	LTTGGG NP VGDKLN	-B-TT- -B -S-SSS	
8CAT(A)	127	SADTVR DP RGFAVK	--SSSS S- -EEEEE	
8CAT(A)	320	KLVLNR NP VNYFAE	EEEEEE -- S-HHHH	

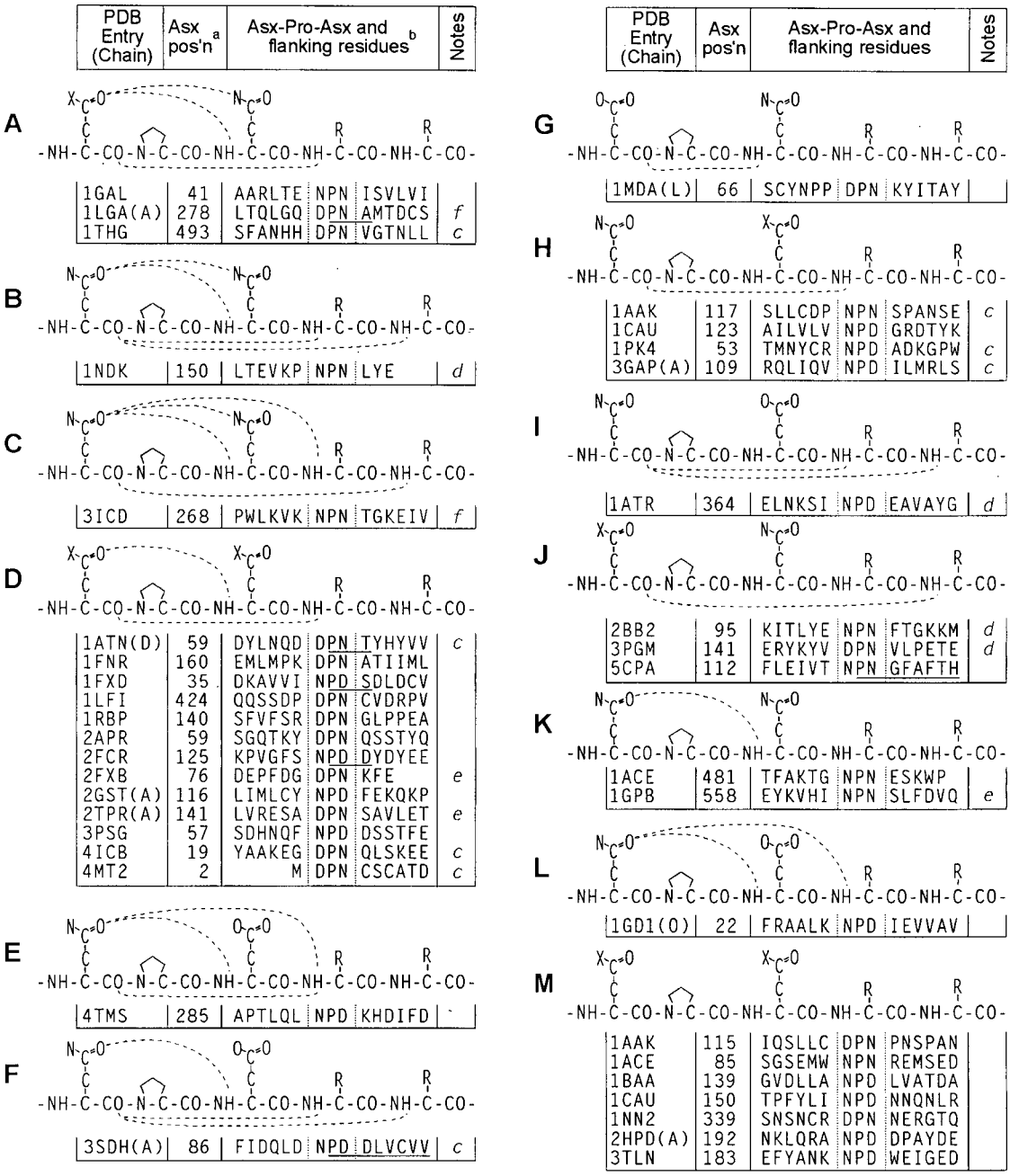


Fig. 3-2. Local H-bonds in Asx-Pro-Asx sequences, Dataset C. Classification of local sidechain-backbone and backbone-backbone H-bonds formed in 39 Asx-Pro-Asx sequences identified in Release 14 of the NRL3D database. Groups A to F correspond to the 51 per cent of (20 of 39) Asx-Pro-Asx sequences that form both sidechain-backbone and backbone-backbone H-bonds, summarized in Figure 3-10, part A; groups G to J, to the 23 per cent (9 of 39) that form only backbone-backbone H-bonds; groups K and L, to the 8 per cent (3 of 39) that form only sidechain-backbone H-bonds; and group M, to the 18 per cent (7 of 39) that form no local H-bonds. Superimposed wireframe models of groups A to F are shown in Figure 3-11(parts A and B); groups G to M in Figure 3-11(C and D).

[a]Position of Asx_i in the sequence Asx_i-Pro-Asx_j. [b]Underlined residues are within 3_{10}- or α-helices [c,d]The indicated $i \leftarrow i+2$ (note *c*) or $i \leftarrow i+3$ (note *d*) backbone-backbone H-bonds does not meet all criteria of WHATIF, but meets criteria of DSSP. [e,f]The indicated sidechain-backbone (note *e*) or sidechain-sidechain (note *f*) H-bond does not meet all criteria of WHATIF, but meets theinteratomic distance criterium.

considerations. (i) H-bond geometry is quite variable (McGregor *et al.* 1987; Ippolito *et al.* 1990; Jeffrey & Saenger 1991; Stickle *et al.* 1992). (ii) Possible biases in protein crystal structure refinement procedures such as the more accurate location of backbone *versus* side chain atoms (Richardson 1981) and stricter constraints for covalent *versus* H-bonds, as well as the relatively low absolute resolution of most structures, suggest that stricter criteria should not be applied, in particular to the identification of sidechain-backbone H-bonds. (iii) The program WHATIF was employed for an independent validation of the mixed criteria. Importantly, most H-bonds identified by DSSP or by the simple distance criterium were also identified by WHATIF, which considers H-bond geometry; differences have been noted in the detailed analyses presented in Figures 3-1 and 3-2. (iv) For Asx-Pro sequences in dataset A, C=O*(i)*←NH$(i+n_{n=3 \text{ or } 4})$ backbone H-bonds identified by DSSP (Figure 3-1) were compared with those identified by the simple length criterium. For 61 of the 79 sequences, the methods agreed; disagreement in 14 of the remaining 18 sequences concerned whether an Asx C=O*(i)* group formed H-bonds with both *versus* only one of NH*(i+3)* and NH*(i+4)*. Importantly, discrepancies between the two methods would not have affected the identification of 36 Asx-Pro sequences as "Asx-Pro turns." (v) Importantly, superimposed wireframe models of structures identified as "Asx-Pro turns" (Figures 3-6 and 3-11) strongly suggest that the criteria employed for H-bond identification were reasonable.

Nitrogen and oxygen atoms are not readily distinguished by protein X-ray crystallography techniques and these two atoms are usually assigned on the basis of hydrogen bonding. Considering this, and that McDonald and Thornton (1995) have found that 15 per cent of Asn sidechains would be more favourably oriented for hydrogen bonding if the nitrogen and oxygen designations were reversed, the estimates of H-bonds involving Asn sidechains would seem conservative.

3.4. RESULTS

Many elements of secondary structure can be defined by their hydrogen bonding patterns. Thus the DSSP (Kabsch & Sander 1983) algorithm defines turns and helices in terms of local H-bonds between backbone C=O and NH groups of residues i and $i+n$ ($3 \leq n \leq 5$) respectively. A single such $i \leftarrow i+n$ H-bond defines a turn (n-turn) while repeating turns define 3_{10}-, α- and π-helices as a series of 3-, 4- and 5-turns. This definition allows that statistical analyses of differences between local H-bonds formed by residues preceding a proline *versus* a non-proline can provide convenient and appropriate measures of proline's influence on local structure. Such an analysis was performed for two independently-selected sets of proteins, datasets A (80 proteins) and B (102 proteins), each broadly representative (Boberg *et al.* 1992; Adzhubei & Sternberg 1993) of all protein structural classes. To avoid bias, two datasets were employed; since the separate analyses yielded similar results, only parenthetical reference has been made to the results for database B.

3.4.1. Residues not preceding proline favor $i \leftarrow i+4$ H-bonds; proline favors $i \leftarrow i+3$ H-bonds

When averaged across all amino acids, residues preceding a non-proline favor $i \leftarrow i+4$ over other local ($i \leftarrow i+n_{n=2,3,5}$) backbone H-bonds (Table 3-Ia). Thus, the 23.4 per cent (25.0%, dataset B) of residues forming $i \leftarrow i+4$ H-bonds is 2.9-fold greater than the 8.2 per cent (8.3%, dataset B) forming $i \leftarrow i+3$ H-bonds, the next favoured conformation. The importance of $i \leftarrow i+4$ H-bonds to protein structure is apparent: they account for 39.6 per cent (41.2%, dataset B) of all (local and non-local) backbone H-bonds. The importance of local H-bonds is similarly evident: 63.5 per cent (64.7%, dataset B) of backbone H-bonds are local, a finding consistent with a previous survey (Stickle *et al.* 1992).

In contrast, residues preceding proline favor $i \leftarrow i+3$ over other local backbone H-bonds (Table 3-Ib): the 21.1 per cent (20.0%, dataset B) of residues forming $i \leftarrow i+3$ H-bonds is 1.5-

Table 3-I. Effect[a] of proline on backbone hydrogen bonding of preceding residue.

	Number of residues forming C=O*(i)*←NH*(i+n)* backbone H-bonds[b], per 100 residues[c]		Per cent of total number of residues forming H-bonds	
	Dataset A	Dataset B	Dataset A	Dataset B
(a) Residue *i*+1 is not proline				
Local H-bonds				
C=O*(i)*←NH*(i+2)*	4.9	4.8	8.3	7.9
C=O*(i)*←NH*(i+3)*	**8.2**	8.3	13.9	13.7
C=O*(i)*←NH*(i+4)*	**23.4**	25.0	**39.6**	41.2
C=O*(i)*←NH*(i+5)*	1.0	1.1	1.7	1.8
Local H-bonds	37.5	39.2	**63.5**	64.7
Non-local H-bonds				
C=O*(i)*←NH*(i+n)*, $n>5$ or $n<0$	21.6	21.4	36.5	35.3
Local + Non-local H-bonds	59.1	60.6	100.0	100.0
(b) Residue *i*+1 is proline				
Local H-bonds				
C=O*(i)*←NH*(i+2)*	4.7	4.3	8.3	7.6
C=O*(i)*←NH*(i+3)*	**21.1**	20.0	37.2	35.4
C=O*(i)*←NH*(i+4)*	**14.2**	15.0	25.0	26.5
C=O*(i)*←NH*(i+5)*	0.2	0.2	0.4	0.4
Local H-bonds	40.2	39.5	70.9	69.9
Non-local H-bonds				
C=O*(i)*←NH*(i+n)*, $n>5$ or $n<0$	16.5	17.0	29.1	30.1
Local + Non-local H-bonds	56.7	56.5	100.0	100.0

[a]Table provides an overview of the effect of proline on local backbone hydrogen bonding by allowing comparison of the fractions of residues at position *i* that form C=O*(i)*←NH$(i+n_{2\le n\le 5})$ backbone H-bonds when position *i*+1 is occupied by (i) non-proline or (ii) proline residues. Numbers in **bold** are referred to specifically in the text of the Results section.

[b]For simplicity, only a single H-bond (the strongest, according to the program DSSP) was counted for each C=O group. Because few C=O groups form multiple H-bonds and similar results (not shown) were thus obtained when up to two H-bonds were counted for each C=O group, I consider that these statistics give an adequate measure of backbone H-bonds.

[c]Summary is based on analysis of (a) 13,900 (dataset A) and 21,395 (dataset B) residues preceding a non-proline residue, and (b) 660 (dataset A) and 1,065 (dataset B) residues preceding proline.

figure shown on following page

Fig. 3-3. Effect of proline on backbone hydrogen bonding. This figure extends the analysis of Table 3-I by allowing comparison -- for individual residues or groups of residues -- of the fractions of residues at position *i* that form C=O*(i)*←NH$(i+n_{2\le n\le 5})$ backbone H-bonds when position $i+1$ is occupied by (i) non-proline or (ii) proline residues. Tabular data in (c) -- the numeric difference between values presented graphically in (a) and (b) -- provides a measure of the "gain" or "loss" of local backbone H-bonds when residues, at position *i*, precede proline *versus* non-proline residues at position $i+1$. A dashed vertical line has been placed arbitrarily at position 40 of the horizontal axis to facilitate comparison of bars in (a) and (b).

Certain residues have been grouped to reduce the volume of data with minimal loss of information. The fractions of each of A, V, I, L, P, M and W that form local H-bonds are smaller when these residues precede proline *versus* non-proline residues; small increases occur in the corresponding fractions for residues C and Y. For residue F, decreases or relatively small increases occur in these fractions in datasets A and B, respectively.

[a]I emphasize that H-bonds formed by sidechains are not enumerated in this figure. The column "Polar atom in sidechain" was included to illustrate the idea that, as discussed in the text, there is a correlation between the position of polar atoms in sidechains and the formation of backbone-backbone H-bonds.

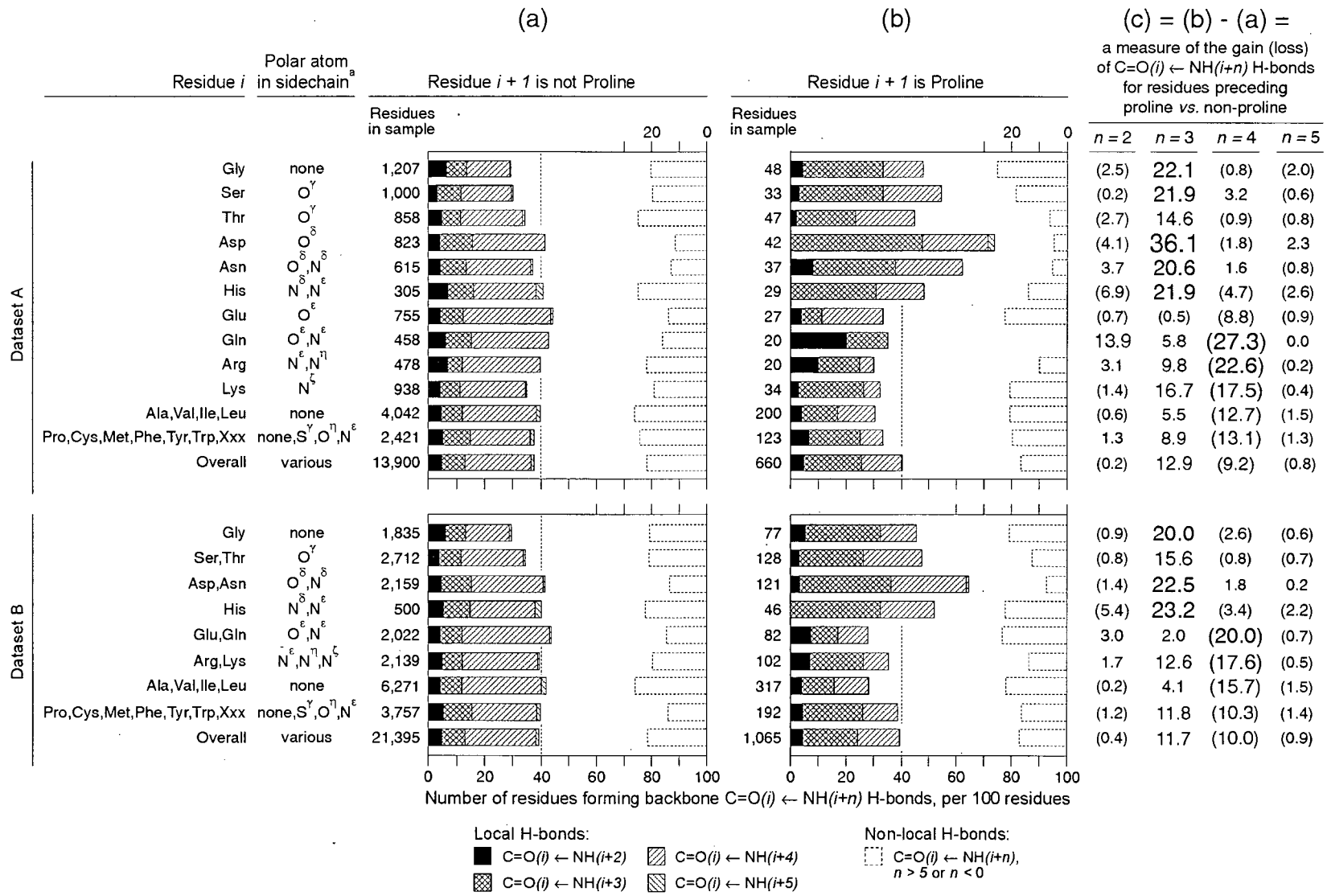

(a)
(b)
(c) = (b) - (a) =
a measure of the gain (loss) of C=O(i) ← NH(i+n) H-bonds for residues preceding proline vs. non-proline
Residue i
Polar atom in sidechain[a]
Residue i + 1 is not Proline
Residue i + 1 is Proline
Residues in sample
n = 2
n = 3
n = 4
n = 5
Dataset A
Gly none 1,207 48 (2.5) 22.1 (0.8) (2.0)
Ser Oγ 1,000 33 (0.2) 21.9 3.2 (0.6)
Thr Oγ 858 47 (2.7) 14.6 (0.9) (0.8)
Asp Oδ 823 42 (4.1) 36.1 (1.8) 2.3
Asn Oδ,Nδ 615 37 3.7 20.6 1.6 (0.8)
His Nδ,Nε 305 29 (6.9) 21.9 (4.7) (2.6)
Glu Oε 755 27 (0.7) (0.5) (8.8) (0.9)
Gln Oε,Nε 458 20 13.9 5.8 (27.3) 0.0
Arg Nε,Nη 478 20 3.1 9.8 (22.6) (0.2)
Lys Nζ 938 34 (1.4) 16.7 (17.5) (0.4)
Ala,Val,Ile,Leu none 4,042 200 (0.6) 5.5 (12.7) (1.5)
Pro,Cys,Met,Phe,Tyr,Trp,Xxx none,Sγ,Oη,Nε 2,421 123 1.3 8.9 (13.1) (1.3)
Overall various 13,900 660 (0.2) 12.9 (9.2) (0.8)
Dataset B
Gly none 1,835 77 (0.9) 20.0 (2.6) (0.6)
Ser,Thr Oγ 2,712 128 (0.8) 15.6 (0.8) (0.7)
Asp,Asn Oδ,Nδ 2,159 121 (1.4) 22.5 1.8 0.2
His Nδ,Nε 500 46 (5.4) 23.2 (3.4) (2.2)
Glu,Gln Oε,Nε 2,022 82 3.0 2.0 (20.0) (0.7)
Arg,Lys Nε,Nη,Nζ 2,139 102 1.7 12.6 (17.6) (0.5)
Ala,Val,Ile,Leu none 6,271 317 (0.2) 4.1 (15.7) (1.5)
Pro,Cys,Met,Phe,Tyr,Trp,Xxx none,Sγ,Oη,Nε 3,757 192 (1.2) 11.8 (10.3) (1.4)
Overall various 21,395 1,065 (0.4) 11.7 (10.0) (0.9)
0 20 40 60 80 100
Number of residues forming backbone C=O(i) ← NH(i+n) H-bonds, per 100 residues
Local H-bonds:
C=O(i) ← NH(i+2)
C=O(i) ← NH(i+3)
C=O(i) ← NH(i+4)
C=O(i) ← NH(i+5)
Non-local H-bonds:
C=O(i) ← NH(i+n), n > 5 or n < 0

Fig. 3-3.

fold greater than the 14.2 per cent (15.0%, dataset B) forming $i \leftarrow i+4$ H-bonds.

These findings hold true for individual residues: for each of glycine, nine polar residues and for groupings of other residues, $i \leftarrow i+4$ backbone H-bonds predominate over $i \leftarrow i+3$ H-bonds by a factor of 2.1:1 (Ser) to 5.3:1 (Arg) when these residues precede a non-proline (Figure 3-3a). Similarly, for most residues (Glu being a notable exception), $i \leftarrow i+3$ H-bonds predominate over $i \leftarrow i+4$ H-bonds by factors as high as 2.2:1 (Asp), 3:1 (Arg) and 4:1 (Lys) when these residues precede proline (Figure 3-3b).

3.4.2. Short polar sidechains in residues preceding proline favor $i \leftarrow i+3$ H-bonds

The effect of proline in favoring $i \leftarrow i+3$ backbone H-bonds is most pronounced for residues with short polar sidechains and for glycine (Figure 3-3b and c). Thus, the fractions of serine, threonine, aspartate, asparagine, histidine and (anomalously) glycine forming $i \leftarrow i+3$ H-bonds increase by 14.6 (Thr) to 36.1 (Asp) per cent (dataset A; Figure 3-3c) when preceding proline *versus* non-proline, while the fractions forming $i \leftarrow i+4$ H-bonds are reduced by only 0.9 (Thr) to 4.7 (His) per cent. Overall "gains" for local H-bonds are greatest for aspartate and asparagine: 73.8 and 62.2 per cent, respectively, of these residues form $i \leftarrow i+n_{2 \leq n \leq 5}$ H-bonds when preceding proline (dataset A, Figure 3-3b).

In contrast, residues with polar atoms more distant from the α-carbon tend to score only modest gains in those fractions forming $i \leftarrow i+3$ backbone H-bonds and to score "losses" of $i \leftarrow i+4$ H-bonds. For example, the fraction of glutamine residues forming $i \leftarrow i+3$ H-bonds increases only 5.8 per cent when glutamine precedes proline *versus* non-proline, while the fractions of glutamate, glutamine, arginine and lysine forming $i \leftarrow i+4$ H-bonds decrease 8.8 to 27.3 per cent when these residues precede proline *versus* non-proline (dataset A, Figure 3-3c). A similar trend occurs with other residues (A, V, I, L, P, M, F and W in one-letter code; detailed data not shown, see Figure 3-3 for summary data). Moreover, the total number of local H-bonds for this heterogenous group of residues is smaller when these residues precede

proline *versus* non-proline. Small increases in local H-bonds occur for cysteine and tyrosine (data not shown).

3.4.3. Asx-Pro favors a combination of sidechain-backbone and backbone-backbone H-bonds

Since the greatest propensity to form local backbone H-bonds occurs with aspartate, asparagine, serine and histidine preceding proline, the analysis was extended to include potentially turn-stabilizing $i \leftarrow i+n_{2\leq n\leq 5}$ H-bonds from the polar sidechains of these residues to the peptide backbone. Significantly, Asx-Pro sequences were found to have a greater tendency to form combined sidechain-backbone and backbone-backbone H-bonds than do Ser-Pro and His-Pro sequences (Figure 3-4). Thus, 68 per cent (64%, dataset B) of Asx-Pro peptides form C=O$(i)\leftarrow$NH$(i+n_{2\leq n\leq 5})$ backbone H-bonds, 49 per cent (52%, dataset B) form $O^{\delta}(i)\leftarrow$NH$(i+n_{n=2,3})$ sidechain-backbone H-bonds and 47 percent (46%, dataset B) form these H-bonds in combination. In contrast, and although 61 per cent (55%, dataset B) of Ser-Pro peptides form C=O$(i)\leftarrow$NH$(i+n_{2\leq n\leq 5})$ backbone H-bonds, only 15 per cent (18%, dataset B) form both sidechain-backbone and backbone-backbone H-bonds. Moreover, those sidechain-backbone H-bonds that are formed are more diverse, involving backbone NH and C=O groups, reflecting alternative roles of serine's polar sidechain as H-bond donor or acceptor. Similarly, only 21 per cent (15%, dataset B) of His-Pro peptides form a combination of complementary (and diverse) sidechain-backbone and backbone-backbone H-bonds.

3.4.4. Similar structures are formed by a variety of Asx-Pro H-bond conformations

The 47 per cent of Asx-Pro sequences (46%, dataset B) that form complementary sidechain-backbone and backbone-backbone H-bonds represents 37 of the 79 Asx-Pro sequences of dataset A (56 of 121, dataset B). Importantly, 36 of these 37 sequences (55 of 56, dataset B), form turn structures stabilized by a limited set of H-bonds among the sidechain O^{δ} and backbone C=O carbonyl oxygens of Asx(i) and the backbone NH of residues i+2, i+3

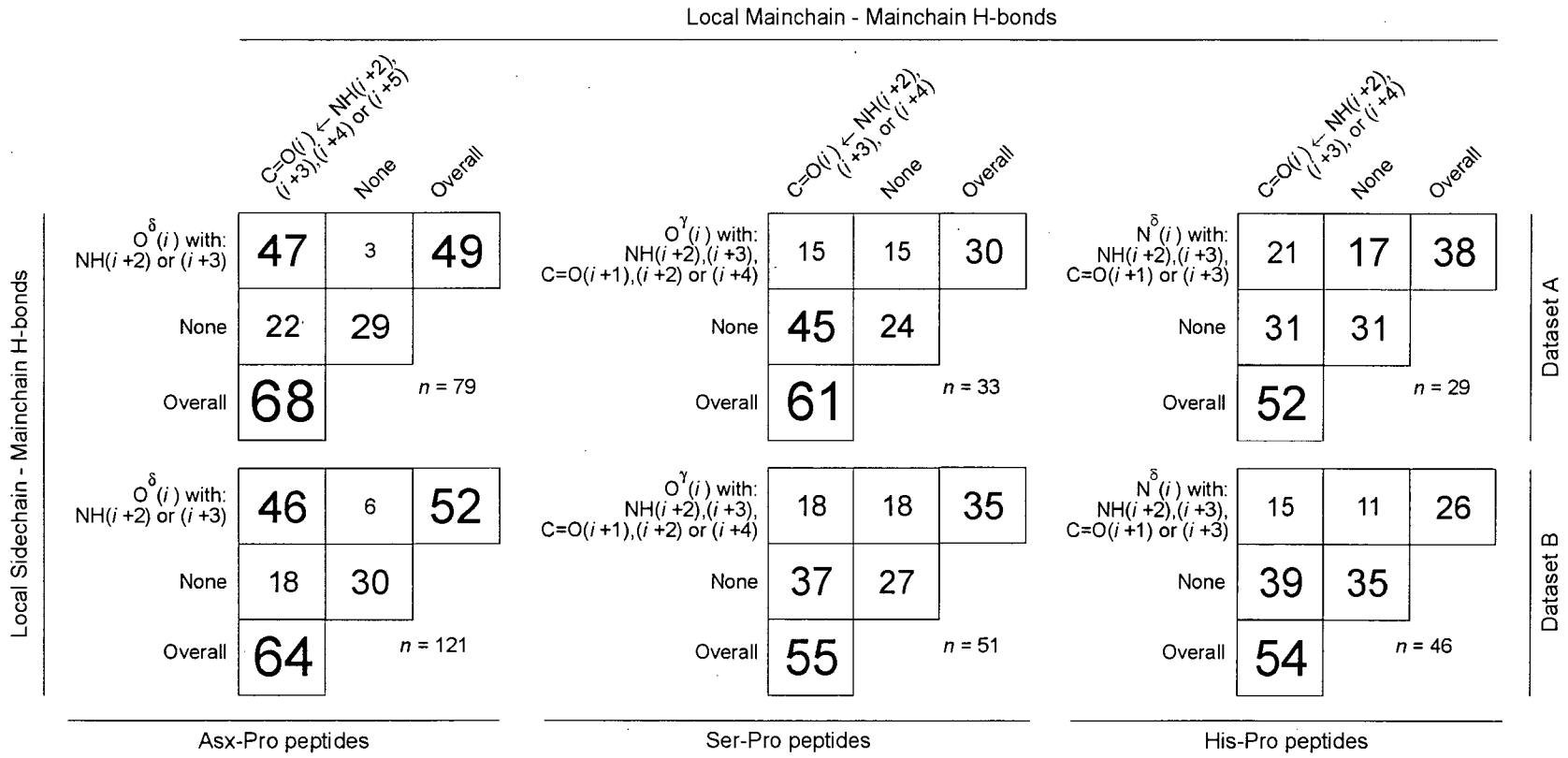
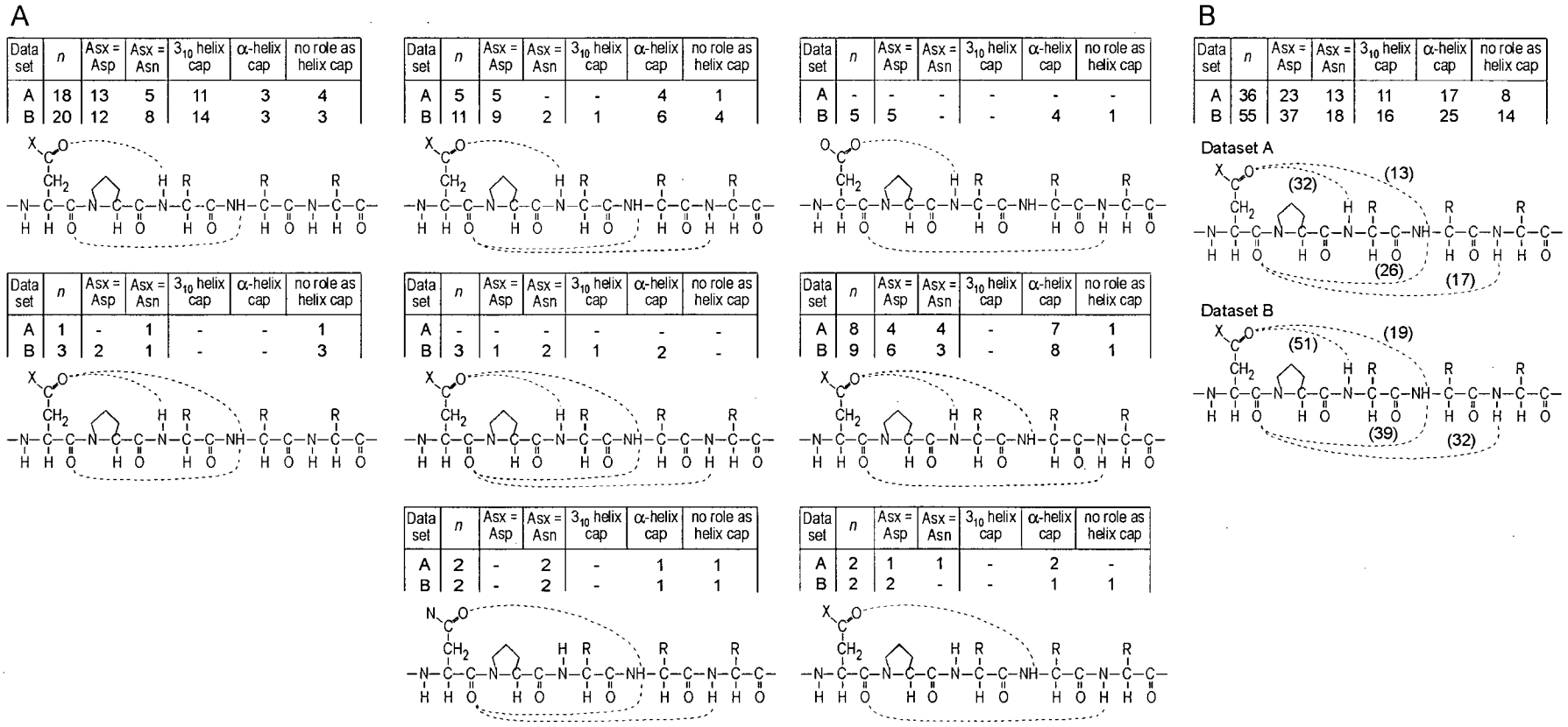


Fig. 3-4. Local H-bonds in Asx-Pro, Ser-Pro and His-Pro sequences. Each matrix shows, per 100 residues, the number of residues that form (i) local sidechain-backbone and backbone-backbone H-bonds in combination, (ii) sidechain-backbone H-bonds only, (iii) backbone-backbone H-bonds only, or (iv) no local H-bonds. All Asx-Pro, Ser-Pro and His-Pro sequences summarized in Figure 3-3 are included, except for one Ser-Pro sequence where poor local structure resolution precluded complete H-bond analysis.



Data set	*n*	Asx = Asp	Asx = Asn	3_{10} helix cap	α-helix cap	no role as helix cap
A	18	13	5	11	3	4
B	20	12	8	14	3	3

Data set	*n*	Asx = Asp	Asx = Asn	3_{10} helix cap	α-helix cap	no role as helix cap
A	5	5	-	-	4	1
B	11	9	2	1	6	4

Data set	*n*	Asx = Asp	Asx = Asn	3_{10} helix cap	α-helix cap	no role as helix cap
A	-	-	-	-	-	-
B	5	5	-	-	4	1

Data set	*n*	Asx = Asp	Asx = Asn	3_{10} helix cap	α-helix cap	no role as helix cap
A	1	-	1	-	-	1
B	3	2	1	-	-	3

Data set	*n*	Asx = Asp	Asx = Asn	3_{10} helix cap	α-helix cap	no role as helix cap
A	-	-	-	-	-	-
B	3	1	2	1	2	-

Data set	*n*	Asx = Asp	Asx = Asn	3_{10} helix cap	α-helix cap	no role as helix cap
A	8	4	4	-	7	1
B	9	6	3	-	8	1

Data set	*n*	Asx = Asp	Asx = Asn	3_{10} helix cap	α-helix cap	no role as helix cap
A	2	-	2	-	1	1
B	2	-	2	-	1	1

Data set	*n*	Asx = Asp	Asx = Asn	3_{10} helix cap	α-helix cap	no role as helix cap
A	2	1	1	-	2	-
B	2	2	-	-	1	1

Data set	*n*	Asx = Asp	Asx = Asn	3_{10} helix cap	α-helix cap	no role as helix cap
A	36	23	13	11	17	8
B	55	37	18	16	25	14

Fig. 3-5. Alternative hydrogen bonding patterns in Asx-Pro turns. Each of the eight "Asx-Pro turn" hydrogen bonding patterns summarized in **A** reflects an alternative set of H-bonds formed among the sidechain O^{δ} and backbone C=O of Asx*(i)* and backbone NH groups of residues *i*+2, *i*+3 and *i*+4. The table associated with each pattern indicates the number (*n*) of turns initiated by each of Asp and Asn, as well as the number of turns that function as 3_{10}- or α-helix caps, or as ß-turns. **B** is a summary of **A**. Numbers in parentheses indicate the number of H-bonds formed among the 36 (dataset A) and 55 (dataset B) Asx-Pro turns identified.

and *i*+4 (Figure 3-5B). The most common H-bonds, individually and in combination, are O^{δ}*(i)*←NH*(i+2)* and C=O*(i)*←NH*(i+3)*. Significantly, the eight hydrogen-bonding patterns adopted by Asx-Pro peptides (Figure 3-5A) form virtually indistinguishable structures (Figure 3-6), suggesting that they form a canonical structure stabilized by alternative hydrogen bonding patterns. For want of a better name, this canonical structure has been called an "Asx-Pro turn".

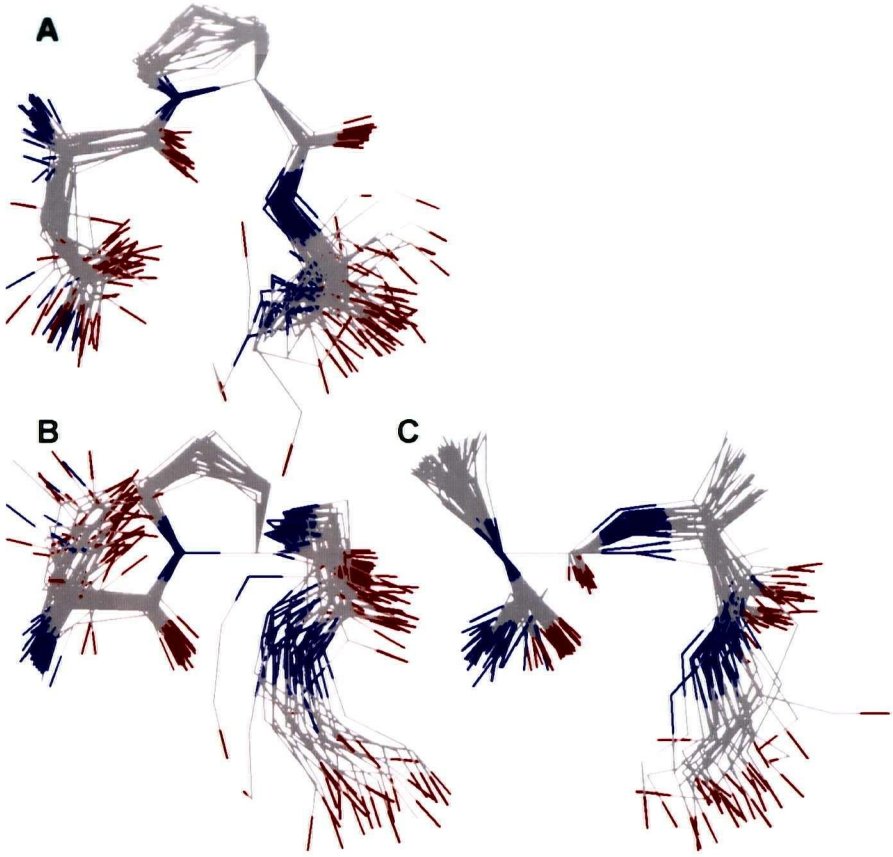


Fig. 3-6. Superimposed wireframe models of Asx-Pro turns. Models were prepared from PDB coordinates of the 36 Asx-Pro-Xxx_1-Xxx_2 turns identified in dataset A and superimposed by hand using general-purpose graphics software. In view **A**, proline N-C^{α} and C^{α}-C bonds lie in the plane of the paper. Rotating the models 90° around the X-axis (away from the viewer) yields view **B** and a subsequent 75° Y-axis rotation (to the viewer's right) yields view **C**. In views **A** and **B**, sidechains of Xxx_1 and Xxx_2 are omitted. In view **C**, the sidechains of Asx and Xxx_1 are omitted and the sidechain of Xxx_2 is truncated at the ß-carbon. Red, oxygen atoms; blue, nitrogen.

3.4.5. Asx-Pro turns have diverse roles

Although most Asx-Pro turns function as 3_{10}- or α-helix caps (Figure 3-5, tabular data),

more than one-fifth of these structures function as ß-turns. No clear association was found between the eight H-bond patterns (Figure 3-5A) and their roles as helix caps or as ß-turns. Similarly, no clear association could be found between Asx-Pro turns and flanking structural environments. Asx-Pro turns are, for example, found between ß-strand and helix structures (Figure 3-1B, 1FX1: position 129), in areas of geometric bends and undefined structure (1GP1(A):142), in a ß-strand/turn/helix/ß-strand configuration (2APR:59) and as a turn connecting two helices (8CAT(A):256).

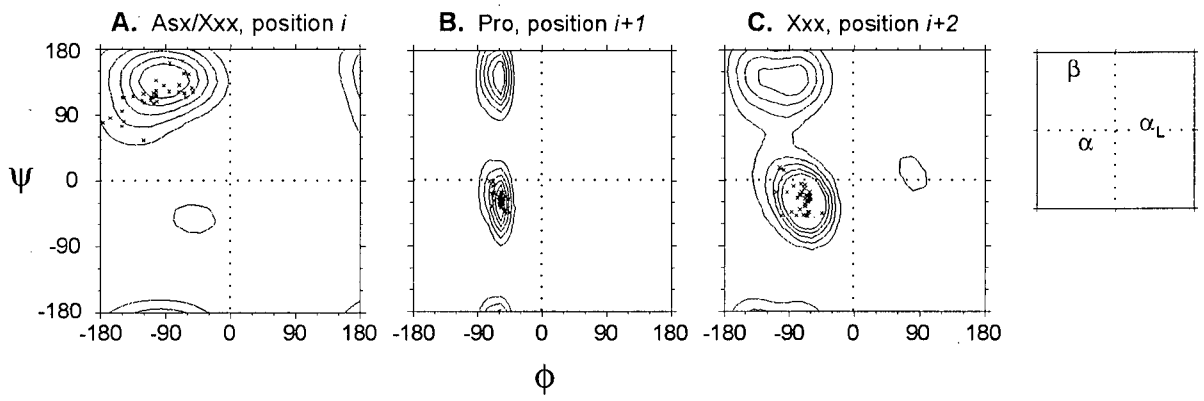


Fig. 3-7. Ramachandran plot of ϕ,ψ dihedral angles in Asx-Pro turns and other proline-containing sequences. "x", individual ϕ,ψ pairs for the 36 Asx-Pro turns identified in dataset A. Contour plots reflect all Xxx-Pro-Xxx sequences of dataset A other than Asx-Pro-Xxx; in each plot, the outermost contour line of each separate area reflects a common low density. Contour plots illustrate the general restriction to ß-conformational space of residues preceding prolines (panel **A**), the narrow range of ϕ angles adopted by prolines (panel **B**) and the restriction of residues generally (whether following proline or not) to α-, ß- and (in some cases) to α_L-conformational space (panel **C**); they reasonably reflect detailed experimental and theoretical plots of permitted ϕ,ψ angles (MacArthur & Thornton 1991; Richardson 1981; Schulz & Schirmer 1979).

3.4.6. Backbone and sidechain geometry of Asx-Pro turns are not unusual

While the geometry of Asx-Pro turns makes intuitive sense, backbone ϕ,ψ angles of residues in Asx-Pro turns were nevertheless compared with those of other proline-containing sequences, and χ_1 angles of Asx in Asx-Pro turns were compared with published values for aspartate and asparagine (McGregor *et al.* 1987), to confirm that Asx-Pro turn geometry is not unusual. Indeed, backbone ϕ,ψ pairs for Asx residues in Asx-Pro turns fall within common

(Schulz & Schirmer 1979; Richardson 1981; MacArthur & Thornton 1991) areas of conformational space (Figure 3-7). As well, most Asx-Pro turns adopt Asx χ_1 angles (Figure 3-8) corresponding to one (the *t* conformation) of two common conformations observed for aspartate and asparagine sidechains (McGregor *et al.* 1987), and all possess Asx χ_1 angles consistent with sidechain conformations observed (Dunbrack & Karplus 1994) for the ϕ,ψ ß-conformation adopted by Asx in these turns.

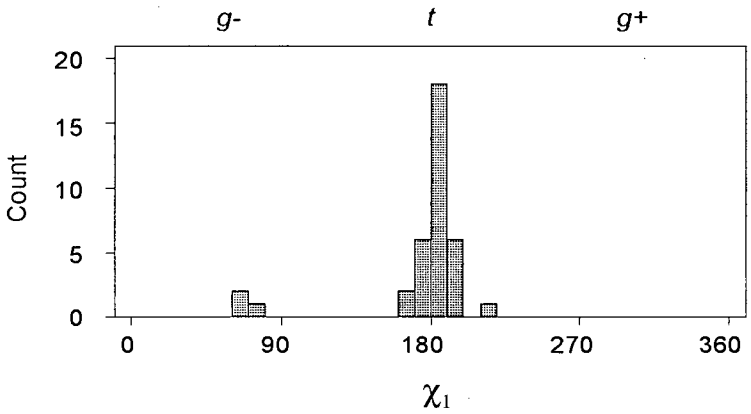


Fig. 3-8. Sidechain χ_1 angles of Asx in Asx-Pro turns. χ_1 angles in the 36 Asx-Pro turns identified in dataset A, converted to the range 0°-360° for comparison with published global distributions of χ_1 angles (McGregor *et al.* 1987). Most angles (n = 33, mean = 185°, s = 9.3°) correspond to the *t* conformation; remaining angles correspond to the *g*- conformation.

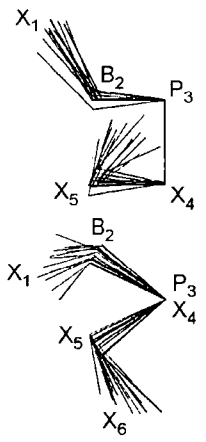


Fig. 3-9. Superimposed α-carbon traces of Asx-Pro helix caps and flanking residues. These views, rotated 90° relative to one another, are of six 3_{10}- and six α-helix caps among 36 Asx-Pro turns (dataset A) that adopt three different hydrogen bonding patterns. X_1, residue immediately preceding helix cap. B_2, Asx, helix capping position. P_3, Pro, first helix position. X_4 to X_6, residues in first helix turn.

A further feature of Asx-Pro turn geometry is noteworthy: although these turns commonly function as helix caps, stabilizing helices in the C-terminal direction, they are helix-terminating in the N-terminal direction (Figure 3-9).

3.4.7. A consensus-derived model for Asn-Pro-Asn agrees with experimental data

Chapter 4 describes studies of the expression, on the surface of filamentous

bacteriophage, of "shaped" peptides containing the sequence Asn-Pro-Asn-Ala (NPNA), tandem repeats of which comprise most of an immunodominant region of the *P. falciparum* circumsporozoite protein. The structure of these repeats has been the focus of several studies, including two that predicted minimum-energy conformations of these repeats (Gibson & Scheraga 1986; Brooks *et al.* 1987). However, neither of these predicted conformations is consistent with NMR studies (Dyson *et al.* 1990) that identified a limited set of H-bonds possible in NPNA-containing peptides. Specifically, these NMR studies allowed that H-bonds could form among the sidechain O^{δ} of Asn*(i)*, backbone C=O of residues *i*-1 and *i*, and backbone NH groups of residues $i+2$ and $i+3$. Later studies (Satterthwait *et al.* 1990) additionally allowed that H-bond(s) could form between the sidechains of asparagine residues flanking proline.

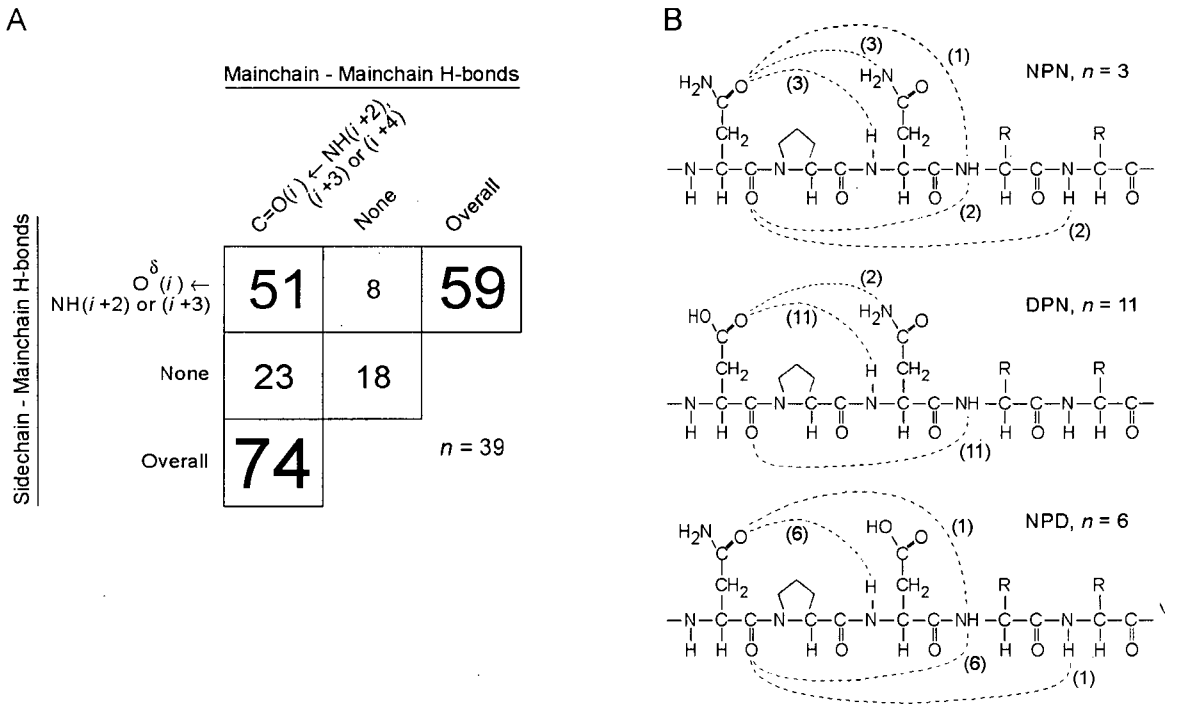


Fig. 3-10. Local H-bonds in Asx-Pro-Asx sequences. The matrix in part **A** shows, per 100 residues, the number of Asx-Pro-Asx sequences identified in dataset C that form (i) local sidechain-backbone and backbone-backbone H-bonds in combination, (ii) sidechain-backbone H-bonds only, (iii) backbone-backbone H-bonds only, or (iv) no local H-bonds. Part **B** provides a summary of H-bonds formed by the 51 per cent of (20 of 39) Asx-Pro-Asx sequences that form Asx-Pro turns. Numbers in parentheses indicate the number of H-bonds formed among the 3 Asn-Pro-Asn, 11 Asp-Pro-Asn and 6 Asn-Pro-Asp sequences.

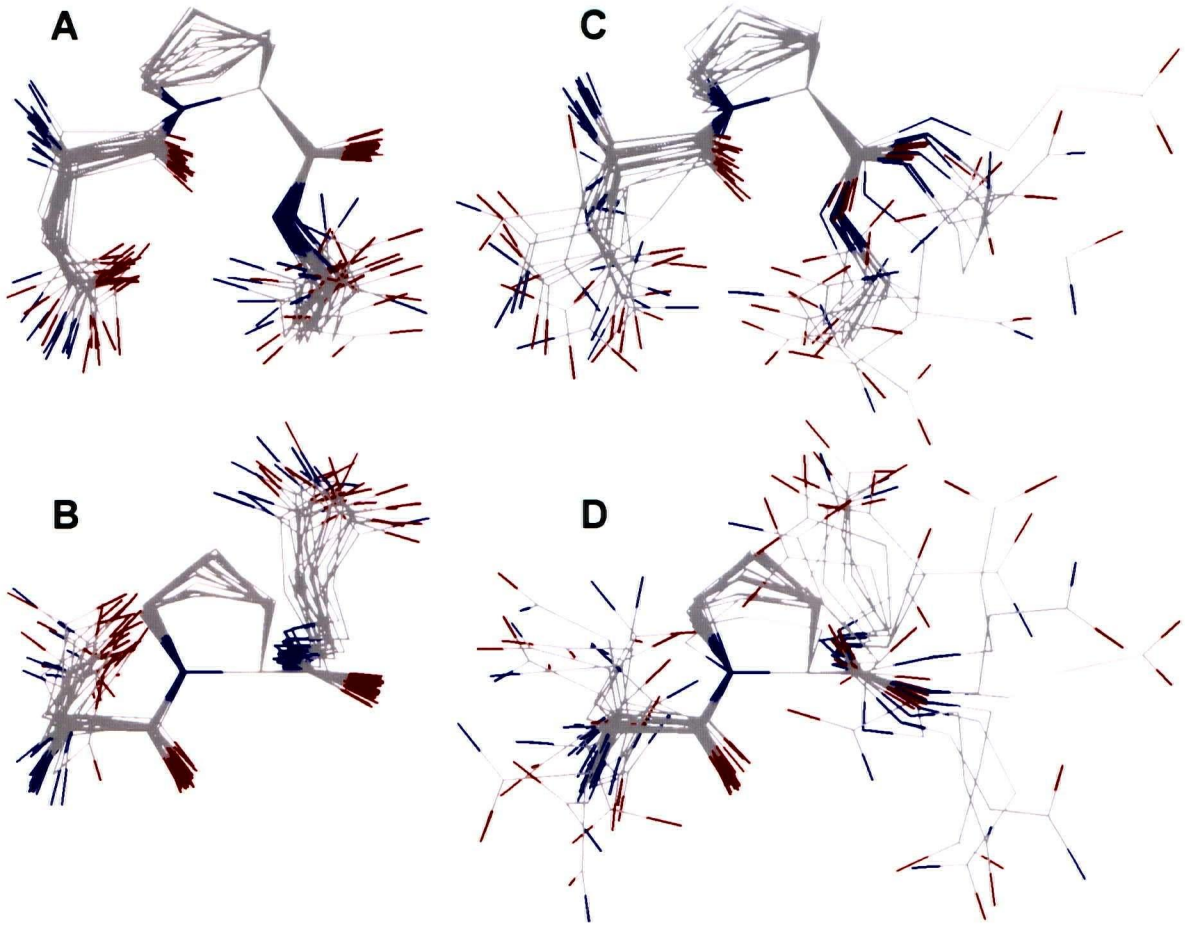


Fig. 3-11. Superimposed wireframe models of Asx-Pro-Asx peptides. Models were prepared from PDB coordinates of 39 Asx-Pro-Asx sequences identified in dataset C and superimposed by hand using general-purpose graphics software. Views **A** and **B** represent 20 Asx-Pro-Asx sequences that form Asx-Pro turns, while views **C** and **D** show 19 other Asx-Pro-Asx sequences. Atoms comprising the C=O group of the 2nd Asx residue are omitted. In views **A** and **C**, proline N-C^{α} and C^{α}-C bonds lie in the plane of the paper. A 90° rotation of these models around the X-axis (away from the viewer) yields views **B** and **D**, respectively. Red, oxygen atoms; blue, nitrogen.

Anticipating that a consensus structure derived from the PDB would accord with these NMR findings and more precisely describe possible hydrogen bonding patterns within NPNA repeats, the PDB was surveyed for structures formed by Asx-Pro-Asx. To allow for Asx-Asx sidechain-sidechain H-bonds, the study was restricted to sequences where both Asx residues were not aspartate. Not unexpectedly, the results (Figure 3-10) are similar to the findings for Asx-Pro: 51 per cent (20 of 39) of Asx-Pro-Asx sequences form H-bonds among sidechain $O^{\delta}(i)$, backbone C=O(i) and backbone NH$(i+n_{2\leq n\leq 4})$ groups. The most common or

"consensus" H-bonds are $O^{\delta}(i) \leftarrow NH(i+2)$ and $C{=}O(i) \leftarrow NH(i+3)$. Importantly, geometries adopted by Asx-Pro-Asx turns are strikingly similar (Figure 3-11).

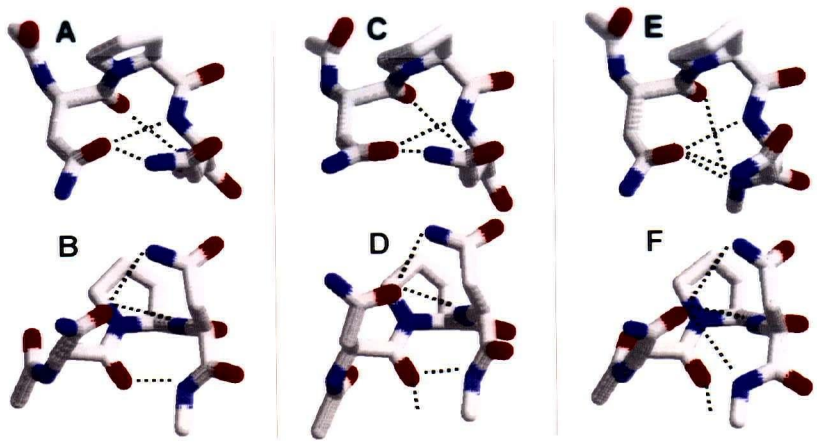


Fig. 3-12. H-bonds in Asn-Pro-Asn sequences that form Asx-Pro turns. The molecular geometries adopted by the three Asn-Pro-Asn sequences identified in dataset C are striking similar. **A** and **B**, PDB entry 1GAL, residues 41-43; **C** and **D**, 1NDK, 150-152; **E** and **F**, 3ICD, 268-270 (Figure 3-2). In views **A**, **C** and **E** proline N-C^{α} and C^{α}-C bonds lie in the plane of the paper. A 90° rotation of these models around the X-axis yields views **B**, **D** and **F** respectively. H-bonds in views **A** and **B** are consistent with those allowed by NMR (Dyson *et al.* 1990) and related (Satterthwait *et al.* 1990) studies of peptides containing Asn-Pro-Asn-Ala repeats.

Although only four sequences were found in which Asx-Asx sidechain-sidechain H-bonds were formed, it may be significant that all of the three Asn-Pro-Asn turns possessed these H-bonds (Figure 3-12). H-bonds in the first of these (Figure 3-12A and B) include the two "consensus" H-bonds described above. That these H-bonds are consistent with NMR (Dyson *et al.* 1990) and related (Satterthwait *et al.* 1990) studies of NPNA-containing peptides suggests that this structure can serve as a consensus-derived partial model for a minimum energy conformation of NPNA.

3.4.8. The proportion of turns *versus* non-turns increases with structure resolution

Structures with relatively low resolution (up to 3 Å) were included in this study in order to analyse a reasonably large dataset: at this low resolution, a significant number of Asx-Pro turns were identified. Notably, however, the proportions of Asx-Pro and Asx-Pro-Asx sequences forming turns increase with improved resolution. Thus, only 45 per cent of Asx-Pro

sequences form turns in structures resolved to ≤3.0 Å, while 63 per cent form turns in structures resolved to ≤1.8 Å (Figure 3-13). Similarly, 51 per cent of Asx-Pro-Asx sequences form Asx-Pro turns in structures resolved to < 3.0 Å, but 73 per cent form turns in structures resolved to ≤1.8 Å (Figure 3-13).

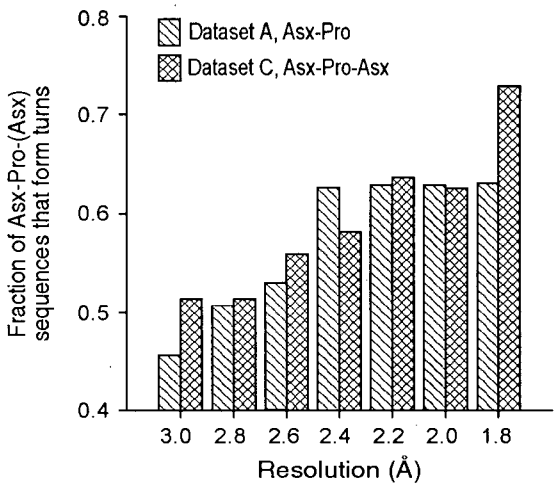


Fig. 3-13. Influence of structure resolution on identification of Asx-Pro turns. Figure shows the fraction of Asx-Pro sequences identified as Asx-Pro turns as a function of structure resolution. Bars summarizing structures resolved to ≤3.0 Å resolution reflect all of the 79 Asx-Pro sequences identified in dataset A and all of the 39 Asx-Pro-Asx sequences identified in dataset C. Bars summarizing higher resolution structures reflect progressively fewer sequences; thus the bars for structures solved to ≤1.8 Å are derived from only 27 Asx-Pro sequences and 11 Asx-Pro-Asx sequences.

3.5. DISCUSSION

This study has shown that, in contrast to a relatively more diverse assortment of conformations adopted by other Xxx-Pro sequences (including those where Xxx = Ser or His), 63 per cent or more of Asx-Pro sequences and 73 per cent or more of Asx_i-Pro-Asx_j (where $Asx_i \neq Asx_j$) sequences, in a variety of flanking structural contexts, adopt strikingly similar conformations. Their common feature is a canonical turn geometry stabilized by a set of alternative H-bonds among the sidechain O^{δ} and backbone C=O of Asx (residue *i*) and backbone NH groups of residues *i*+2, *i*+3 and *i*+4. That alternative H-bonds are possible with changes in geometry would seem to reduce the sensitivity of Asx-Pro turns to perturbation, allow greater flexibility in local packing and greater freedom to substitute neighbouring residues, whether by evolutionary mechanisms or deliberate engineering.

That a majority of Asx-Pro sequences form these turn structures is not surprising. Anecdotal experience, combined with studies of turn-forming propensities of these residues

considered separately, dictate that they "must" form turns. Indeed, the molecular geometry of an Asx-Pro turn makes almost intuitive sense and would seem to reflect simple molecular modelling concepts such as (a) proline-induced and more general steric constraints that limit backbone and sidechain combinations to a small number of preferred choices, and (b) those characteristics of asparagine and aspartate sidechains (highly polar O^{δ} atoms, sidechain length and geometry) that make these residues uniquely suited to forming turn-stabilizing, alternative sets of H-bonds. In contrast, the polar sidechains of serine, threonine and histidine are presumably too short, lack a required conformational freedom or (compared to aspartate) are too versatile, being able to function as H-bond donor or acceptor. Although we might predict that glutamate and glutamine can substitute for aspartate and asparagine, the side chains of the former residues behave quite differently (Figure 3-3), presumably because they fit different geometries and -- having an extra degree of rotational freedom-- must give up more entropy to adopt fixed conformations.

Given that Asx-Pro turn geometry makes sense and reflects an inherent stability, these structures will expectedly find application in protein engineering, such as in design of shaped peptides where long-range interactions need not be considered, and in design of turns and helix caps in proteins. Some limited support for this idea derives from the finding that a consensus structure reflecting hydrogen bonds within Asx-Pro turns can serve as a reliable, partial model for NPNA-containing peptides derived from the *P. falciparum* circumsporozoite protein, for it agrees with published experimental findings (Dyson *et al.* 1990; Satterthwait *et al.* 1990). Chapter 4 provides further experimental support for this idea and illustrates the application of these findings to the expression of shaped, "Asx-Pro turn"-containing peptides on the surface of filamentous phage.

Chapter 4

Recognition of phage-expressed peptides containing Asx-Pro sequences by monoclonal antibodies produced against *Plasmodium falciparum* circumsporozoite protein

4.1. ABSTRACT

The immunodominant region of the *Plasmodium falciparum* circumsporozoite protein is comprised mainly of a series of tetrapeptide repeats that can -- depending on the starting cadence chosen -- be described as $(NANP)_n$, $(ANPN)_n$, $(NPNA)_n$ or $(PNAN)_n$ in one-letter amino acid code. Data from several studies suggest that the NPNA cadence alone is structurally correct, in that each NPNA tetrapeptide effectively forms a structural unit initiated by an Asx-Pro turn. To explore this idea further, and to assess the immunological relevance of peptide conformation as it relates to the cadence of these tetrapeptide repeats, ELISA was employed to compare the abilities of monoclonal antibodies (mAbs) produced against *P. falciparum* sporozoites to recognize repeat-related heptapeptides expressed on the surface of filamentous phage. Having included representatives of both NANP and NPNA cadences, and other peptides in which the number and location of Asx-Pro sequences varied, the study provides evidence that Asx-Pro sequences play an important role in peptide conformation and antibody recognition, that peptide conformation is influenced by the cadence of the tetrapeptide repeats, and that peptide conformation is important to the abilities of these mAbs to recognize their epitopes.

4.2. INTRODUCTION

Antigens of *Plasmodium falciparum* (Pf) and other species of *Plasmodium* -- the causative agents of malaria -- have been the focus of extensive research, the end goal of which

is prevention or treatment of a disease that kills one to two million people each year (Cox 1991). A significant portion of this work has focussed on the immunodominant repeat region of the circumsporozoite (CS) protein, which forms the protein coat of the infective (sporozoite) form of the parasite. In *P. falciparum* this region consists of tandemly repeating tetrapeptides, predominantly Asn-Pro-Asn-Ala (NPNA in one-letter code) or, in the cadence more commonly cited, Asn-Ala-Asn-Pro (NANP). Vaccines based on B-cell responses to these repeats have been unsuccessful and it appears that an effective malaria vaccine must include both B- and T-cell epitopes from several stages of the parasite's life cycle (Nussenzweig & Nussenzweig 1989; Cox 1991; Kaslow 1993; Long 1993; Nardin & Nussenzweig 1993). The CS repeats and related sequences are candidates for inclusion in such a "cocktail" vaccine.

Although the folded structure of *P. falciparum* CS protein has not been determined, it is known that antibodies against peptides comprised of a small number of tetrapeptide repeats recognize CS protein and, conversely, that anti-Pf antibodies from infected individuals recognize these or similar peptides (Ballou *et al.* 1985; Zavala *et al.* 1985, 1986; Del Giudice *et al.* 1987a, 1987b, 1989). This suggests that repeat-containing peptides adopt or sample conformations found in native CS protein and that study of such peptides can provide helpful insights. Several studies support these ideas.

Circular dichroism measurements showed (Fasman *et al.* 1990) that CS repeats are rich in ß-turn content. A more extensive study (Dyson *et al.* 1990) employed nuclear magnetic resonance (NMR) spectroscopy of repeat-containing peptides of the cadences $(NANP)_{n\leq 3}$ and $(NPNA)_n$ to infer the existence of transient turn-like conformations stabilized by hydrogen bonds within a structural unit defined by the sequence NPNA. A followup study (Satterthwait *et al.* 1990) used peptides in which an ethylene bridge linked sidechains of asparagine residues flanking proline and showed that antibodies against these constrained $\overline{\text{NPNA}}$ repeats recognized sporozoites.

Further insight is provided in the analysis presented in Chapter 3 (see also Wilson & Finlay 1997) showing that 63 per cent or more of Asx-Pro sequences and 73 per cent or more of Asx_i-Pro-Asx_j (where Asx_i and Asx_j are not both Asp) sequences in a variety of flanking structural contexts adopt a canonical turn conformation. This "Asx-Pro turn" is stabilized by alternative hydrogen bonds among the sidechain O^{δ} and backbone C=O of Asx*(i)* and the backbone NH of residues $i+2$, $i+3$ and $i+4$. In structures formed by the specific sequence Asn-Pro-Asn, additional hydrogen bonds may form between asparagine sidechains. Because the most common hydrogen bonds within Asx-Pro and Asn-Pro-Asn turns are a subset of those allowed by NMR (Dyson *et al.* 1990) and related (Satterthwait *et al.* 1990) studies of NPNA-containing peptides, it was proposed (Chapter 3) that a consensus structure reflecting these hydrogen bonds can serve as a partial model for a minimum energy conformation of the NPNA tetrapeptide repeat (Figure 4-1).

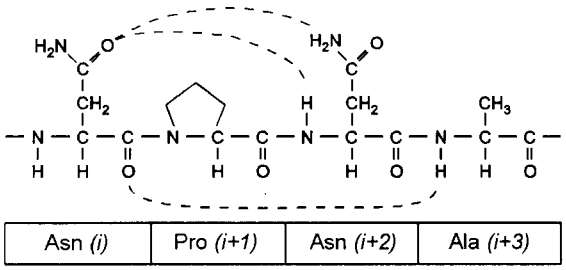


Fig. 4-1. Postulated hydrogen bonds within the sequence Asn-Pro-Asn-Ala. Model is based on experimental (Dyson *et al.* 1990; Satterthwait *et al.* 1990) and statistical (Chapter 3) studies, as reviewed in the text.

That proline is important to the structure of the CS repeats, and that Asn-Pro and Asp-Pro have somewhat interchangeable structural roles, are also suggested by amino acid sequences of CS repeats in diverse *Plasmodium* species, for most contain proline and about half contain Asn-Pro or Asp-Pro. Thus, CS proteins of *P. falciparum* strains contain 30 to 40 occurrences of NPNA, NPNV and DPNA in a ~10:1:1 ratio (Dame *et al.* 1984; Del Portillo *et al.* 1987; Lockyer & Schwarz 1987; Campbell 1989; Caspers *et al.* 1989; Davis *et al.* 1992); these tetrapeptides also occur in CS protein of the chimpanzee parasite *P. reichenowi* (Lal & Goldman 1991). A strain of the rodent parasite *P. yoelii* has 17 DPNA sequences in $DPNAP^{V}_{R}A$ repeats (Colomer-

Gould & Enea 1990), and among strains of *P. berghei* (rodent host) the ratio of Asp-Pro to Asn-Pro is about 2:1 in $N^{P}_{A}NDP^{P}_{A}PP$ octapeptides (Eichinger *et al.* 1986; Lanar 1990; Lockyer *et al.* 1990). In contrast, repeats in *P. vivax* (human host), *P. simium* (monkey) and *P. knowlesi* (monkey) and some *P. cynomolgi* (primate host) strains contain Gln-Pro (Arnot *et al.* 1985, 1988, 1990; Sharma *et al.* 1985; Goldman *et al.* 1993); repeats in other *P. cynomolgi* strains and in *P. malariae* (human host) contain asparagine, aspartate or glutamine but lack proline (Arnot *et al.* 1985, 1988, 1990; Sharma *et al.* 1985; Galinski *et al.* 1987; Lal *et al.* 1988; Goldman *et al.* 1993).

Apart from the study by Satterthwait *et al.* (1990), the immunological relevance of peptide conformation as it relates to cadence of the repeats has not been addressed. This issue was addressed in the current study by using ELISA to compare the abilities of monoclonal antibodies (mAbs) produced against CS protein (Wirtz *et al.* 1987) to recognize phage-displayed CS repeat-related heptapeptides. This involved employing a small library of phage-expressed heptapeptides that included representatives of both NANP and NPNA cadences, and peptides in which the number and location of Asx-Pro sequences were varied, to examine (i) the importance of tetrapeptide cadence and of Asx-Pro sequences to peptide conformation, and (ii) the influence of peptide conformation on the abilities of anti-CS protein mAbs to recognize their epitopes.

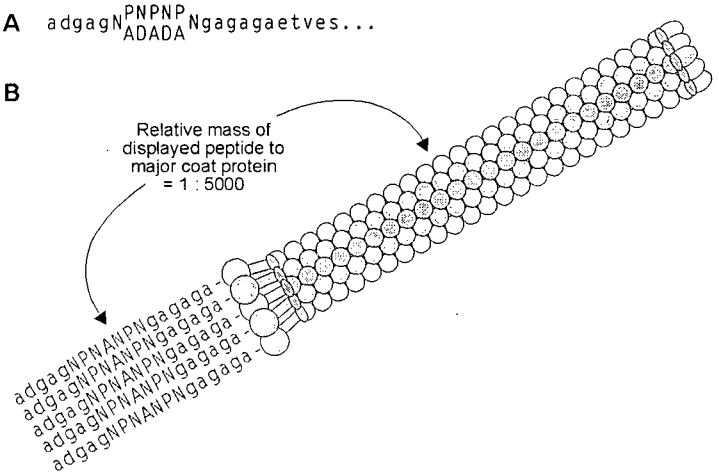


Fig. 4-2. Phage-displayed CSP-related peptides. **A.** Schematic representation of 32 variant peptides and flanking vector and linker residues. Further detail of the construction of these peptides is presented in Figure 7-8 (Chapter 7). **B.** cartoon of the "native" peptide NPNANPN displayed on the N-termini of the four or five copies of pIII present on each virion. The displayed heptapeptide represents ~0.02% of total virion protein.

4.3. RESULTS

Epitopes recognized by six of the eight mAbs employed in this study were previously characterized by Burkot *et al.* (1991) using a "peptide scanning" approach (Geysen *et al.* 1987) that employed overlapping four- to eight-residue peptides derived from both repeat and non-repeat regions of *P. falciparum* CS protein. Considering their findings and the different current objectives of assessing the importance of CS tetrapeptide repeat cadence, the current study employed a set of phage clones, each displaying one of 32 heptapeptide sequences ($N^{PNPNP}_{A\,DA\,DA}N$; see Figures 4-2 and Chapter 7, Figure 7-8) as an insertion near the N-terminus of pIII, a minor phage coat protein present in four or five copies per virion. This 32-member library included the two native peptides NPNANPN and NANPNAN -- each beginning with a different tetrapeptide cadence -- and variant peptides (not found in CS protein) such as NPNPDPN (rich in Asx-Pro) and NANANAN (lacking in proline).

4.3.1. ELISA overview

ELISA was used to measure the abilities of mAbs produced against *P. falciparum* sporozoites to recognize phage-displayed library heptapeptides as well as the phage-displayed 11-residue peptide PNANPNANPNA which, because of sequence overlap, contains the two native heptapeptides NPNANPN and NANPNAN. Although all of the eight anti-Pf mAbs readily recognized a positive control -- a recombinant protein comprised of 32 major (NPNA) and minor (NPNV, DPNA) CS repeats (antigen Pf-R32tet32) -- these mAbs varied widely in their abilities to recognize shorter, phage-expressed peptides (Figures 4-3 and 4-4). Based on these recognition abilities, the mAbs form four groups.

Group 1 mAbs Pf4C11.6 and Pf1G3.4 recognized no phage-expressed peptides (Figure 4-3). The single Group 2 mAb Pf5C1.1 recognized the 11-residue phage-displayed native peptide PNANPNANPNA but neither of the shorter native peptides NPNANPN and NANPNAN (Figure 4-3). Group 3 mAbs Pf2A10 and Pf5G5.3 and Group 4 mAbs Pf1B2.2,

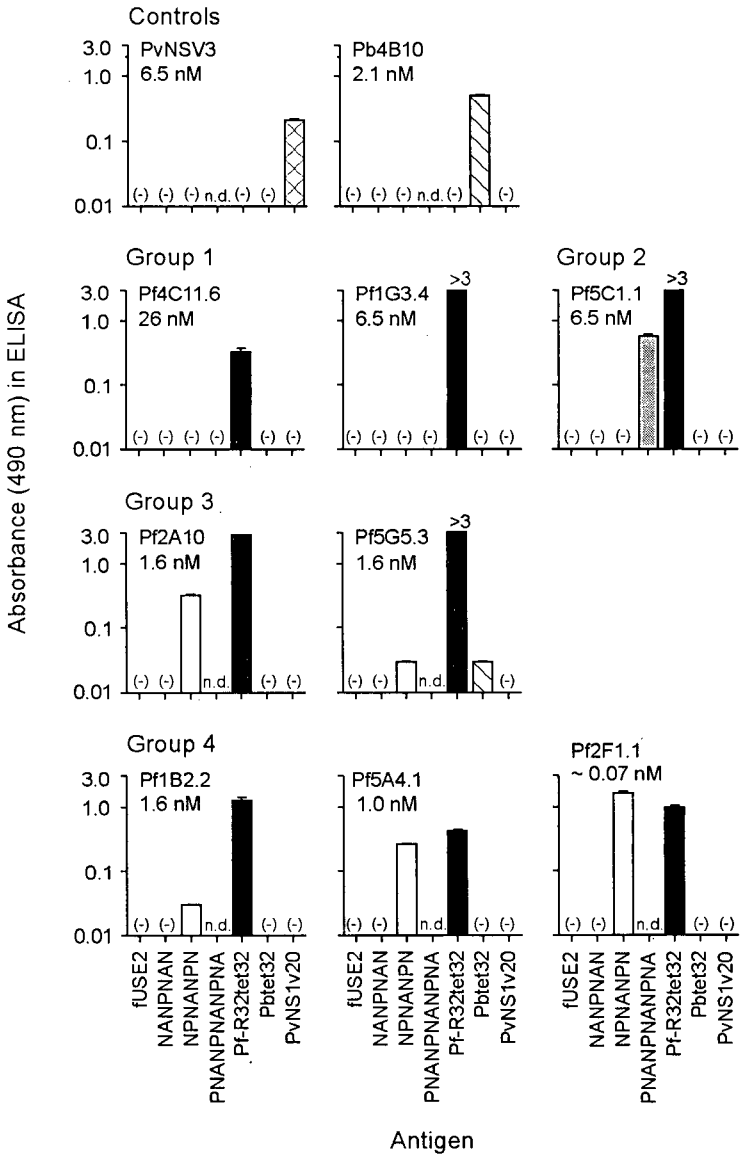


Fig. 4-3. ELISA with recombinant CS protein and phage-displayed native peptides, using ELISA method A (Materials and Methods, §2.16.2) in which CS peptide-displaying virions or recombinant CS protein bound to wells of Immulon plates were probed with mAbs in solution. **Results.** Bar height reflects recognition in ELISA of bacteriophage fUSE2 (negative control), CS peptide-displaying phage (NANPNAN, NPNANPN, PNANPNANPNA) and recombinant CS protein (Pf-R32tet32, Pbtet32 and PvNSv20) with mAbs that recognize CS protein. **Monoclonal antibodies**. PvNSV3 and Pb4B10 (controls) were produced against sporozoites of *P. vivax* and *P. berghei* respectively; mAbs prefixed with "Pf" were produced against *P. falciparum* sporozoites. mAbs were used at the indicated concentrations. **Antigens**. fUSE2, phage display vector without insert, and thus not displaying a CSP-related peptide (negative control). NANPNAN, NPNANPN and PNANPNANPNA, virions displaying the indicated CSP-related peptides. Pf-R32tet32, *P. falciparum* recombinant CS protein (rCSP). Pbtet32, *P. berghei* rCSP. PvNS1v20, *P. vivax* rCSP. **Values**. Means and standard errors for triplicate wells are shown. "-", no significant signal above background. "n.d.", not determined.

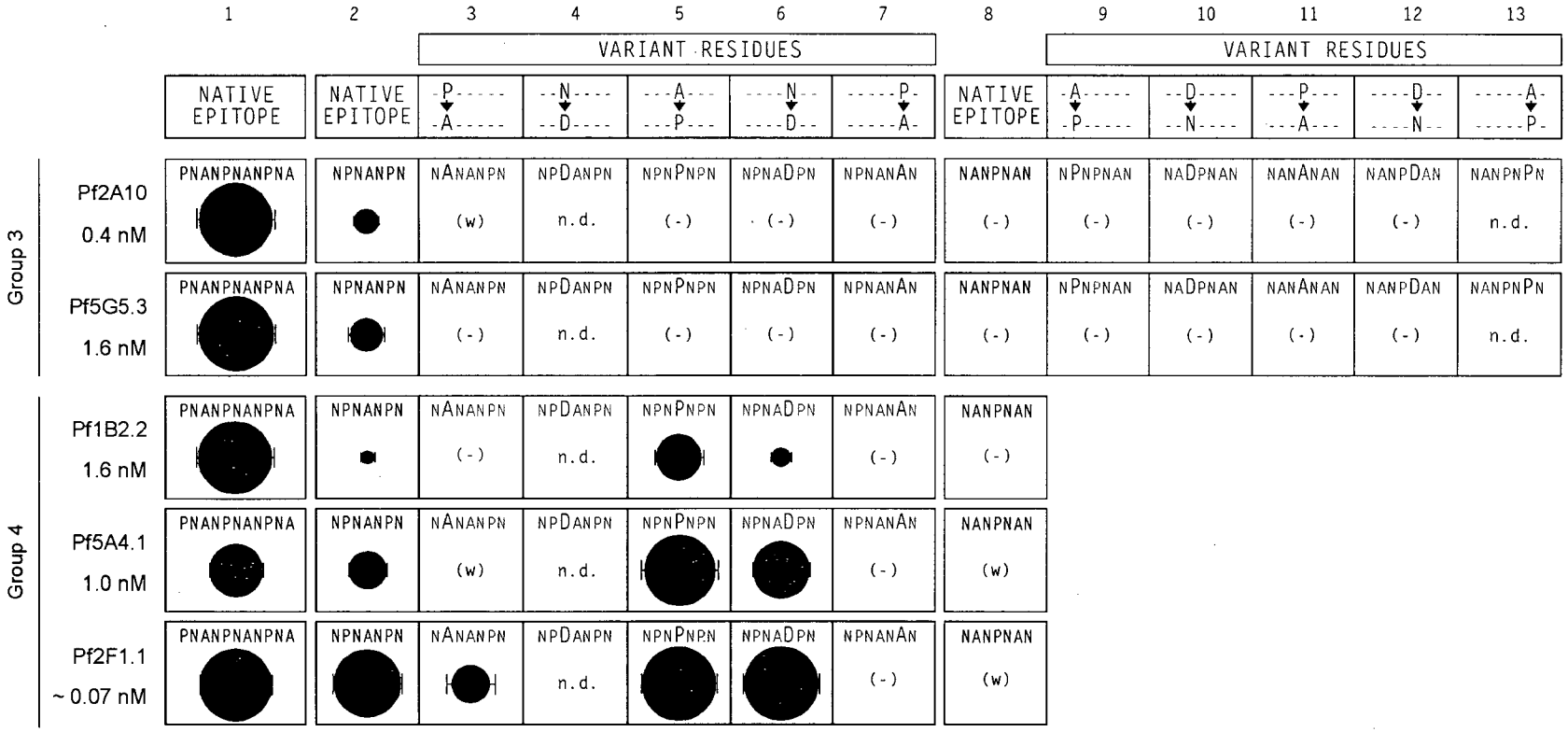


Fig. 4-4. ELISA with phage-displayed native peptides and single-residue variants, using ELISA method A (Materials and Methods, §2.16.2) in which CS peptide-displaying virions coated onto wells of Immulon plates were probed with mAbs in solution, employed at the indicated concentrations. **Results.** Diameter of dot is proportional to absorbance value in ELISA; these are shown relative to the greatest absorbance value observed with each mAb. Width including error bars = mean + two standard errors. Most means are of triplicate wells ($n = 3$) coated with virions of a single clone or a pool of two independently isolated and processed clones displaying the same peptide. Other means are of triplicate wells of each of two independently isolated, processed and assayed clones displaying the same peptide ($n = 6$). "w", weak signal, observed only with higher concentrations of antibody. "-", no significant signal above background. "n.d.", not determined. **Antigens.** Column 1, virions displaying the 11-residue peptide PNANPNANPNA. Columns 2 and 8, virions displaying native heptapeptides NPNANPN and NANPNAN respectively. Columns 3 to 7, variant peptides that differ from NPNANPN by a single residue. Columns 9 to 13, variant peptides that differ from NANPNAN by a single residue.

Pf5A4.1 and Pf2F1.1 recognized not only PNANPNANPNA but NPNANPN as well (Figures 4-3 and 4-4). Significantly, these mAbs did not recognize NANPNAN, which begins with a different tetrapeptide cadence than NPNANPN. Group 3 mAbs recognized none of the non-native peptides tested; these included single-residue variants of both NPNANPN and NANPNAN (Figures 4-3 and 4-4). In contrast, Group 4 mAbs recognized many non-native peptides, including not only single-residue variants of NPNANPN (Figure 4-4) but two- and three-residue variants as well (Figure 4-5).

4.3.2. Variant peptides recognized by Group 4 mAbs

By considering that most heptapeptides in the library can be defined in terms of a "parent" peptide in which a single residue has been substituted, and arranging data accordingly (Figure 4-5), there was a surprising consistency in the recognition of non-native peptides by Group 4 mAbs, in that the same substitutions in different parent peptides affected ELISA absorbance values in a consistent manner. For example, peptides with alanine in position 2 were consistently recognized by Pf5A4.1 less efficiently than otherwise identical peptides with proline in the same position (Figure 4-5B, column 2 *versus* column 1). Conversely, Pf5A4.1 recognized NPNADPN, NANADPN, NPNPDPN and NANPDPN (column 5) more strongly than their respective parent peptides (column 1). Similar, consistent increases or decreases in peptide recognition by Pf5A4.1 occurred with substitutions in peptide positions 3, 4 and 6 (Figure 4-5B, columns 3, 4 and 6 *versus* column 1).

4.3.3. Similar effects of residue substitutions on recognition by Group 4 mAbs

Pf1B2.2 (Figure 4-5A), Pf5A4.1 (Figure 4-5B) and Pf2F1.1 (Figure 4-5C) were affected in similar ways by these substitutions although the degree of effect varied (Figure 4-6). Thus, each mAb preferred or required proline rather than alanine in peptide positions 2, 4 or 6; each was intolerant of aspartate in position 3; and each preferred aspartate rather than asparagine in position 5. It is noteworthy that the peptides most strongly recognized by all three mAbs,

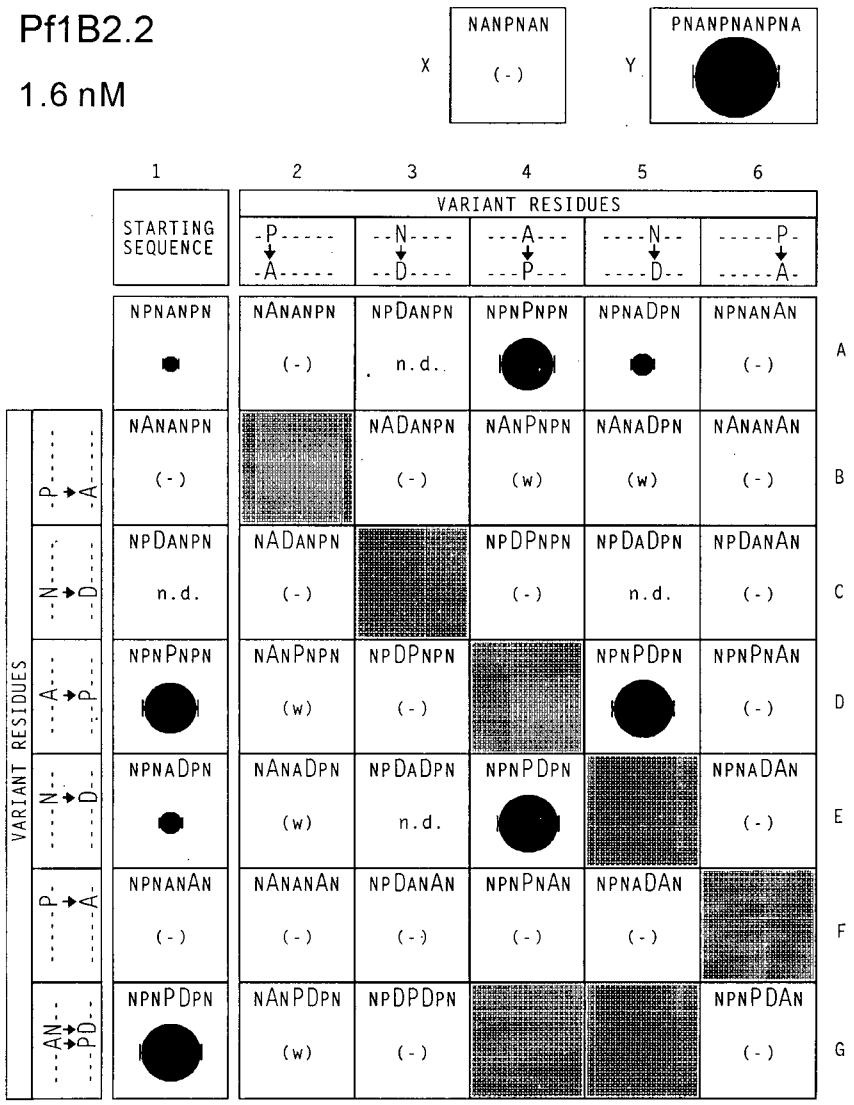


Fig. 4-5A. Group 4 mAb Pf1B2.2 ELISA with phage-displayed native peptides as well as one-, two- and three-residue variants, using ELISA method A (Materials and Methods, §2.16.2), in which CS peptide-displaying virions coated onto wells of Immulon plates were probed with mAbs in solution, employed at the indicated concentrations. **Results.** Diameter of dot is proportional to absorbance value in ELISA, as explained in the legend to Figure 4-4. **Antigens.** Cells **X** and **Y**, virions displaying native peptides NANPNAN and PNANPNANPNA, respectively. **Column 1, Row A = cell 1A**, virions displaying native peptide NPNANPN. All other peptides in columns 1 through 6, rows A to G are one-, two- and three-residue variants of NPNANPN. For example, peptides in cells 1B through 1F are single-residue variants of NPNANPN (cell 1A) while the peptide in cell 1G is a two-residue variant of NPNANPN. As well, each peptide in columns 2 to 6 is a single-residue variant of a peptide occupying a cell of the same row in column 1. Thus, each of the peptides in cells 2A, 3A, 4A, 5A and 6A are single-residue variants of the peptide in cell 1A. Similarly, each of the peptides in cells 2D, 3D, 5D and 6D are single-residue variants of the peptide in cell 1D. Figures 4-5B (Pf5A4.1) and 4-5C (Pf2F1.1) have similar layouts. Text provides examples designed to aid in interpreting Figure 4-5B (Pf5A4.1).

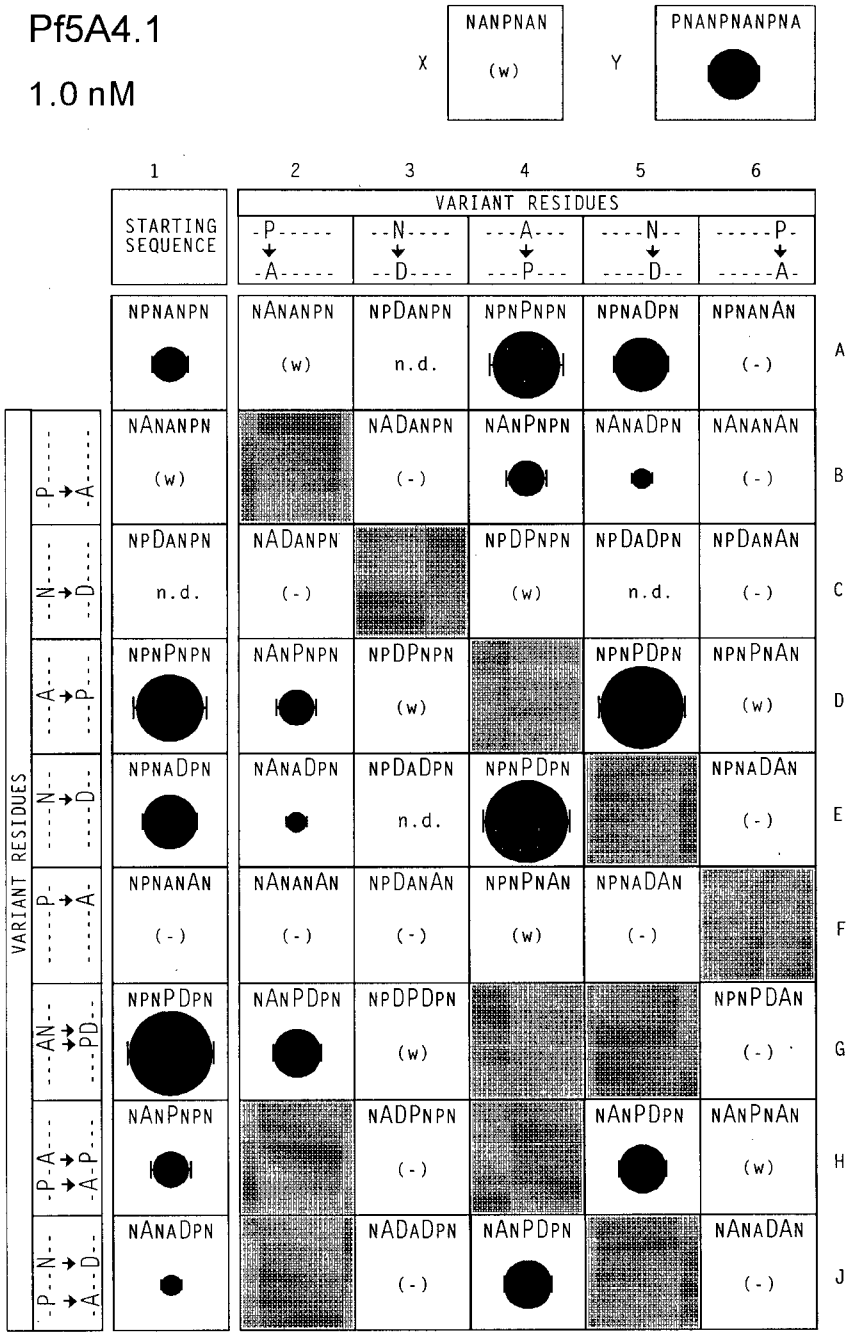


Fig. 4-5B. Group 4 mAb Pf5A4.1 ELISA with phage-displayed native peptides as well as one-, two- and three-residue variants. See legend to Figure 4-5A.

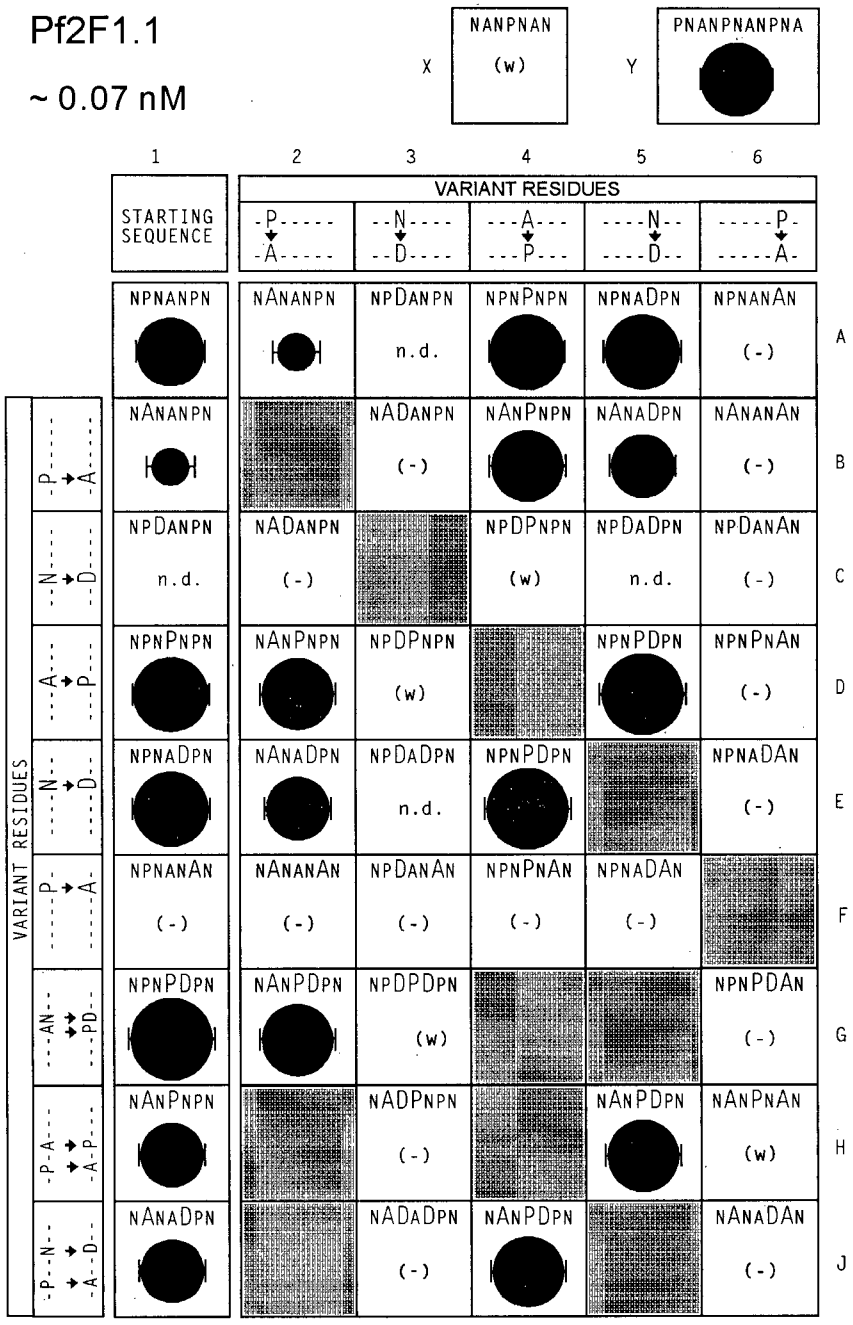


Fig. 4-5C. Group 4 mAb Pf2F1.1 ELISA with phage-displayed native peptides as well as one-, two- and three-residue variants. See legend to Figure 4-5A.

NPNPDPN (Figures 4-5A, B and C, cells 1G, 4E and 5D; see also the top two rows of Figure 4-6) and NPNPNPN (cells 1D and 4A), contain three Asx-Pro sequences each, and that neither of these sequences is found in native CS protein.

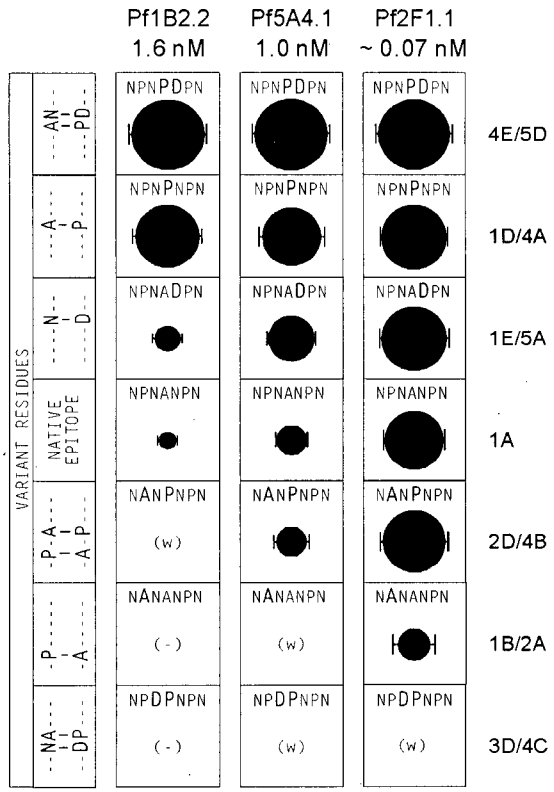


Fig. 4-6. Comparison of ELISA results for Group 4 mAbs (Pf1B2.2, Pf5A4.1 and Pf2F1.1). Derived from data shown in more detail in Figure 4-5. For each mAb, ELISA absorbance values have been standardized relative to those obtained with peptide NPNPDPN (top row). The right-most column shows cell numbers of Figure 4-5. Diameter of dot is proportional to absorbance in ELISA; these and other aspects of the figure layout are explained in the legend to Figure 4-4.

4.3.4. Bivalent Pf1B2.2 and Pf5A4.1 *versus* decavalent Pf2F1.1

Bivalent Pf1B2.2 and Pf5A4.1 (both IgG) exhibited the greatest variation in response to residue substitutions, accepting or rejecting some substitutions much more strongly than others (Figure 4-6). In contrast, decavalent Pf2F1.1 (IgM) showed only small variation in response to the same substitutions, presumably because the combination of decavalent Pf2F1.1 and multivalent virions would allow Pf2F1.1 to bind with relatively high avidity to phage-displayed peptides for which individual Pf2F1.1 antigen-combining sites might have relatively low affinity.

4.3.5. Controls for ELISA

The above results were obtained from ELISA in which virions immobilized on the

surfaces of microtiter plate wells were probed with antibody in solution. Key assumptions were that equal quantities of different clones bound to the wells, that the displayed heptapeptides did not influence the manner in which virions bound, and that the displayed peptides were uniformly accessible to antibodies. These assumptions were tested in three ways.

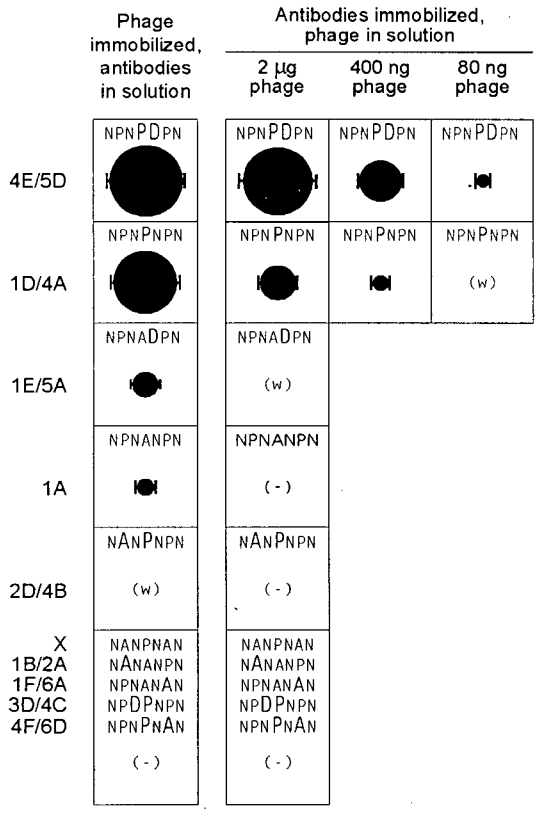


Fig. 4-7. Group 4 mAb Pf1B2.2, comparison of ELISA methods A and B. Comparison of recognition of CS peptide-displaying virions by mAbs against CS protein, using alternative ELISA methods. In Method A (Materials and Methods, §2.16.2.a) microtiter plate wells were coated with virions, and virions were probed with mAbs in solution. In Method B (2.17.1) biotinylated mAbs bound to streptavidin-coated wells were probed with virions in solution. Cell numbers in the left-most column correspond to those of Figure 4-5A. Diameter of dot is proportional to absorbance in ELISA; these and other aspects of the figure layout are explained in the legend to Figure 4-4.

First, ELISA was performed in which, instead of probing immobilized virions with antibodies in solution, immobilized MAbs were probed with virions in solution. Importantly, the assays included peptides that varied widely in their recognition by α-CSP mAbs. For both mAbs tested, Pf1B2.2 (Figure 4-7) and Pf2A10 (data not shown), the different assays yielded similar results, suggesting that peptides displayed on virions bound to wells were as accessible to antibodies as those displayed on virions in solution.

As well, for each clone, bound virions were assayed by coating microtiter plate wells in the manner employed for ELISA and assaying (colorimetric BCA assay) virion protein remaining after the ELISA washing step. For most clones, bound protein was similar, ranging

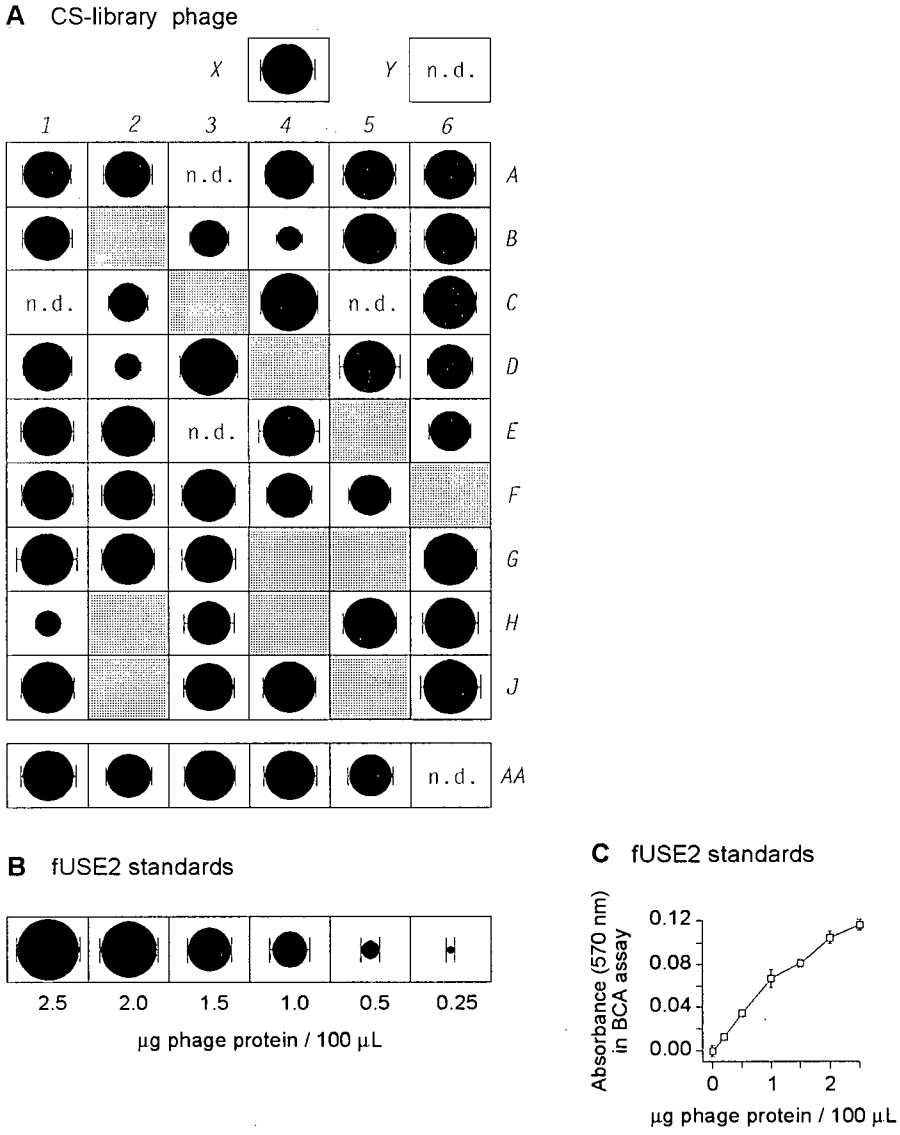


Fig. 4-8. Relative quantities of virions bound to microtiter plate wells, determined by a BCA protein assay (Materials and Methods, §2.16.2.c). **A.** Assay was performed using the same samples used for the ELISA shown in Figures 4-4 and 4-5. Phage clones represented in cells X, Y, and A1 through J6 are the same as those in the corresponding cells of Figure 4-5. Cells A1-A6 and AA1-AA6 correspond to cells in rows 2-7 and 8-13, respectively, of Figure 4-4. Diameter of dot is proportional to absorbance value, and absorbance is proportional to the amount of bound virion protein. Width including error bars = mean + two standard errors. Most means are of triplicate wells (n = 3) coated with virions of a single clone or a pool of two independently isolated and processed clones displaying the same peptide. Other means are of triplicate wells of each of two independently isolated, processed and assayed clones displaying the same peptide (n = 6). **B, C.** Vector fUSE2, assayed concurrently, shown in alternative formats. Mean + two standard errors are shown in **B**, mean ± two standard errors in **C**.

from 83% to 118% of the overall mean (Figure 4-8A). However, for the clone displaying NANPNPN, bound protein was 50% of the overall mean (Figure 4-8A, cells B4 and D2). Based on a followup assay designed to examine the effects of varying the concentration of virions in coating buffer (Figure 4-9), ELISA absorbance values for virions displaying NANPNPN would have been expected to increase less than two-fold for a two-fold increase in bound virions; such an increase would not affect interpretation of the results.

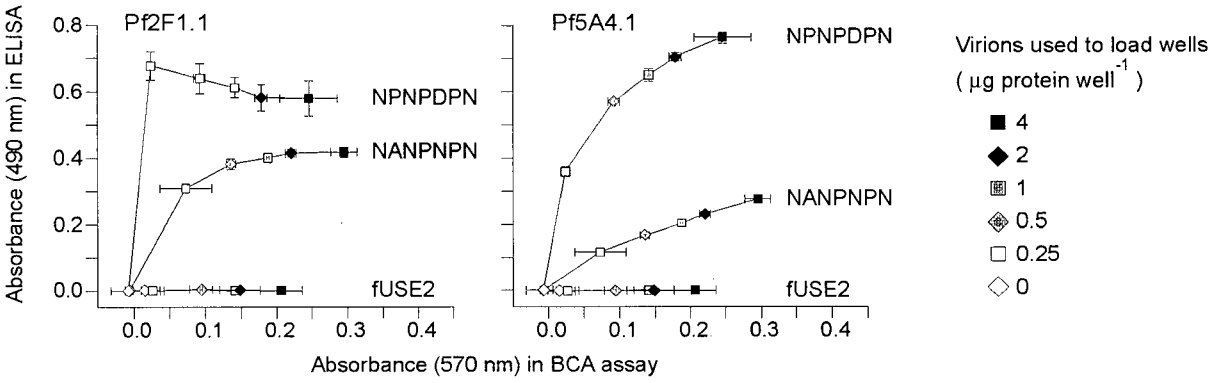


Fig. 4-9. Effect on ELISA signal of varying the quantity of virions loaded into and bound to Immulon plates. These ELISA were performed in the same manner as those of Figures 4-4 and 4-5 (Materials and Methods, §2.16.2.d), except that wells were coated with varying quantities of virions. The phage clone displaying NANPNAN bound poorly to Immulon plates (Figure 4-8); the clone displaying NPNPDPN was included as a control.

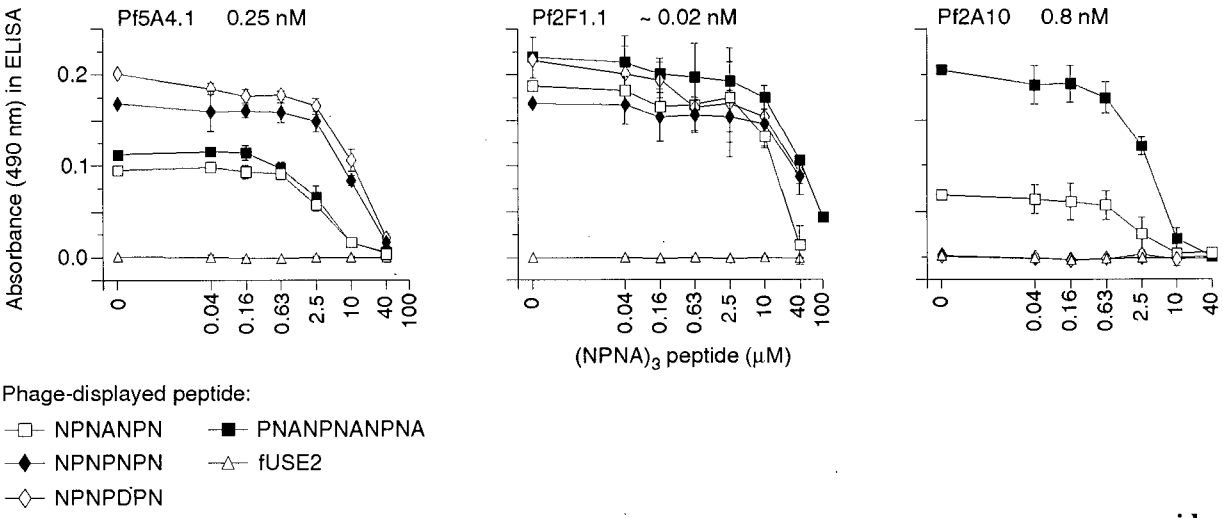


Fig. 4-10. Inhibition of ELISA signal by preincubation of α-CSP mAbs with $(NPNA)_3$ peptide. Virions displaying the indicated peptides were probed with α-CSP mAbs that had been pre-incubated with varying concentrations of the synthetic peptide $(NPNA)_3$ (Materials and Methods, §2.16.2.e). Pf5A4.1 (IgG), Pf2A10 (IgG) and Pf2F1.1 (IgM) were used at the indicated concentrations. Values shown are mean ± two standard errors.

Finally, synthetic $(NPNA)_3$ peptides were found to inhibit binding of mAbs Pf5A4.1 and Pf2A10 to virions displaying different heptapeptides in a manner consistent with the idea that different phage-displayed peptides are equally able to compete with free $(NPNA)_3$ for antibody binding sites (Figure 4-10).

4.4. DISCUSSION

4.4.1. Importance of peptide conformation and tetrapeptide cadence

Burkot *et al.* (1991) have shown that the minimum native NANP- or NPNA-related sequence recognized by mAbs Pf2A10, Pf1B2.2 and Pf2F1.1 is NANPN, a sequence included in both of the phage-expressed peptides NPNANPN and NANPNAN. Given the findings that Pf2A10, Pf1B2.2 and Pf2F1.1 better recognize NPNANPN than NANPNAN (Figures 4-3, 4-4 and 4-5), it follows that NANPN adopts different conformations in these phage-expressed peptides, and that the mAbs prefer NANPN in the conformation provided by phage-displayed NPNANPN. Given also that these mAbs were produced against and recognize sporozoites (Wirtz *et al.* 1987; Burkot *et al.* 1991) and thus recognize native protein conformation, it follows that phage-displayed NPNANPN adopts or samples native CS conformation better than does NANPNAN, which begins with a different cadence.

This argument makes most sense in context of a model (Figure 4-1) that follows (i) from NMR and related experimental studies of $(NANP)_{n\leq 3}$ and $(NPNA)_n$ peptides (Dyson *et al.* 1990; Satterthwait *et al.* 1990) that indicate that NPNA but not NANP forms a structural unit, and that NPNA structural units interact (*viz.*, that the first asparagine of each NPNA interacts with the preceding structural unit); and (ii) from statistical analyses of hydrogen bonds within Asx-Pro sequences (Chapter 3) that suggest that the NPNA structural unit is largely defined by the initial asparagine and proline such that the sequences NPNA, NPNX and NPXX are likely to adopt similar conformations. From this model, it follows that the native conformation of NANPN can be provided by interacting NPNX structural units formed within phage-displayed

NPNANPN (*viz.*, GAGNPNA-NPNGAG, where GAG sequences link NPNANPN to phage protein pIII; see Figure 4-2) but not by the combination of NA with a single NPNA structural unit formed within NANPNAN (*viz.*, GAGNA-NPNA-NGAG). It follows that NPNANPN would be better recognized than NANPNAN by mAbs Pf2A10, Pf1B2.2 and Pf2F1.1.

Table 4-I. Summary[a] of α-CSP mAb recognition of CS repeat-derived peptides.

		Burkot *et al.* (1991)			This study			
		Minimum $(NANP)_n$-derived sequence recognized[d]	Peptide[b]		Peptide[b]			Pf-R32tet32[c]
Group	mAb		NANPNAN	NPNANPN	NANPNAN	NPNANPN	PNANPNANPNA	$[(NANP)_{15}NVDP]_2$
1	Pf1G3.4	PNAN[e]	+	++	-	-	-	++
	Pf4C11.6	n.d.	n.d.	n.d.	-	-	-	++
2	Pf5C1.1	NANPN	+++	++	-	-	++	++
3	Pf2A10	NANPN	++	+++	-	+	++	++
	Pf5G5.3	PNANP	-	+++	-	+	++	++
4	Pf1B2.2	NANPN	++	+++	-	+	++	++
	Pf5A4.1	n.d.	n.d.	n.d.	w	+	++	++
	Pf2F1.1	NANPN	++	+++	w	++	++	++

[a]Summary of Figures 4-3 and 4-4 of the current study and Figure 3B in Burkot *et al.* (1991).

[b]Peptides were presented in different contexts in the two studies. In the current study, peptides were expressed as insertions in the pIII protein of filamentous phage and were separated from pIII residues by Gly-Ala-Gly linker residues (Figure 4-2); in the study of Burkot et al. (1991) N-termini of peptides were acetylated and peptide-like linkers were used to attach peptide C-termini to polypropylene rods.

[c]described in the legend to Figure 4-3.

[d]as identified by Burkot *et al.* or as interpreted from their published data.

[e]Pf1G3.4 best recognizes sequences containing PNV, derived from the CS variant repeat region; it poorly recognizes sequences derived only from the $(NANP)_n$ repeats (Burkot *et al.* 1991).

"+++","++" and "+", relative strength of ELISA signal. "w", absorbance not significantly above background, but weak signal observed with higher concentration of mAb. "-", not significantly above background. "n.d.", not determined.

4.4.2. Peptide acetylation may confound epitope analysis

Regrettably, this model may not entirely fit with the previous work of Burkot *et al.* (1991). Although the current results agree generally with theirs (see Table 4-I for examples), it was puzzling that Burkot *et al.* found that Pf2A10, Pf1B2.2 and Pf2F1.1 reacted well with

acetylated peptides beginning with the NANP sequence, for the current study found that these mAbs recognized NANP-initiated phage-displayed peptides poorly or not at all. Conceivably, these disparate findings may reflect use of different antibody concentrations, different local concentrations of peptide or other methodological issues that point to a greater sensitivity but lesser selectivity of their approach compared to that of the current study. Alternatively, disparaties may reflect dissimilar conformations adopted by identical peptides in different local contexts. The peptides employed by Burkot *et al.* were acetylated at their N-termini (Geysen *et al.* 1987; Burkot *et al.* 1991), while the phage-displayed peptides were separated from the phage pIII protein N-terminus by the sequence ADGAG (Figure 4-2). N-terminal acetyl groups have been shown to function well as N-terminal helix caps in peptides (Chakrabartty *et al.* 1993) -- roughly equivalent to asparagine in unacetylated peptides -- suggesting that N-terminal acetyl groups form acetyl-backbone hydrogen bonds not unlike the sidechain-backbone hydrogen bonds identified for Asx-Pro turns (Chapter 3).

That acetyl groups have this role in proteins is suggested by the molecular geometries of N-acetylated proteins deposited in the Brookhaven protein data bank (PDB; Bernstein *et al.* 1977); the geometries suggest that pairs of hydrogen bonds involving the N-terminal acetyl and first residue backbone C=O groups (acceptors) and backbone NH groups of residues 2, 3, 4 or 5 (donors) stabilize turns (PDB entries 1ATN, chain D; 1CTD, chain A) or initiate α-helices (entries 1AL1, 1COS and 3ADK). That acetyl groups influence peptide conformation may explain anomalous data of Burkot *et al.* that showed that increasing a peptide's length could substantially reduce its recognition by antibody. For example, their ELISA absorbance values for Pf1B2.2 recognition of PNANPN were 1.4- to 2.9-fold greater than the values for NPNANPN, PNANPNA, NPNANPNA, ANPNANPN and PNANPNAN. As well, they found that Pf5C1.1 recognized NANPN more strongly than eight of the nine longer peptides that contained NANPN; in the extreme case, ELISA absorbance values for recognition of NANPN

were more than four-fold greater than those for NPNANPNA and ANPNANPN. That the present findings are consistent with NMR and related experimental and statistical studies described earlier suggests that the phage-displayed peptides adopt conformations that better reflect native protein folding than do the acetylated peptides employed by Burkot *et al.*

4.4.3. Is peptide length important?

Although NMR data (Dyson *et al.* 1990) indicate that flanking NPNA repeats interact, it is not clear whether these interactions are important to antibody recognition. For interacting repeats, one would predict that the middle repeats within a long peptide are more likely to adopt native conformation than those at the N- and C-termini. Longer peptides should thus better reflect native conformation than shorter peptides and, if conformation is important to antibody recognition, be better recognized by antibody. Both data from the current study as well as previous studies of CSP tetrapeptide-related peptides (Ballou *et al.* 1985; Zavala *et al.* 1985) provide some support for this idea. Specifically, all of the eight mAbs studied recognized Pf-R32tet32 (which contains 30 repeats of NANP and 2 minor repeats of NVDP), only six recognized the phage-displayed 11-residue peptide PNANPNANPNA and only five of these recognized the heptapeptide NPNANPN (Figures 4-3 and 4-4; Table 4-I). Moreover, three of the mAbs that recognized both PNANPNANPNA and NPNANPN recognized PNANPNANPNA much more strongly (Table 4-I). However, two factors confound this analysis. First, all of the six mAbs studied by Burkot *et al.* recognized not only sequences derived solely from NANP repeats, but also sequences derived from a combination of NANP and NVDP (Burkot *et al.* 1991). This likely accounts for some of the greater recognition of Pf-R32tet32, which contains the latter sequences. Second, PNANPNANPNA contains two occurrences of the minimum sequences (PNAN, NANPN and PNANP; Table 4-1) recognized by many of these mAbs. This may account for some of the increased recognition of PNANPNANPN compared to NPNANPN, which contains only one occurrence of each of these

minimum sequences. Nevertheless, substantial MAb-to-MAb differences in the ratios of absorbance values for PNANPNANPNA *versus* NPNANPN (Figure 4-3) suggest at least some role for peptide length in influencing conformation.

4.4.4. Structural implications of the recognition of variant peptides by Group 4 mAbs

Several aspects of the results with Group 4 mAbs Pf1B2.2, Pf5A4.1 and Pf2F1.1 (Figures 4-5A to C, 4-6 and 4-7) are important. First, that each of these mAbs prefers aspartate over asparagine in peptide position 5 supports the idea that NPNA sequences form "Asx-Pro turns" (Chapter 3) by adopting the hydrogen bonds shown in Figure 4-1. Specifically, by considering that position 5 is the expected point for initiation of a second NPNA structural unit, then the two sidechain O^{δ} atoms of an aspartate at this position can be expected to more readily form the hydrogen bonds required for native conformation (Figure 4-1) than can the single O^{δ} of a native asparagine residue. As well, that each of these mAbs has an absolute requirement for asparagine rather than aspartate in position 3 supports the existence of a proposed hydrogen bond between asparagine sidechains (Figure 4-1). Because this hydrogen bond involves the N^{δ} of Asn_3 in the sequence Asn_1-Pro_2-Asn_3, aspartate in this position (i.e., Asn_1-Pro_2-Asp_3) cannot substitute for asparagine for it lacks a N^{δ} atom. Finally, that each of these mAbs requires or prefers proline over alanine in positions 2, 4 and 6 supports the idea that proline plays an important role in restricting conformational freedom (Chapter 3) so as to favor formation of these hydrogen bonds.

4.4.5. Multiple Asx-Pro repeats as immunogens

For vaccine development, it is noteworthy that all three mAbs recognized the non-native sequences NPNPDPN and NPNPNPN more strongly than any other peptides studied, suggesting that these and other peptides containing multiple Asx-Pro repeats may be useful as immunogens. Interestingly, although NPNPDPN is not a native sequence, it contains the native sequence NPDPN which is found in a short region of CS protein adjacent to the major

tetrapeptide repeats (Burkot *et al.* 1991). That phage-displayed NPNPDPN was recognized by Group 4 mAbs, while NPDPNPN was not, underscores the importance of cadence when engineering candidate immunogens. In this context, it is noteworthy that the SPf66 cocktail vaccine (Patarroyo *et al.* 1987, 1988; Alonso *et al.* 1994) links blood stage antigens with the sequence NANP rather than the structure-defining NPNA.

4.4.6. Concluding remarks

These findings provide insight into the properties of epitopes recognized by eight anti-Pf mAbs, particularly Pf1B2.2, Pf5A4.1 and Pf2F1.1. More importantly, these results taken in context of a model (Figure 4-1) derived from experimental and statistical studies suggest (i) that the immunodominant repeats of *P. falciparum* CS protein adopt turn-like conformations largely defined by the sequence NPNA and stabilized by hydrogen bonds that derive from a unique geometry imposed by the initial Asn-Pro residues; (ii) that this conformation is influenced by the cadence of the tetrapeptide repeats; and (iii) that this conformation is important to the abilities of the mAbs we studied to recognize their epitopes. Given the paucity of data on the structure of CSP, and the lack of agreement between theoretical predictions (Gibson & Scheraga 1986; Brooks *et al.* 1987) of the structure of the immunodominant tetrapeptide repeats and NMR (Dyson *et al.* 1990) and related (Satterthwait *et al.* 1990) experimental data, the consensus-derived model of NPNA structure (Figure 4-1) should provide a useful basis for continued study of CS protein.

Chapter 5

Biopanning random peptide libraries (RPLs) with monoclonal antibodies raised against *Plasmodium falciparum* circumsporozoite protein suggests inherent practical limitations of RPLs

5.1. ABSTRACT

Two phage-display libraries, displaying 6- and 15-residue random peptides (RPL_6 and RPL_{15}), were biopanned with two monoclonal antibodies (mAbs) raised against *Plasmodium falciparum* circumsporozoite protein (CSP), with the goals of (i) confirming earlier findings that Asx-Pro sequences are important to the abilities of these mAbs to recognize their epitopes, (ii) providing insight into the epitope specificities of these mAbs, and (iii) gaining experience that would help determine the feasibility of more complex applications of random peptide libraries in the study of pathogenesis. Although, for one of the mAbs, biopanning RPL_6 identified a peptide sharing relatively high sequence similarity with its known epitope, neither this nor the other biopannings yielded an expected collection of sequences related to CSP sequences. These findings are discussed in context of inherent practical limitations of phage-displayed random peptide libraries that, notwithstanding the successes of early pioneering studies, limit their utility and broad applicability.

5.2. INTRODUCTION

Phage displayed random peptide libraries (RPLs) have been successfully employed to identify ligands for antibodies and a variety of other molecules (Chapter 1). An example of this is provided in the study by Conley *et al.* (1994), in which biopanning a 15-residue RPL with an anti-HIV-1 (human immunodeficiency virus-1) mAb led to the identification of 20 mAb-reactive clones that could readily be classified into four sequence groups (xxDKW,

xLD$^{R}/_{K}$W, ExD$^{R}/_{K}$W and ELD$^{R}/_{K}$W) that (i) allowed identification of ELDKW (a sequence in the native protein) as the epitope recognized by the α-HIV-1 mAb, and (ii) provided information about residues that were most important to binding. A theme common to this and certain other studies, particularly the pioneering RPL papers (Cwirla *et al.* 1990; Devlin *et al.* 1990; Scott & Smith 1990), is that biopanning a random library can readily lead to a set of related sequences that as a *collection* provide insight not obtainable by methods that lead to a single sequence or by approaches such as epitope scanning.

In context of this idea, it seemed possible to extend the earlier analysis of mAbs raised against *Plasmodium falciparum* circumsporozoite protein (CSP). As shown in Chapter 4, these mAbs bind most effectively when their epitopes are presented in context of an Asx-Pro turn (Chapter 3). An expectation was that that biopanning RPLS with α-CSP mAbs would confirm this finding by identifying a set of mAb-reactive Asx-Pro-containing sequences and provide additional insight into the epitope specificities of these mAbs. The α-CSP mAbs chosen for a first trial, Pf2A10 and Pf1B2.2, vary in their residue requirements. Of the heptapeptides examined (Chapter 4), Pf2A10 recognized only NPNANPN, while Pf1B2.2 recognized sequences of the form NPN$^{A}/_{P}{}^{N}/_{D}$PN and, more weakly, three of four NAN$^{A}/_{P}{}^{N}/_{D}$PN sequences. As well, experiences gained in this simple model were expected to be useful in determining the feasibility of employing RPLs in more complex applications related to pathogenesis, such as in the identification of phage-displayed peptides able to block bacterial adhesion or inhibit entry of bacteria into eukaryotic cells.

{ Pf2A10 (narrow specificity); Pf1B2.2 (broader specificity) } x { 1.2 µg; 120 ng; 12 ng; 1.2 ng } x { 6-mer library; 15-mer library } = 16 biopans

Fig. 5-1. Overview of 16 biopans. Four quantities of each of two α-CSP mAbs were employed to biopan two phage-displayed random peptide libraries.

5.3. RESULTS

Two RPLs, displaying 6- and 15-residue peptides (RPL_6 and RPL_{15}), were biopanned. Each library was biopanned with four quantities of each of two α-CSP mAbs (Figure 5-1). Briefly, this involved incubating library aliquots with biotinylated antibodies overnight before capturing the resulting antibody-phage complexes in streptavidin-coated microtiter plate wells. Three rounds of biopanning were carried out, and after each round the enriched fractions were amplified by infecting *E. coli* and harvesting virions produced during overnight culture.

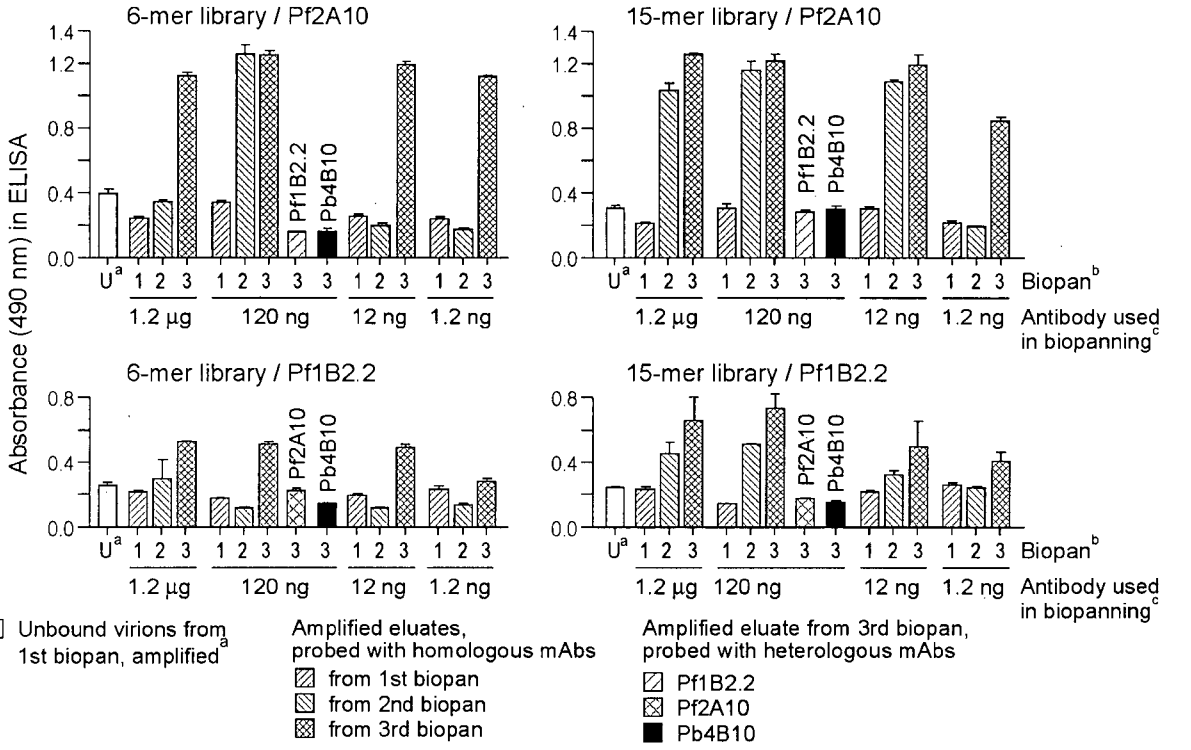


Fig. 5-2. Apparent enrichment for target clones with multiple rounds of biopanning. For each set of biopans (*i.e.,* each combination of library [6- or 15-residue] and mAb [Pf2A10 or Pf1B2.2]), amplified outputs from each round of biopanning were assayed by capture ELISA (§2.17.2, Materials and Methods) in which (i) virions were captured by biotinylated mAbs immobilized in streptavidin-coated microtiter plate wells, and (ii) captured virions were quantified with protein A-purified rabbit α-f1(1°) pAbs and peroxidase-conjugated 2° antibodies. Each of the four sets of assays included, as controls, (i) unbound virions recovered from the first round of biopanning, (ii) output from the 3rd round of biopanning with 120 ng mAb, probed with two heterologous mAbs.

[a]Unbound virions are those recovered from the first biopan before eluting target clones; these were amplified in *E. coli* K91-Kan and used as negative controls. [b]Output from 1st, 2nd or 3rd round of biopanning. [c]Quantity of antibody that had been employed in biopanning.

5.3.1. Progressive enrichment for target clones with each round of biopanning

Enrichment for target clones with each round of biopanning was shown by ELISA (Figure 5-2) performed in a manner parallelling biopanning, that is by capturing enriched fractions with biotinylated mAbs immobilized in streptavidin-coated microtiter plate wells. Enrichment was evident for each of the sixteen biopans, although greater and earlier enrichment was seen for biopannings with 1.2 μg and 120 ng of mAb than with 12 and 1.2 ng. Results varied little among biopannings with 1.2 μg and 120 ng, except for the Pf2A10/RPL_6 pannings. Here earlier enrichment was seen with 120 ng than with 1.2 μg. This may have been due -- in the biopannings with 1.2 μg -- to excess of antibody over the biotin-binding sites of streptavidin, resulting in loss of target clones (and thus sequence diversity) during the first biopan. Considering this and the common belief that biopanning with a relatively larger quantity of antibody leads to the identification of target clones with relatively lower affinity but of greater sequence diversity (*e.g.*, Stoute *et al.* 1995), fractions enriched with 120 ng (*versus* those enriched with lesser quantities) were chosen for further analysis.

5.3.2. Screening of 96 clones yielded only a limited set of sequences

Twenty-four clones chosen from the third round of each biopanning with 120 ng were screened in a preliminary ELISA (Figures 5-3 and 5-4). To avoid artefacts of nonspecific binding involving biopanning system components (Smith & Scott 1993; see also Chapter 1), an alternative ELISA format was chosen. Here, peptide-displaying virions coated onto plates were probed with mAbs in solution. ELISA signals were relatively weak with Pf2A10 (Figure 5-3) and relatively high mAb concentrations (≥4-fold more concentrated than the lowest dilutions used in the work reported in Chapter 4) were required for signal development with both mAbs, suggesting that the putative clones reacted less well than those employed in earlier work (Chapter 4). Importantly, controls suggested that mAb-reactive clones had been identified. Thus, for Pf2A10/RPL_6 biopanning (Figure 5-3), at least eight clones gave ELISA signals

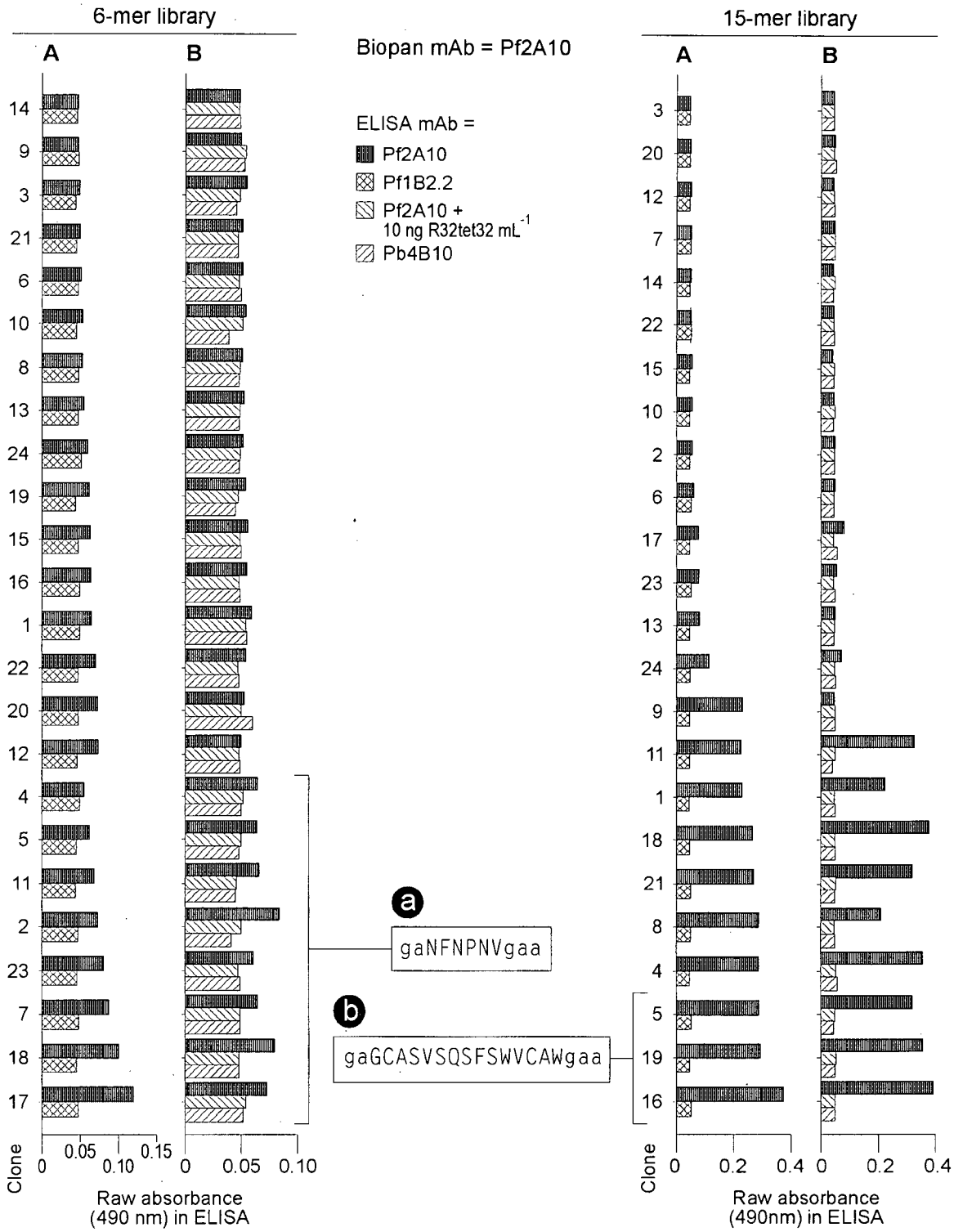


Fig. 5-3. Initial screen of clones selected from third round of biopanning with Pf2A10. Twenty-four clones from each of the 6-and 15-residue RPLs biopanned with 120 ng mAb Pf2A10, screened by ELISA Method A (§2.16.3) for recognition by (i) Pf2A10, (ii) Pf2A10 pre-incubated with recombinant CS protein (R32tet32) and (iii) heterologous mAbs Pf1B2.2 and Pb4B10. Raw absorbance values, based on single wells, are shown. Characteristic background A_{490} for similar ELISAs is 0.05. **(a)** and **(b)** identify the two unique sequences possessed by the 11 clones that were sequenced; these are discussed further in the text.

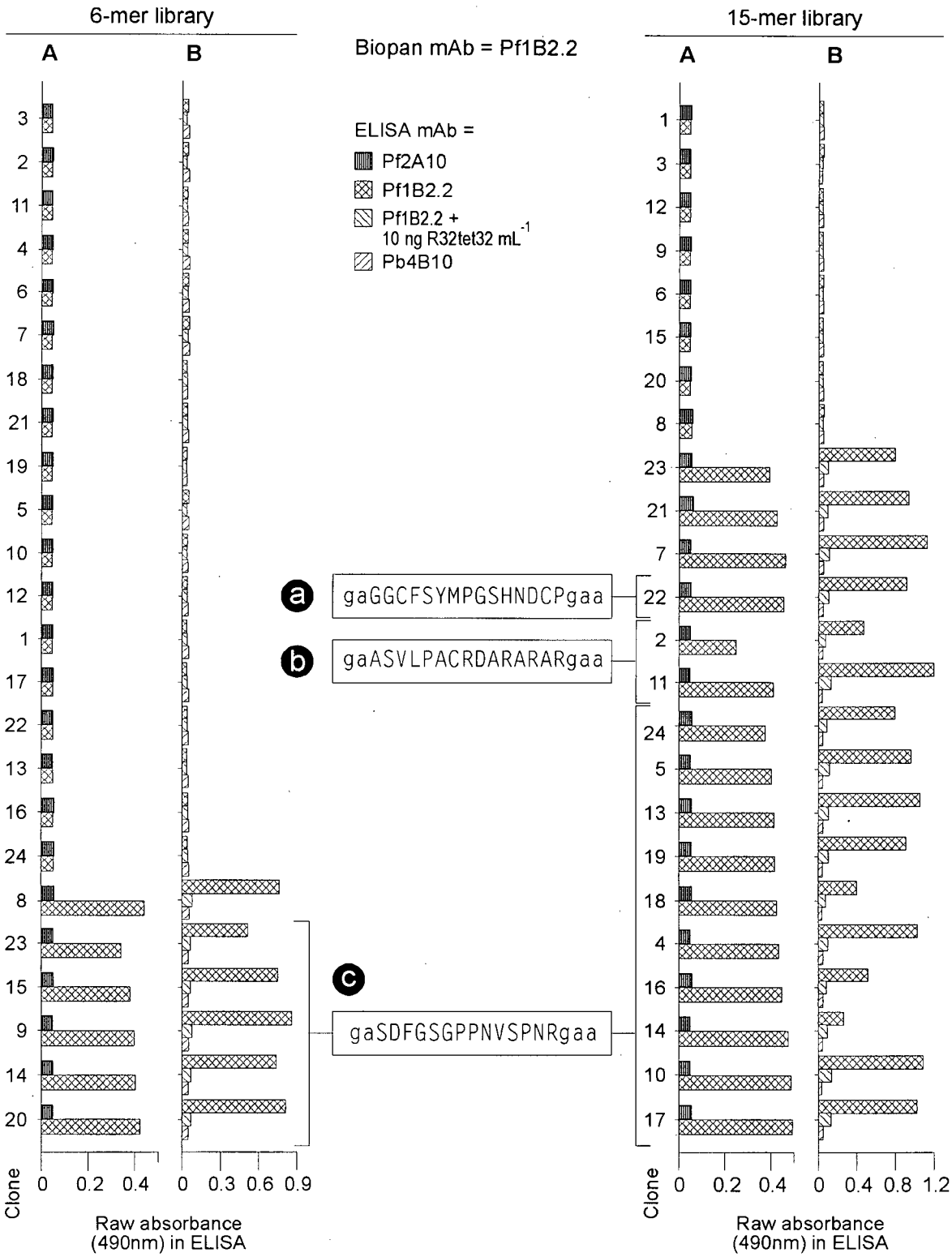


Fig. 5-4. Initial screen of clones selected from third round of biopanning with Pf1B2.2. In assays parallelling those of Figure 5-3, 24 clones biopanned from each of the 6-mer and 15-mer random peptide libraries with 120 ng mAb Pf1B2.2 were screened for recognition by (i) Pf1B2.2, (ii) Pf1B2.2 pre-incubated with recombinant CS protein (R32tet32) and (iii) heterologous mAbs Pf2A10 and Pb4B10. Further details of ELISA are provided in the legend to Figure 5-3. **(a)**, **(b)** and **(c)** identify the three unique sequences possessed by the 18 clones that were sequenced; these are discussed further in the text.

above common background levels with Pf2A10 but not with the Pf1B2.2 or Pf2A10 preincubated with recombinant CSP; for Pf2A10/RPL_{15}, nine clones reacted in this manner. A corresponding screen for RPLs biopanned with Pf1B2.2 (Figure 5-4) identified six RPL_6 and sixteen RPL_{15} antibody-reactive clones.

In contrast to expectations, sequencing identified only a small number of unique clones (Figures 5-3 and 5-4). Strikingly, each of the eight sequenced Pf2A10-reactive RPL_6 clones sequenced possessed the same sequence, NFNPNV (Figure 5-3, *a*), and each of the three Pf2A10-reactive RPL_{15} clones possessed GCASVSQSFSWVCAW (Figure 5-3, *b*). The former sequence resembles native CSP sequences such as NANPNA and NANPNV. Although the latter sequence is interesting (it contains a pair of Cys residues which suggest disulfide-bond potential, the motif Ser-X-Ser-X-Ser-X-Ser, and three aromatic amino acids, Phe and two Trp), it lacks similarity to CSP sequences.

It was puzzling that a common *14*-residue sequence (Figure 5-4, *c*) was found in the (putative) output from panning RPL_6 and RPL_{15} with Pf1B2.2. Although it is tempting to dismiss this sequence (SDFGSGPPNVSPNR) as a contaminant, it accounted for all of the five Pf1B2.2-reactive RPL_6 clones sequenced and 10 of the 13 RPL_{15} clones and its sequence is somewhat similar to CSP sequences PNANPN, PNVDPN and PND-PNR. Two other sequences found in the Pf1B2.2/RPL_{15} biopanning are interesting for their nature but are dissimilar to CSP protein sequences. One of these, GGCFSYMPGSHNDCP (Figure 5-4, *a*), resembles the earlier-described Pf2A10/RPL_{15} sequence in that it contains a pair of Cys residues as well as Ser and aromatic amino acids. The second of these (Figure 5-4, *b*) contains the motif RXARARAR. Considering the results in context of issues discussed below, and that during the course of this work Stoute *et al.* (1995) published a more extensive study that similarly involved biopanning a hexapeptide RPL with Pf2A10, the study was not extended beyond these preliminary screens.

5.4. DISCUSSION

The selection of a *14*-residue peptide from (apparently) both RPL_6 and RPL_{15} is troubling. Paradoxically, its sequence is somewhat similar to CSP sequences. Since both RPL_6 and RPL_{15} were constructed with fUSE5 and were obtained from investigators who collaborate and share recombinant phage materials, it seems possible (but unlikely) that the 14-residue sequence represents a contaminant common to fUSE5 stocks used to prepare these libraries. Alternatively (seemingly more likely in context of procedures employed for these studies), it may represent an unexpected recombinant that arose in either RPL_6 or RPL_{15} and which, by cross-contamination during biopanning, was thus recovered from both libraries.

That this work was not fundamentally flawed is suggested by the reasonable agreement of the single Pf2A10/RPL_6 sequence identified, NFNPNV, with the sequence NXNPQ shared by five Pf2A10-reactive phage clones identified by Stoute *et al.* (1995) in their related biopanning study. Although I at first found it curious that NFNPNV and the NXNPQ sequences more closely resembled NANPN than a sequence reflecting the NPNA structural cadence described in Chapter 4, I later considered that a hexapeptide, by design, cannot both begin with the NPNA structural cadence and contain NANPN, the minimal native epitope for Pf2A10 (Burkot *et al.* 1991): a heptapeptide is required. Non-specific binding may have led to the other sequences, for it is possible to affinity-select target clones that bind to biopanning system components (Smith & Scott 1993; see also Chapter 1). That both Pf2A10 and Pf1B2.2 yielded compositionally similar (Cys, Ser and aromatic amino acids) RPL_{15} clones suggests that these may fall into this category; indeed, certain phage-displayed peptides containing aromatic residues bind to plastic (Adey *et al.* 1995).

5.4.1. Practical issues

This study did not yield an expected collection of related sequences, but rather only a limited set, some with no sequence similarity to the previously characterized (Chapter 4 and

references cited therein) epitopes for the mAbs employed. For example, only a single seqence was identified after biopanning and subsequent screening of RPL_6 with Pf2A10. Conceivably, the identified peptide represents the most avidly-binding target sequence. This was not tested, such as by comparing recognition by Pf2A10 of the identified sequence with recognition of a native CS sequence. Notably, identification of the single most strongly recognized peptide is not a specific goal of many studies that employ biopanning, including the present study. Rather, as illustrated by the results of Stoute *et al.* (1995), who similarly biopanned a RPL_6 with Pf2A10 and identified five CS-related sequences, an expected and desired result is that a number of related sequences be identified, such that analysis of these sequences can allow identification of residues critical to antibody-antigen recognition. In this context, identification of a single sequence provides little information. As well, even in studies where identification of the single most strongly recognized peptide *is* desirable, it seems likely that many investigators will prefer to identify a number of related sequences by biopanning and to subsequently, by ELISA, compare their reactivities. Intuitively, results obtained in this way would seem more reliable.

5.4.2. Biopanning results are essentially unpredictable

The appeal of the pioneering RPL studies cited in Chapter 1 derived in large part from both the technical simplicity of the methodology and the nature of the results obtained, particularly of the common theme that biopanning would lead to a collection of related sequences. Yet as reviewed in Chapter 1 and demonstrated here, these results are not always achieved. Several explanations have been offered for this (*e.g.*, Böttger & Lane 1994; Burritt *et al.* 1996), including that (i) the target peptide may be missing from a library, (ii) some antibodies may recognize conformation-dependent or discontinuous epitopes that cannot be mimicked by library peptides, or (iii) the target peptide may be longer than library peptides.

Conceptually, there is a more fundamental difficulty with biopanning, a practical

consequence of (i) variations among biopanning applications and among target peptides of the affinities of an antibody (or other ligate) for target peptides, and (ii) the idea that repeated rounds of biopanning may -- depending on these variables and the specificity of an antibody for its ligand -- lead to many, a few or only a single peptide with the highest achievable affinity for the antibody. Commonly, the goal is not a single peptide but rather a collection. Thus, success derives from (i) expecting the system to work ineffectively and yield the desired collection, or (ii) by tuning the system to lead to a "reasonable" number of target peptides with varying but noticeable degrees of sequence similarity, and with varying affinities for antibody, so that meaningful conclusions can be drawn regarding critical residues or the most strongly-recognized peptide confidently identified. This involves determining such parameters as optimal antibody concentrations, the number of rounds of biopanning, and the numbers of affinity-selected clones to be screened and sequenced. Expectedly, some times this can be worked out quickly, other times not.

Such difficulties are acknowledged indirectly in the literature, in reports of unsuccessful attempts to find mAb-reactive peptides (Chapter 1). Although difficulties are discussed (*e.g.*, Lane & Stephen 1993; Burritt *et al.* 1996), there are no clear paths to success and finding the right strategy may require considerable experimentation. Thus, Wells (1996) described the strategy that yielded the earlier-reviewed (Chapter 1; Livnah *et al.* 1996; Wrighton *et al.* 1996) phage-displayed peptide able to activate the erythropoietin receptor as "try everything you can think of". The successful strategy, in considering the interaction of methodology with affinity and avidity, employed both high-valency (pVIII) and low-valency (pIII) libraries and several elution methods. In summary, both the current study and those of others (Chapter 1) suggest that routine employment of RPLs does not always enjoy those traits that were suggested by the success of early RPL papers, and which account for much of the utility of RPLs: technical simplicity, ease of use and assurance of results.

Chapter 6

Antigenic analysis of *Bordetella pertussis* filamentous hemagglutinin using phage display libraries and rabbit α-FHA polyclonal antibodies

6.1. ABSTRACT

Although important progress has been made in developing efficacious acellular vaccines to protect against disease caused by *Bordetella pertussis*, continued efforts to improve efficacy are expected to be hampered by a lack of understanding of the antigenic makeup of key *B. pertussis* virulence factors, including the multifaceted adhesin filamentous hemagglutinin (FHA). To identify antigenic regions of FHA, a series of phage display libraries were constructed using random fragments of the 10 kbp *Eco*RI fragment of *B. pertussis fhaB*. Affinity-selection with rabbit α-FHA polyclonal antibodies and subsequent characterization of antibody-reactive clones displaying FHA-derived peptides identified fourteen antigenic regions, each containing one or more epitopes. A number of reactive clones mapped within regions known to contain FHA adhesin domains. The corresponding epitopes may prove to be of protective value. This work also illustrates the potential utility of gene fragment and genomic phage display libraries in other applications.

6.2. INTRODUCTION

6.2.1. *Bordetella pertussis* vaccines

The gram-negative bacterium *Bordetella pertussis* is a human respiratory pathogen that is responsible for more than 355,000 deaths annually -- mostly of unimmunized young children in developing countries -- from the disease known as whooping cough, or pertussis (Cherry 1996). While efficacious whole cell vaccines (WCVs) against this organism have been used since the 1930s, the temporal association of the administration of pertussis WCVs with sudden

infant death syndrome (SIDS) and what was termed "pertussis vaccine encephalopathy" led to debate over vaccine safety (Brennan *et al.* 1992; Cherry 1996). Notwithstanding that several studies found no causative role of WCVs in neurological disease or SIDS, poor public acceptance and the relatively severe reactogenicity of WCVs provided incentive (Brennan *et al.* 1992; Cherry 1996) for the development of a number of acellular vaccines (ACVs).

B. pertussis produces a number of virulence factors that are released extracellularly or located at the outer surface of the bacterium and which are suitable candidates for inclusion in ACVs (Brennan *et al.* 1992; Edwards 1993; Rappuoli 1994; Brennan & Shahin 1996). Current ACVs include one or more of these factors. Purified toxoided or non-toxic genetically engineered pertussis toxin is a key component of all ACVs, for this dual-function adhesin/toxin is an important virulence factor, immunogenic in animals and humans, protective in animal models (Brennan *et al.* 1992; Brennan & Shahin 1996) and has been commonly believed to be the principal mediator of disease symptoms (but see Parton 1996). Purified filamentous hemagglutinin (FHA), an important multi-functional adhesin that is both cell-associated and secreted into the external milieu, is included in most ACV formulations and appears to improve efficacy, while another *B. pertussis* adhesin, pertactin, has been less commonly included (in one case unintentionally: Parton 1996). Systematic evaluation of ACVs in clinical trials (*e.g.*, Edwards *et al.* 1995; Decker & Edwards 1996) conducted to identify the most efficacious combinations of components have identified vaccines that are safe and less reactogenic than WCVs. Ongoing research (for reviews, see Brennan *et al.* 1992; Brennan & Shahin 1996) appears directed towards improving the long-term efficacy of ACVs by exploring alternative methods and routes of vaccine delivery and in furthering our currently poor understanding of the immune response to *B. pertussis* antigens. It seems likely that recombinant proteins comprised of various protective antigen subcomponents will form the basis of future vaccines. In such work, an understanding of the antigenic makeup of vaccine components is of fundamental

importance.

6.2.2. Role of FHA in pathogenesis

B. pertussis pathogenesis (reviewed in Weiss & Hewlett 1986; Rappuoli 1994; Brennan & Shahin 1996; Parton 1996; see also Fernandez & Weiss 1994) involves an almost redundant set of diverse adhesins and toxins. These include FHA, pertactin and BrkA (all of which contain RGD sequences), fimbriae, pertussis toxin (which functions as both toxin and adhesin), the bifunctional adenylate cyclase-hemolysin, tracheal cytotoxin and dermonecrotic toxin. Several of these have been incorporated into ACVs, including FHA, the focus of this study. The importance of FHA derives from its key role in adherence, for it possesses at least three adhesin domains that -- as reviewed below -- mediate *B. pertussis* attachment to, and in some cases uptake into, several host cell types. Although FHA has not been tested as a single-component ACV and its efficacy in humans is accordingly unknown, it has been shown to elicit protective immune responses in animal models (reviewed in Brennan *et al.* 1992; Brennan & Shahin 1996).

A model of *B. pertussis* infection adapted from Locht *et al.* (1993) -- and necessarily ignoring the role of other adhesion molecules -- serves as a useful frame of reference for illustrating the continued role of FHA throughout the infection cycle. Infection begins when aerosol-borne *B. pertussis* enter the upper airways and temperature-induced signalling through the BvgS/BvgA two-component regulatory system (Scarlato *et al.* 1993; Rappuoli 1994) leads to expression of the first of two temporally-separated groups of virulence factors. Not unsurprisingly the first group of products, FHA and fimbriae, mediate *B. pertussis* adherence to host cells. It is likely that an identified lectin-like domain of FHA which mediates binding to ciliated cells (and to macrophages; Prasad *et al.* 1993) is important at this stage. Interestingly, it has been shown that purified FHA mediates attachment to ciliated cells of not only *B. pertussis* but also certain strains of *Haemophilus influenzae* (Tuomanen 1986), suggesting that

secreted FHA can function as a bivalent bridge between host and bacterium.

Once cells are attached, the second temporally-expressed group of virulence factors -- pertussis toxin, tracheal cytotoxin and the bifunctional adenylate cyclase-hemolysin toxin -- mediate the local and systemic damage (Weiss & Hewlett 1986; Brennan *et al.* 1992; Rappuoli 1994) associated with pertussis disease. Since these toxins mediate substantial changes in the respiratory epithelium, the abilities of the FHA heparin-binding domain (Hannah *et al.* 1994; Menozzi *et al.* 1994) to bind to targets other than ciliated cells are likely important at this stage. Possible targets include common sulphated glycoconjugates of epithelial cell surfaces and of respiratory mucus.

The persistence of pertussis infection may be partly attributable to FHA, for the ability of *B. pertussis* to employ its FHA RGD motif to bind to CR3 integrins and enter macrophages (Relman *et al.* 1990; Saukkonen *et al.* 1991; see also Ishibashi *et al.* 1994) without triggering an oxidative burst would seem to allow immune system evasion and the establishment of an intracellular reservoir. Indeed, about half of the total bacterial population recovered from lungs in a rabbit model of infection appeared to reside within macrophages (Saukkonen *et al.* 1991). Binding activities of FHA that facilitate HeLa cell invasion may also contribute to persistence of infection (Ewanowich *et al.* 1989)

As expected for a molecule with multiple adhesin domains, FHA is large and complex (Locht *et al.* 1993; Makhov *et al.* 1994). Synthesized as a 367 kDa precursor (FhaB), it is translocated to the periplasm by a process that may involve the Sec apparatus and exported through the outer membrane by an accessory protein, the product of *fhaC* (Arico *et al.* 1993; Jacob-Dubuisson *et al.* 1996; Renauld-Mongenie *et al.* 1996). N-terminal processing appears to remove the first 71 residues (Jacob-Dubuisson *et al.* 1996) and subsequent cleavage of the C-terminal third of FhaB yields the 220 kDa mature FHA molecule (Arico *et al.* 1993). Although the role of the cleaved C-terminal region is not known, it has been suggested (Renauld-Mongenie

et al. 1996) that it serves as an intramolecular chaperone preventing premature folding of the mature FHA protein into its proposed (Makhov *et al.* 1994) mature rod-like conformation.

Of the three identified adhesin domains, that of the RGD triplet ($FHA_{1097-1099}$) was readily located within the primary amino acid sequence of FHA (Relman *et al.* 1989). The heparin-binding domain has recently been found to be within (Hannah *et al.* 1994) a 422-residue span ($FHA_{442-863}$), and the lectin-like binding domain has been mapped (Prasad *et al.* 1993) to a 139-residue region ($FHA_{1141-1279}$). Comparison of the sequences containing the lectin-like binding domains of pertussis toxin S2 subunit and FHA suggested that the 19-residue sequence $FHA_{1224-1242}$ may contain residues important in binding (Prasad *et al.* 1993). As suggested by several studies, these adhesin domains represent candidates for inclusion in recombinant subcomponent vaccines.

Sequence comparisons have also identified regions with homology to molecules that, like the FHA RGD motif, interact with the leukocyte integrin CR3 (Sandros & Tuomanen 1993). These are (i) $FHA_{1407-1417}$, which is homologous to C3bi, and (ii) $FHA_{1979-1984}$ and $FHA_{2062-2068}$, which are apparent mimics of functional regions of the coagulation component Factor X (Rozdzinksi *et al.* 1995). To the extent that these molecules truly mimic C3bi and Factor X, we can assume that they do so for the benefit of *B. pertussis*. For example, peptides derived from the Factor X homologs of FHA were shown to inhibit Factor X binding to neutrophils and prolong clotting time, and to prevent transendothelial migration of leukocytes *in vitro* (Rozdzinksi *et al.* 1995). Thus, *prima facie*, antibodies against these mimics would be of value. On the other hand, in an apparent extension of this mimicry, mAbs that recognize (i) $FHA_{1141-1279}$, which contains the earlier-described lectin-like binding domain, and (ii) FHA $_{2013-2110}$ which includes one of the Factor X homologs, have been shown to bind to cerebral microvessels, to interfere with transmigration of leukoctyes into cerebrospinal fluid, and (for one of these mAbs) to induce a dose-dependent reversible increase in permeability of the

blood-brain barrier (Tuomanen *et al.* 1993). While the implications of these interactions to pertussis disease are not known, inclusion of these regions in recombinant vaccines may be undesirable (Tuomanen *et al.* 1993).

6.2.3. Rationale for this study

The antigenic makeup of FHA has been the focus of only a single, limited report which identified the approximate locations of four epitopes recognized by nine mAbs within a 1200-residue span (Delisse-Gathoye *et al.* 1990) of the ~2165-residue mature FHA molecule. Yet the importance of FHA as an adhesin makes it an important subject for antigenic analysis. Given this, and that FHA is encoded by a relatively large (10.8 kbp; Relman *et al.* 1989; Domenighini *et al.* 1990) gene and can thus serve as a non-trivial model for genomic PDLs, rabbit α-FHA polyclonal antibodies were employed to affinity-select and characterize antibody-reactive clones from phage libraries displaying peptides encoded by *fhaB*-derived gene fragments with the goals of (i) providing an antigenic analysis of FHA and (ii) demonstrating the feasibility of employing genomic PDLs in other applications.

Table 6-I. Peptides displayed by FHA-70 and -80 libraries.

Library[a]	Expected size of displayed peptide (amino acid residues)
70-A, 80-A	10-25
70-B, 80-B	25-50
70-C, 80-C	50-100
70-D, 80-D	100-200
70-X, 80-X, 80-Y	no insert

[a]The "70" and "80" libraries were constructed with vector fDRW70 and vector series fDRW8*nn*, respectively. Details of library construction are presented in Chapter 3.

6.3. RESULTS

6.3.1. Library construction and evaluation

The construction and evaluation of two sets of libraries (FHA-70 and FHA-80)

displaying peptides encoded by fragments of the 10 kbp *Eco*RI fragment of *B. pertussis fhaB* are described in Chapter 7 (§7.3.9, §7.3.10), which reviews issues in library construction. The use of multiple libraries (Table 6-I), each displaying peptides derived from a specific size range of DNase I-generated fragments, reflected a strategy (§7.3.7) intended to maximize library diversity and minimize clonal competition. Experiments intended to assess sequence diversity and other properties of the FHA-70 libraries were inconclusive (§7.3.9.b, §7.3.9.c) and no attempt was made to assess the FHA-80 libraries (§7.3.10).

Target clones reactive with rabbit α-FHA polyclonal antibodies were selected from these libraries by a single round of biopanning with four 10-fold serially different quantities of pAbs (§7.3.11). That biopanning was successful was indicated by recovery of a greater number of virions from libraries panned with 3.6 μg and 360 ng of pAbs than from control libraries or libraries panned with smaller quantities of pAbs (Chapter 7, Figure 7-30). Success in affinity selecting target clones from FHA-70 libraries was also confirmed in related assays (Chapter 7, Figures 7-31 and 7-32).

6.3.2. Selection of clones for characterization

(a) Choice of antibodies. α-FHA pAbs were purified from three rabbit sera obtained

---

Table 6-II. Summary of antibody-reactive clones selected[a] from FHA-70 libraries.

Library	Dilution of FN2/4 pAbs (*E. coli*-absorbed, protein A-purified)	Number of clones selected
70-A	1 : 8,000	15
	1: 32,000	11
70-B	1: 8,000	5
70-C	1: 128,000	13
70-D	1: 32,000	12
	Total number of clones selected	56

[a]As described in the text, clones were selected from plaque lifts that employed, for each library, the greatest antibody dilution. Selected plaques were those yielding the greatest color intensity. The fractions of positive *versus* total plaques at each antibody dilution are illustrated in Figure 7-32 (Chapter 7).

with two preparations of purified FHA (Materials and Methods, §2.13.3). A preparation of "native" FHA was used to raise sera FN1/4 and FN2/4, while a preparation of "SDS-denatured" FHA was used to raise FS1/4. To maximize the numbers of target clones recovered, biopannings (§7.3.11) employed a pool of *E. coli*-absorbed, Protein A-purified FN2/4 and FS1/4. Because preliminary plaque lifts suggested that virtually all of the affinity-selected clones were recognized by antibody, a smaller number of clones were targeted in later plaque lifts that employed FN2/4 alone, and clones were chosen from among those that reacted most strongly with antibodies as judged by color intensity at a given antibody dilution or by any visible reactivity at a high antibody dilution.

(b) Selection of FHA-70 library clones. Using these criteria, a total of 56 antibody-reactive clones (Table 6-II) were selected from plaque lifts of biopanning output from FHA-70 libraries (see Chapter 7, Figure 7-32). These clones were purified by several rounds of plaque purification involving, for each round, plaque excision and preparation of new plaque lifts. During this procedure, five antibody-reactive clones were lost. Whether this was due to insert instability, incorrect excision of a non-reactive plaque or other mishap was not determined.

Table 6-III. Summary of antibody-reactive clones identified in FHA library 80-A by plaque lifts.

Quantity of pAbs[a] employed in biopanning	Dilution of pAbs[b] used in plaque lifts	Number of antibody-reactive (*r*) / total (*t*) plaques: r / t (%)	Number of clones chosen for characterization
3.6 µg	1 : 8,000	107 / 357 (30%)	0
360 ng		111 / 321 (35%)	
3.6 µg	1: 32,000	54 / 534 (10%)	58
360 ng		45 / 456 (10%)	
3.6 µg	1 : 128,000	0 / 379 (0%)	0
360 ng		0 / 321 (0%)	

[a]equimolar pool of *E. coli*-absorbed protein A-purified FN2/4 and FS1/4 rabbit α-FHA pAbs.
[b]protein A-purified FN2/4 pAbs.

(c) Selection of FHA-80-A clones. Plaque lifts (Table 6-III) of biopanning output (Figure 7-30, Chapter 7) from the small-fragment library 80-A were performed to identify an additional 58 antibody-reactive clones for characterization. Using the criteria described earlier for identifying highly-reactive clones, clones were chosen from plaque lifts developed with a 1:32,000 dilution of purified FN2/4 pAbs (Table 6-III). Clonal purification was accomplished by limiting dilution transduction of virions from excised plaques (Materials and Methods, §2.15.3.b). As described later, three of these antibody-reactive clones were subsequently lost.

6.3.3. Sequencing

(a) Agreement with published sequence data. Sequences of most cloned *fhaB*-derived inserts agreed with the published sequence[9], and in particular there was no evidence of insertions or deletions. Single base substitutions were found in three clones (clones I-a, XI-c and XI-d), and a single base substitution was found in a vector sequence, at the cloning site; these are identified in Figure 6-1. Two clones contained inverted sequences. The first of these (I-b) additionally contained a second non-inverted fragment encoding an FHA peptide, while the second ("clone 30") encoded a peptide unrelated to FHA. This latter clone was employed as a control in later assays.

(b) Apparent importance of FHA regions I, III and XI. Sequencing showed that the 51 FHA-70 and 58 FHA-80-A clones encoded 31 unique vector-FHA peptide sequences and that the displayed peptides could be mapped to fourteen regions of the primary amino acid sequence of FHA (Figures 6-1 and 6-2). More than half of the clones mapped to region I. Within this region, a 126-residue peptide (I-a) was represented by 11 siblings, while a shorter 11-residue peptide (I-d) was represented by 22 siblings. Relatively high numbers of clones were also found for regions III (7 siblings for each of 2 unique clones) and XI (8 siblings for clone XI-a, 9 for XI-d). These findings suggested that regions I, III and XI contained

---

[9]based on comparison with *B. pertussis fhaB* gene and translated FHA protein sequences in GenBank entry M60351.

Figure shown on following page

Fig. 6-1. Sequences of antibody-reactive FHA-70 and -80 clones. **A**. Nucleotide and implied amino acid sequences of *B. pertussis fha*B-derived inserts and flanking vector sequences of 110 clones selected, by means of plaque lifts employing α-FHA pAbs, from libraries 70-A, -B, -C, -D and 80-A. **B.** Position of each unique clone within the 3,591-residue *fhaB* gene product (FHA).

[a]The number of clones identified that were found to share a common combination of vector- and *fha*B-derived sequence, *i.e.,* the number of "siblings" of each *unique* clone.

[b]Nucleotide sequences have been abbreviated for convenience; numbers enclosed in square parentheses "{*n*}" indicate the numbers of bases omitted. All omitted bases correspond to the *fhaB* sequence of GenBank accession number M60351. Upper case letters identify sequences identical to the published *fhaB* sequence; these include some vector-derived bases that adventitiously correspond. Lower case letters identify sequences that do not correspond to the published *fhaB* sequence; although most of these are vector sequences, certain of these (clones I-a, XI-c and XI-d) are *fha*B-derived.

[c]Starting position within the published *fhaB* sequence of the nucleotide sequence shown in upper case in the adjacent column ("Sequence of *fha*B-derived insert ....").

[d]Amino acid sequence derived from the nucleotide sequence shown in the column "Sequence of *fha*B-derived insert ....". Sequences in upper case letters are identical to those derived from the published *fhaB* nucleotide sequence; these include some vector-derived residues that adventitiously correspond. Lower case letters identify sequences that do not correspond to the published *fhaB* sequence. These derive from nucleotide sequence mismatches; see note *b* above.

[e]Starting position, within the 3,591-residue *fhaB* gene product, of the amino acid sequence shown in upper case in the adjacent column ("Encoded peptide ....").

[f]The *fha*B-derived fragment of "clone 30" is inserted in the opposite orientation to that required for production of an FHA-related peptide. This clone, although initially selected by virtue of its apparent recognition by α-FHA pAbs, was employed as a negative control in ELISA and immunoblots.

A

Clone ID	Sib-lings[a]	Sequence of *fha*B-derived insert and flanking vector bases[b]	Posi-tion[c]	Encoded peptide and flanking residues[d]	Posi-tion[e]
I-a	11	GGTGCTGGTGCGCTGTCGGTAAGCGCTGGCGGAAACCTGC GGGCGAAgcAATTG..[328]..CGActggcgcaggtgct	1732	gaGALSVSAGGNLRAkqLVSS AQLEVRGQREVALDDASSARGMTVVAAGALAARNLQSKGAIGVQG GEAVSVANANSDAELRVRGRGQVDLHDLSAARGADISGEGRVNIG RARSDSDVKVSAHGALStgagaga	494
I-b	4	agcccgatcgcgccgcgcgggtggcgggcgacaactatttt cgCAGGTGGATCTG..[33]..TCCGGCGgcgctggatct	1987	spiaprgwrattisQVDLHDLSAARGADISGgags	579
I-c	3	ggcccCGACCTGAG..[31]..GGGCGgggcgctggatct	2001	gpDLSAARGADISGEGRgags	584
I-d	22	ggcccCCTGAGCGC..[19]..TCCGGgggcgtcgggact	2004	gpLSAARGADISGgvgt	585
I-e	4	ggCCTGAGCGCAGC..[16]..TCCGGgggcgctggatct	2004	gLSAARGADISGgagsga	585
I-f	7	ggcccCCTGAGCGC..[28]..GGGCGgggcgttgggtct	2004	gpLSAARGADISGEGRgvgs	585
I-g	2	ggcccCCTGAGCGC..[28]..GGGCGgggcgctggatct	2004	gpLSAARGADISGEGRgags	585
I-h	3	ggcccGAGCGCAGC..[22]..GAGGGGggcgccggtact	2007	gpSAARGADISGEGgagt	586
II	1	ggcccCGCCTTGTC..[16]..GCCCTCGGTGCGggcgct	2100	gpALSIDSMTALGAga	617
III-a	7	ggtgctggtGCCAT..[52]..GCGTctggcgcaggtgct	2272	gAGAMTVRDVAAAADLALQAGDAsgaga	672
III-b	7	ggtgctggtGCCAT..[55]..TTGCctggcgcaggtgct	2272	gAGAMTVRDVAAAADLALQAGDALpgaga	672
III-c	2	ggcccGCGAGACGT..[22]..CTGCAGGgcgtcgggact	2283	gpRDVAAAADLALQgvgt	678
IV	2	ggcCCGATCGTGGT..[25]..GCCGGgggcgctggatct	3940	GPIVVEAGELVSHAGGags	1229
V	1	ggccCCACCAAAGG..[31]..TCTCCGggcgctggatct	4262	gpTKGEMQIAGKGGGSPgags	1338
VI-a	1	ggcccCGATTACAC..[22]..CTGGCGGgcgtcgggact	4467	gpDYTVSADAIALAgvgt	1406
VI-b	1	ggcccCACCGTCTC..[22]..GCGCAGGgcgccggtact	4473	gpTVSADAIALAAQgagt	1408
VII	1	ggccCCAACAAGAT..[16]..TTGCAGGgcgttgggtct	4565	gpNKIRLMGPLQgvgs	1439
VIII	3	ggtgctggtgcTAT..[52]..AAGGctggcgcaggtgct	4827	gagaITVTSRGGFDNEGKMESNKagaga	1526
IX-a	1	ggcCCGCACCTGCG..[22]..GGGCAgggcgtcgggact	5077	gPHLRNTGQVVAGqgvgt	1609
IX-b	1	ggcccGGTCGTAGC..[16]..ATCATgggcgtcgggact	5100	gpVVAGHDIHImgvgt	1617
X-a	1	ggcccCAACAAGCT..[19]..GAGGACATGCAgggcgct	5595	gpNKLGRIRAGEDMqga	1782
X-b	1	ggcccCAACAAGCT..[46]..ATCGAGggcgtcgggact	5595	gpNKLGRIRAGEDMHLDAPRIEgvgt	1782
XI-a	8	ggtgctggtgcGCT..[34]..TATCCtggcgcaggtgct	6102	gagaLDQNRYEYIWGLYPgaga	1951
XI-b	1	ggcccGGACCAGAA..[25]..TATCagggcgttgggtct	6105	gpDQNRYEYIWGLYqgvgs	1952
XI-c	1	ggtgctggtgcCAA..[99]..GTATCaGACCTACACCGA ATGGTCGGTGAATA..[132]..ACActggcgcaggtgct	6036	gagaNRAGLSPATWNFQSTYELLDYLLDQNRYEYIWGLYqTYTEW SVNTLKNLDLGYQAKPAPTAPPMPKAPELDLRGHTLESAEGRKIF GEYtgaga	1929
XI-d	9	ggtgctggtgcCGAGTACATTTGGGGGCTGTATCaGACCT ACACCGAATGGTCG..[166]..AGGCtggcgcaggtgct	6120	gagaEYIWGLYqTYTEW SVNTLKNLDLGYQAKPAPTAPPMPKAPELDLRGHTLESAEGRKIF GEYKKLQGEYEKAgaga	1957
XI-e	1	ggtgctggtgcCAC..[155]..AGCctggcgcaggtgct	6150	gagaTEW SVNTLKNLDLGYQAKPAPTAPPMPKAPELDLRGHTLESAEGRKIF GEYKKpgaga	1967
XII	1	ggtgctggtgcCGT..[194]..GGCctggcgcaggtgct	6381	gagaVHDQLGQRYGKALGGMDAETKEVDGIIQEFAADLRTVYAKQ ADQATIDAETDKVAQRYKSQIDAVRpgaga	2044
XIII	1	ggtgctggtgcCGG..[64]..TTGCctggcgcaggtgct	6819	gagaGRPEGLKIGAHSATSVSGSFDALpgaga	2190
XIV	1	ggtgctggTGCGCT..[79]..CCGGCtggcgcaggtgct	6927	gagALAAVLVNPHIFTRIGAAQTSLADGAAGPAgaga	2226
clone 30	1	ggtgctggtgCAAC..[53]..CCCCctggcgcaggtgct	1333[f]	gagatlladaiftapppatftappgagaga	Note f

B

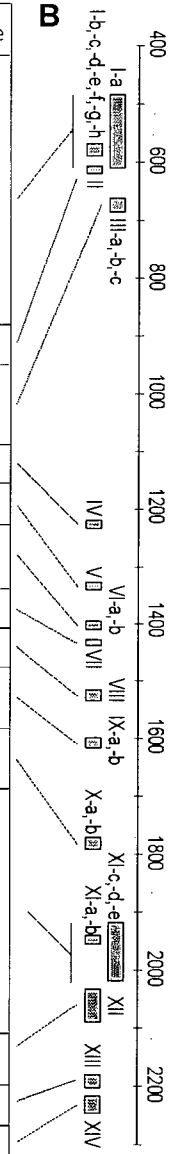


Fig. 6-1.

important antigenic determinants of FHA.

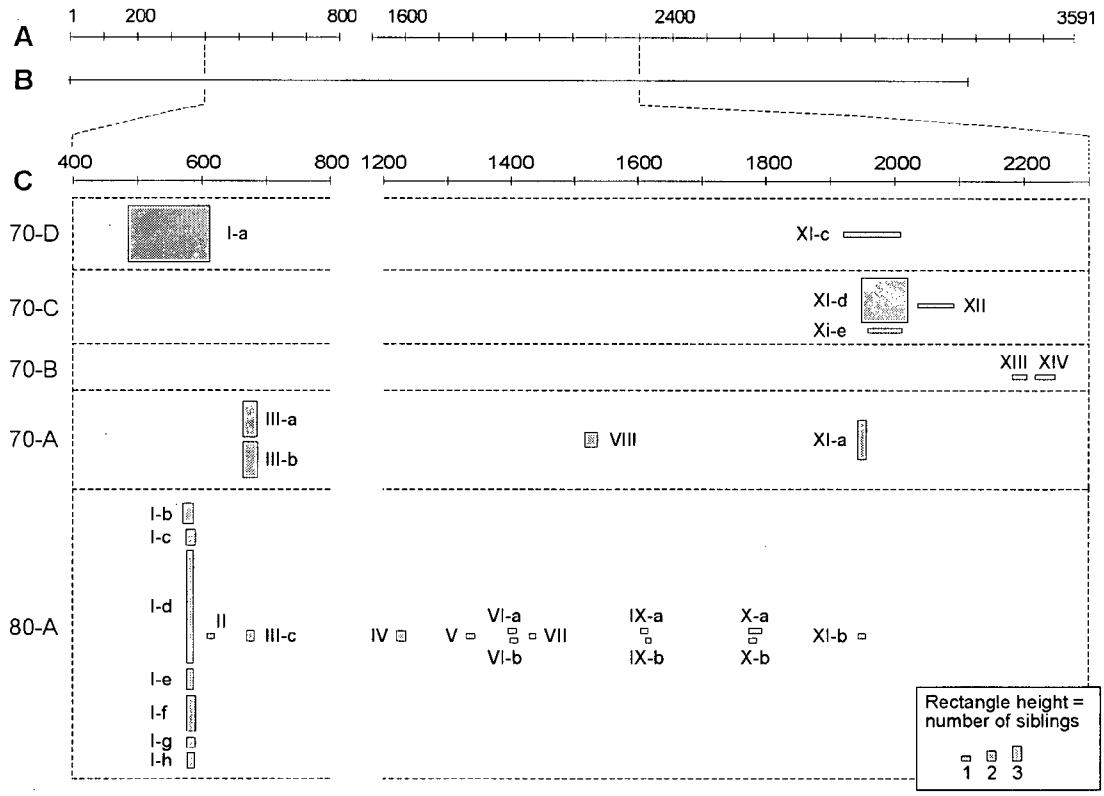


Fig. 6-2. Distribution of antibody-reactive FHA-70 and -80 clones. **A.** The 3,591-residue *B. pertussis fhaB* gene product (FhaB). Mature FHA represents the N-terminal two-thirds of this sequence, nominally 2,165 residues (Makhov *et al.*, 1994). **B.** Region included in the 10 kbp *Eco*RI restriction fragment employed in FHA-70 and -80 library construction. **C.** Positional distribution of sequences of antibody-reactive clones selected from libraries 70-A, -B, -C, -D and 80-A. Each unique clone (as defined by the combination of vector- and *fha*B-derived sequences) is represented by a filled rectangle. The horizontal position and width identifies the location of the encoded peptide within the primary amino acid sequence of FHA, while the height represents the number of siblings of a unique clone; *viz.*, the number of clones that were found to share a common vector- and *fha*B-derived sequence.

6.3.4. Preliminary ELISA

(a) FS1/4 and FN2/4 recognize FHA differently. ELISA of the selected clones with FS1/4 and FN2/4 (Figure 6-3B) confirmed the idea that regions I, III and XI contained important antigenic determinants of FHA, and additionally suggested that regions XII and XIII were important. Notably, FS1/4 and FN2/4 recognized different groups of clones, suggesting that the nature of the FHA immunogens used to elicit these antibodies may be important. This is discussed later.

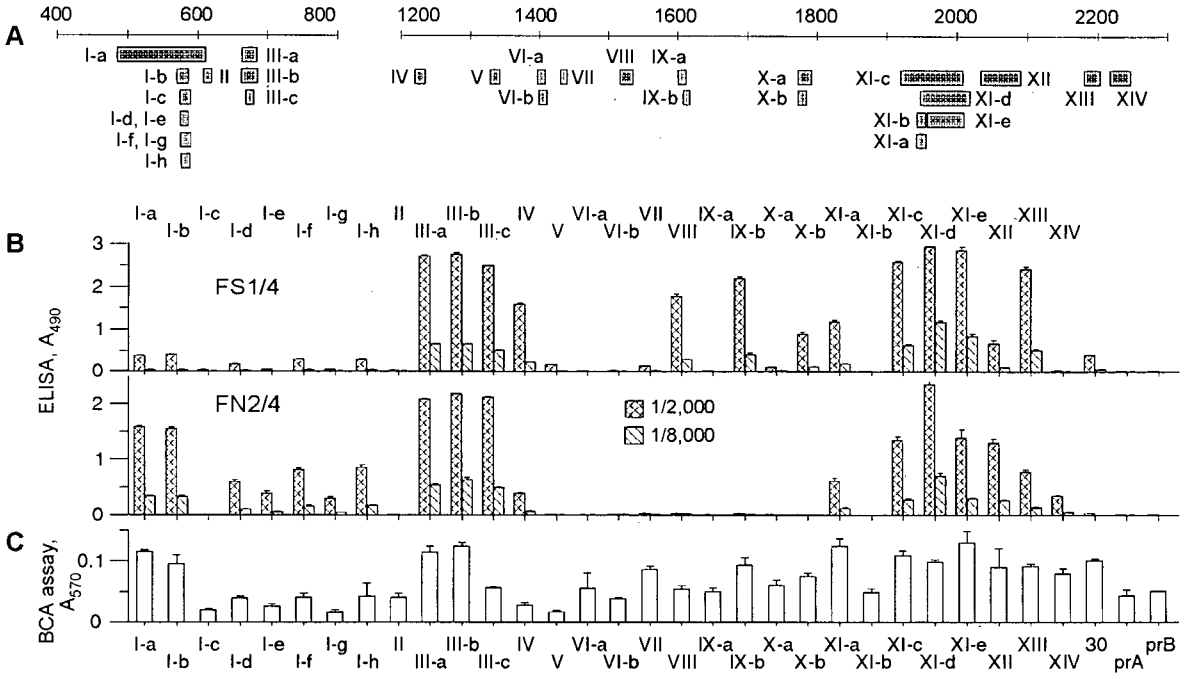


Fig. 6-3. Early ELISA of FHA-70 and -80 clones. A. Position of the peptide encoded by each unique clone (identified by Roman numerals + optional lower case letter) within FHA. **B.** Summary of concurrent ELISAs of two-stage PEG-precipitated virions (single clones or pools of siblings for a unique clone) probed with the indicated dilutions of *E. coli*-absorbed FN2/4 and FS1/4 α-FHA pAbs. Each well was coated with the same nominal quantity (1 μg) of total virion protein. **C.** Concurrent BCA protein assay performed with the goal of demonstrating uniform binding of all clones to microtiter plate wells. **B, C.** As controls for variability, two siblings of each of clones I-b, I-e, III-c and XI-e were assayed. Results are shown only for the sibling with the greater binding to wells and the correspondingly greater ELISA signal. For I-b, ELISA and BCA values varied up to 6-fold; for I-e, 1.5-fold; III-c, ~20-fold; XI-d, ~1.4-fold. This variability is discussed in the text.

Values shown are means of triplicate wells ± 2 standard errors. Clone 30 (see legend to Figure 6-1) was employed as a negative control. Negative control clones "prA" and "prB" are pseudo-revertants of vector fDRW70 and accordingly do not display an *fha*B-derived peptide. n.d., not determined.

(b) Lack of uniformity in binding of FHA library clones to Immulon plates. An earlier study (Chapter 4) of variants of the *P. falciparum* circumsporozoite protein tetrapeptide repeats had shown that CsCl-purified virions bound uniformly (clone-to-clone variability was low) to wells of Immulon plates. Although CsCl preparations serve to remove a relatively large quantity of non-virion material, they are somewhat impractical for more than a small number of clones. Thus, the ELISAs shown in Figure 6-3 were performed with virions purified only by precipitation with polyethylene glycol. Regrettably, up to 8-fold differences in binding were

seen among samples (Figure 6-3C) and there appeared to be some correlation between ELISA values and the quantities of virions bound to wells (Figure 6-3, compare B and C). Importantly, ELISA values could be reduced by preincubating α-FHA pAbs with purified FHA (data not shown; a similar experiment is described later), and there was no significant reactivity between virions and secondary antibodies (data not shown). Nevertheless, the apparent correlation of ELISA values with binding did suggest that clone-to-clone differences in ELISA values, or lack of ELISA signal, may have been in part artefactual.

(c) Differences in binding may reflect differences among displayed peptides. As explained in the legend to Figure 6-3, differences in binding among siblings of the same clone suggested that these differences reflected a lack of sample-to-sample uniformity in methods of virion purification. However, other data suggested that differences in the nature of the displayed peptides also accounted for lack of uniformity of binding. Specifically, an assay of binding patterns of selected clones (Materials and Methods, §2.16.4.c, data not shown; a later assay is shown in Figure 6-7) showed that patterns of binding were similar for peptides within a given region, but significantly different from region to region. Interpretable comparisons of FHA library clones thus required both care in sample preparation and an alternative method of immunocharacterization.

(d) Uniform binding of FHA library clones to nitrocellulose. Accordingly, one or more siblings of each antibody-reactive clone were propagated, harvested and purified by procedures in which exceptional care was taken to minimize sample-to-sample differences in preparation (Materials and Methods, §2.15.4). When these preparations were assayed for binding to nitrocellulose by an immunoblot employing α-f1 pAbs, binding varied over a less than two-fold range (Figure 6-4), a clear improvement over the eight-fold differences seen with earlier preparations and Immulon plates.

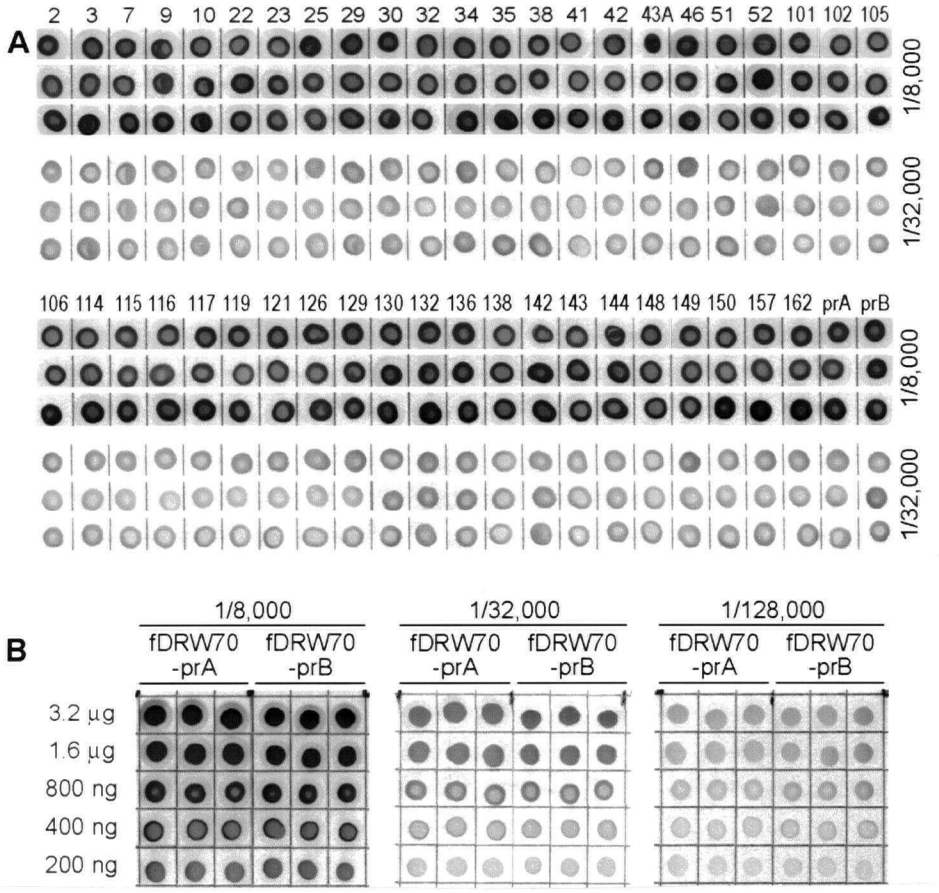


Fig. 6-4. Uniform binding of FHA-70 and -80 clones to nitrocellulose. **A.** Triplicate ~2 µL samples (800 ng protein) of virions of each of the indicated clones and of vector fDRW70 pseudorevertants A and B ("prA" and "prB") were probed with the indicated dilutions of protein A-purified rabbit α-f1 pAbs and alkaline phosphatase-conjugated 2° Abs to demonstrate uniformity of virion binding to nitrocellulose. Samples were applied according to the template shown in Figure 6-4; scanned computer images of the blots were "cut and pasted" into the format shown here. **B.** Triplicate 2 µL samples of fDRW70 pseudorevertant virions were probed with α-f1 pAbs to demonstrate titerability of immunoblot signal.

6.3.5. Immunoblots and ELISA

Accordingly, immunoblots probed with FS1/4, FN2/4 as well as FN1/4 (a third serum, not employed in biopanning or screening) were used in subsequent analyses. One such blot is shown in Figure 6-5. A composite immunoblot -- constructed by cutting and pasting sections from these blots with computer graphics software -- is shown in Figure 6-6. Importantly, variability among triplicate samples and among siblings was minimal and all but three (non-reactive) clones showed titerable reactivity with antibodies. The three non-reactive clones were fDRW8*nn* recombinants and the reactivity of these had not been confirmed after clonal

purification by limiting dilution transduction of virions from excised plaques. As with the five earlier-described initially-reactive fDRW70 clones that were lost, loss of reactivity of the fDRW8*nn* clones may have been due to mishap or insert instability.

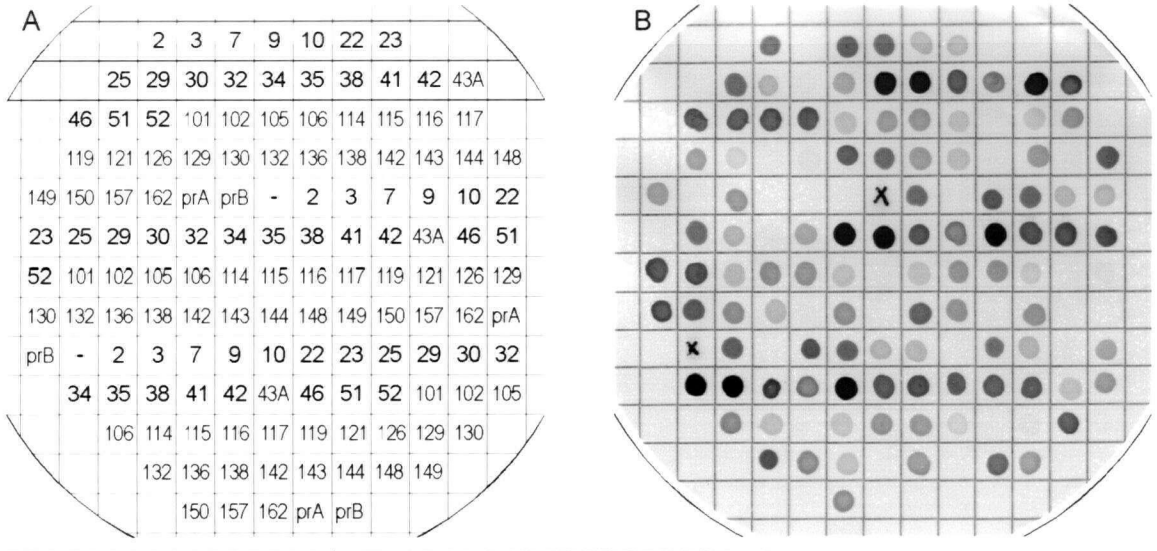


Fig. 6-5. Example immunoblot of FHA-70 and -80 clones. Triplicate 2 µL (800 ng virion protein) samples of each indicated clone (numbered 2 to 162) and controls (fDRW70 pseudo-revertants "prA" and "prB") were applied in the pattern shown in **A**. Goal of this pattern was to demonstrate that the position of a sample did not, as was seen in some early blots, influence color intensity. The blot illustrated in **B**, derived from probing FHA library clones with a 1:8,000 dilution of FN2/4 antibodies, is included in Figure 6-6.

(a) Patterns of recognition. The patterns of recognition of FHA library clone by FS1/4, FN2/4 and FN1/4 were striking (Figure 6-6), reflecting marked differences between regions (*e.g.*, region I *versus* III) and more subtle differences within regions (*e.g.* clones I-a and I-b *versus* I-c to I-h). These differences are discussed later. Notably, in spite of concerns about artefacts in the earlier ELISAs, immunoblots yielded results qualitatively similar to those obtained earlier with ELISA.

(b) Confirmation of variable binding to Immulon plates. ELISA was preferred over immunoblots in a "competition" assay (described in the following section) intended to demonstrate specificity of interaction between displayed peptides and antibodies. Because new virion samples had been prepared in a manner that may have reduced variability of binding to

Figure shown on following page

Fig. 6-6. Immunoblots of antibody-reactive FHA-70 and -80 clones. **A**. Relative position within the 3,591-residue *B. pertussis fhaB* gene product (FHA) of each of the antibody-reactive clones assayed by means of immunoblots with α-FHA pAbs and serum. **B**. Thirty unique clones (I-a, I-b, ..., XIV) were assayed. In many cases (*e.g.*, clone ID I-a) more than one sibling (*e.g.*, clone No. 43A, 46 and 52) were assayed to assess sample-to-sample (sibling-to-sibling) variability. Controls included clone No. 30 (see legend to Figure 6-1) and two pseudorevertants of fDRW70, shown as "prA" and "prB". For each clone or control, triplicate 2 μL samples (800 ng protein) applied to nitrocellulose were probed with the indicated dilutions of (a) *E. coli*-absorbed FS1/4 and FN2/4 α-FHA pAbs, crude FN1/4 α-FHA serum, (b) as a control to assess virion quantities bound to nitrocellulose, protein A-purified α-f1 serum, and (c) no 1° antibody, as a control for recognition of virions by 2° antibody alone. Only the first of the triplicates are shown for these latter controls. Samples were applied according to the template shown in Figure 6-5. After blot development, scanned computer images were "cut and pasted" into the format shown here. Two sets of non-concurrent immunoblots, (i) and (ii), are shown. Immunoblot set (i) was performed concurrently with that shown in Figure 6-5.

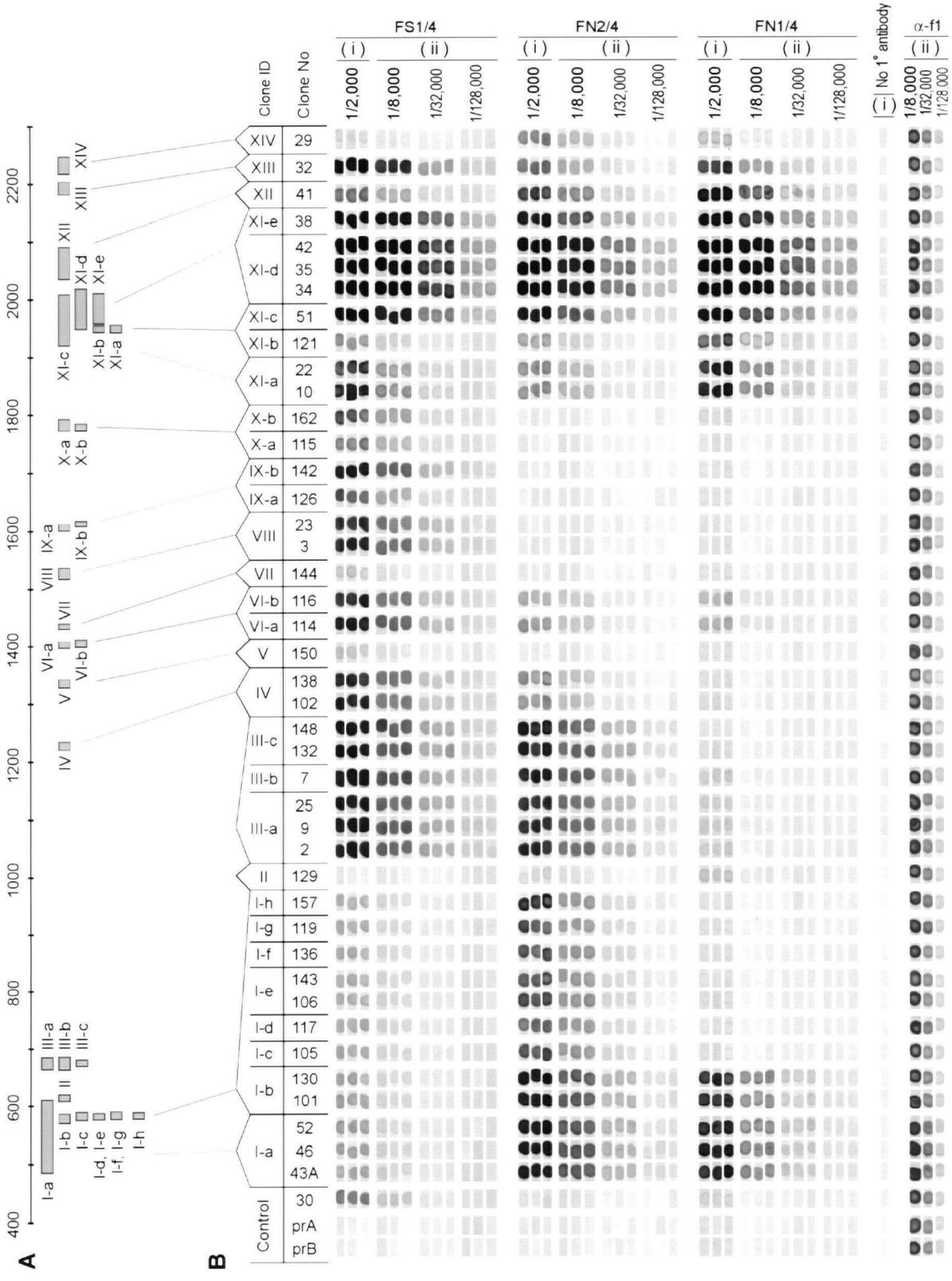

A
B
400
600
800
1000
1200
1400
1600
1800
2000
2200
I-a
I-b
II
I-c
I-d, I-e
I-f, I-g
I-h
III-a
III-b
III-c
IV
V
VI-a
VI-b
VII
VIII
IX-a
IX-b
X-a
X-b
XI-c
XII
XI-d
XI-e
XI-b
XI-a
XIII
XIV
Clone ID
Clone No
FS1/4
FN2/4
FN1/4
(i)
(ii)
1/2,000
1/8,000
1/32,000
1/128,000
(-) No 1° antibody
α-f1
XIV 29
XIII 32
XII 41
XI-e 38
XI-d 42 35 34
XI-c 51
XI-b 121
XI-a 22 10
X-b 162
X-a 115
IX-b 142
IX-a 126
VIII 23 3
VII 144
VI-b 116
VI-a 114
V 150
IV 138 102
III-c 148 132
III-b 7
III-a 25 9 2
II 129
I-h 157
I-g 119
I-f 136
I-e 143 106
I-d 117
I-c 105
I-b 130 101
I-a 52 46 43A
Control 30 prA prB

Fig. 6-6

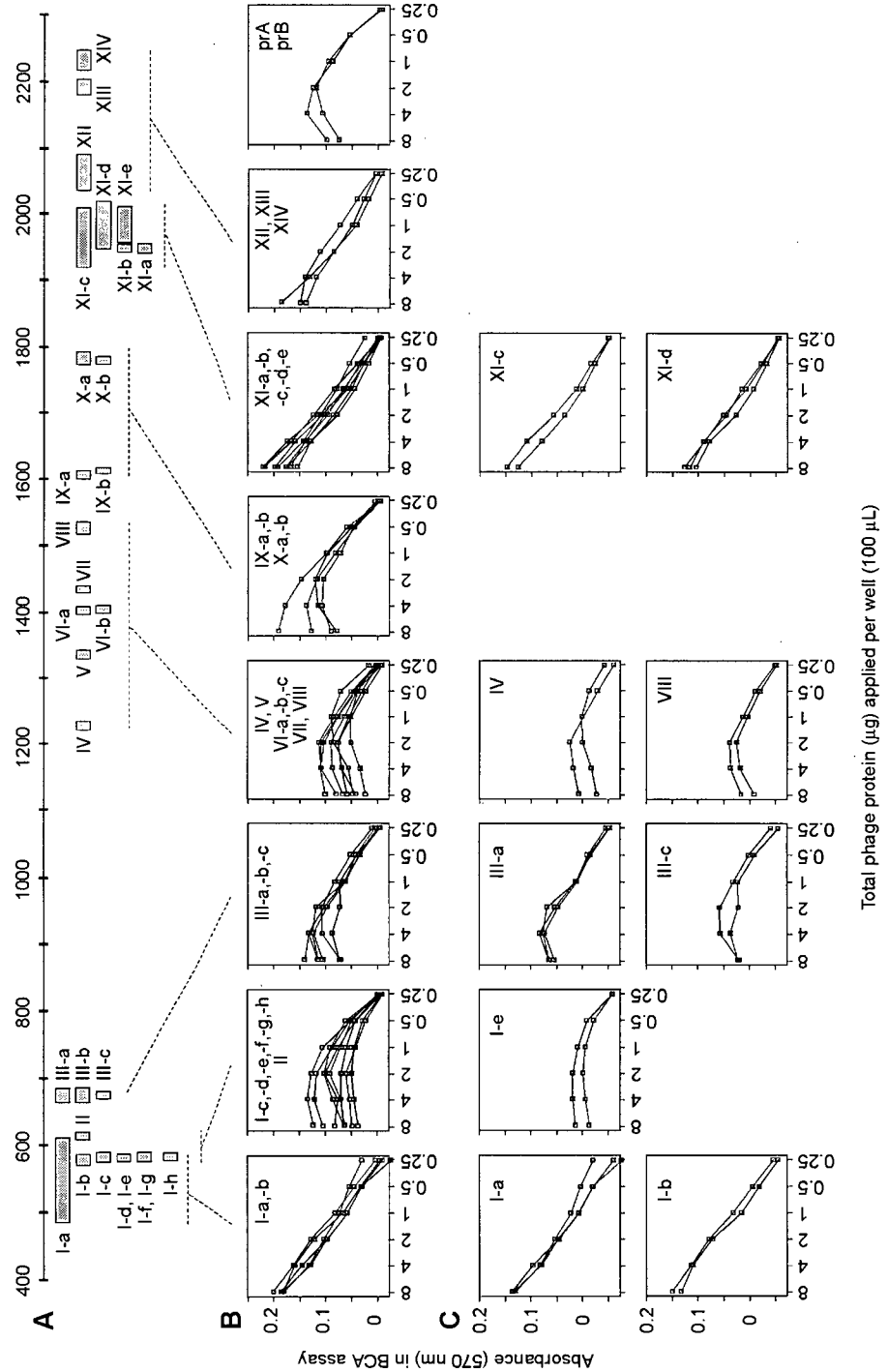


Fig. 6-7. Variable binding of FHA-70 and -80 clones to Immulon plates. **A**. Position within FHA of the peptide encoded by each unique clone. **B**. BCA protein assay performed for virions of 44 siblings of 30 unique clones as well as two pseudorevertants of fDRW70 ("prA" and "prB"). Assay was performed to determine quantities of virions remaining bound to wells of Immulon-2 plates, after wells were coated with the indicated quantities of PEG-precipitated and filtered virions and washed in the manner employed for ELISA. Clones have been grouped by FHA region and by pattern of binding. Values shown are means of duplicate wells. **C**. These graphs, included in larger groupings in **B**, are intended to show sibling-to-sibling reproducibility for two or more siblings of a unique clone

Figure shown on following page

Fig. 6-8. Inhibition of ELISA signal by preincubation of α-FHA antibodies with purified FHA. **A**. Relative position within the 3,591-residue *B. pertussis fhaB* gene product (FHA) of each of the antibody-reactive clones assayed by ELISA with α-FHA pAbs.

B. Thirty unique clones (I-a, I-b, ..., XIV) were assayed. In some cases (*e.g.*, clone ID III-a) more than one sibling (*e.g.*, clone No. 2, 9 and 25) were assayed in order assess sample-to-sample variability. Control phage included clone No. 30 (see legend to Figure 6-1) and a pseudorevertant of vector fDRW70, shown as "prA". Each clone was probed with *E. coli*-absorbed FN2/4 and FS1/4 pAbs. These were prepared by preincubating 1:5,000 dilutions of antibodies with heparin-sepharose affinity-purified FHA at final concentrations of 0, 0.8, 5 and 30 μg mL^{-1} for 2.75 h at 37°C and subsequently diluting these mixtures two-fold immediately before use in ELISA. Values shown are means of duplicate wells.

C. In a concurrent assay, virions bound to plates after washing as in ELISA were assayed by a BCA protein assay. Values show are means of triplicate wells ± two standard errors.

D. For comparison, the first of the three triplicate samples of the immunoblots shown in Figure 7-7 are shown here.

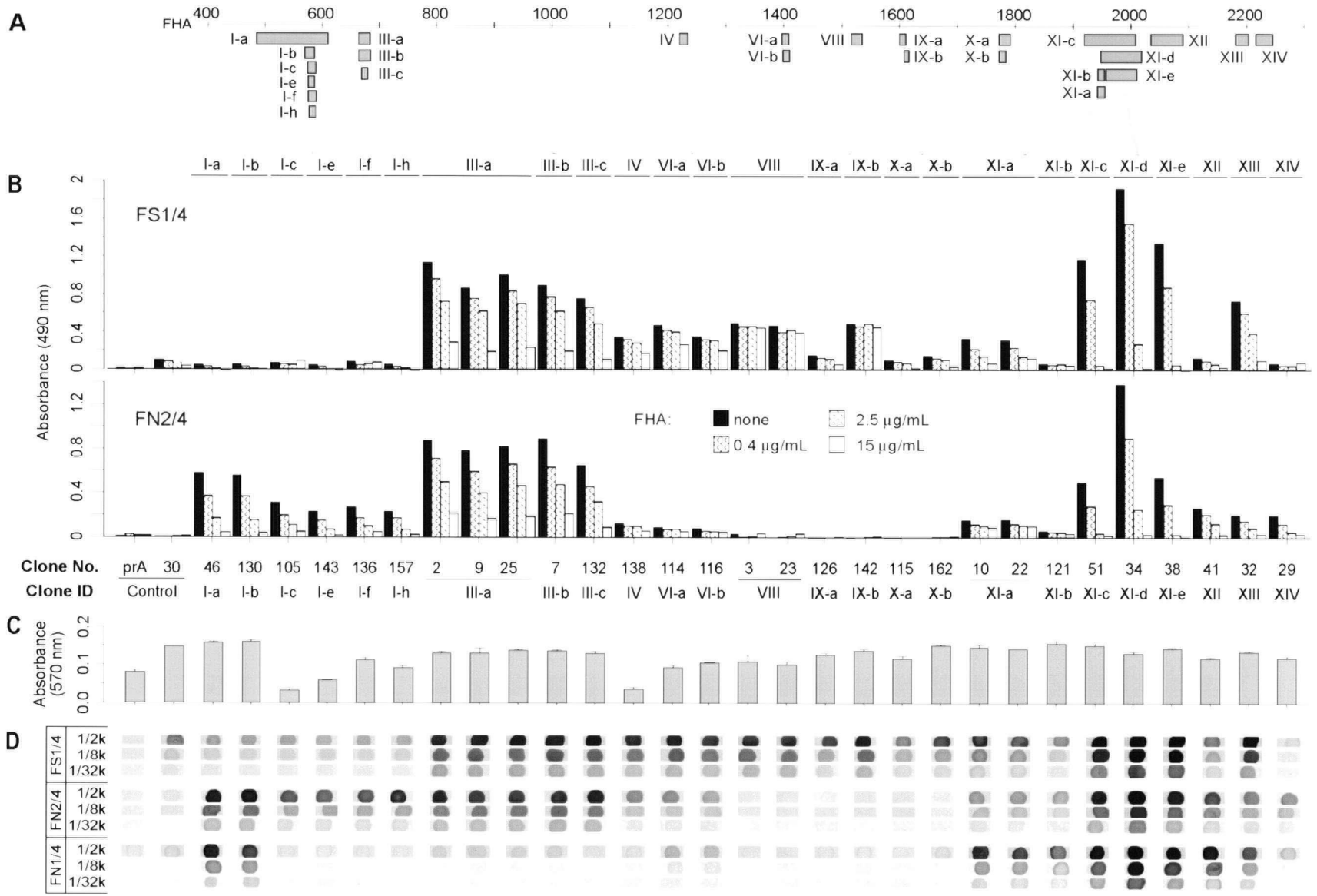

A
FHA
400
600
800
1000
1200
1400
1600
1800
2000
2200
I-a
I-b
I-c
I-e
I-f
I-h
III-a
III-b
III-c
IV
VI-a
VI-b
VIII
IX-a
IX-b
X-a
X-b
XI-c
XI-d
XI-b
XI-e
XI-a
XII
XIII
XIV
B
FS1/4
FN2/4
Absorbance (490 nm)
FHA:
none
2.5 μg/mL
0.4 μg/mL
15 μg/mL
Clone No.
prA 30 46 130 105 143 136 157 2 9 25 7 132 138 114 116 3 23 126 142 115 162 10 22 121 51 34 38 41 32 29
Clone ID
Control I-a I-b I-c I-e I-f I-h III-a III-b III-c IV VI-a VI-b VIII IX-a IX-b X-a X-b XI-a XI-b XI-c XI-d XI-e XII XIII XIV
C
Absorbance (570 nm)
D
FS1/4
FN2/4
FN1/4
1/2k
1/8k
1/32k

Fig. 6-8

Immulon plates, this binding was accordingly re-assessed (Figure 6-7). Although binding still varied in a way that appeared to reflect peptide composition, sibling-to-sibling differences were minimal (Figure 6-7C) and it appeared that binding differences could be minimized by coating plates with ≤ 2 μg virion protein per well (Figure 6-7B).

(c) Inhibition of ELISA signal by preincubation of FN2/4 and FS1/4 with purified FHA. For selected clones, a "competition" ELISA (Figure 6-8B) was performed using purified FN2/4 and FS1/4 that had been preincubated with varying quantities of purified FHA, and with purified FN2/4 and FS1/4 alone. A concurrent assay of bound virions (Figure 6-8C) showed that the range of variability had been reduced from the previously observed 8-fold range (Figure 6-3) to roughly 4.5-fold, and that most clones bound within an even narrower range. In spite of this variability, ELISA results for antibody samples not preincubated with FHA (Figure 6-8B, solid bars) were consistent with those obtained by immunoblotting (Figure 6-8D).

Preincubation of antibodies with FHA inhibited their recognition of most clones, although the effectiveness of this varied among regions. Thus, recognition of most region I and XI clones -- including the strongly recognized clones XI-c, XI-d and XI-e -- was readily titered to near-zero with increasing concentrations of FHA. Recognition of less-reactive region III, IV and VI clones was also inhibited, but not to quite the same low levels. Recognition of region VIII, IX and X clones was only marginally inhibited. These findings are discussed later.

6.3.6. Signal peptide cleavage predictions

As reviewed in the Chapter 1, there is evidence that the N-terminal mature region of a protein can interfere with signal peptidase processing and it was accordingly important to determine whether the peptides encoded by the 30 antibody-reactive clones were in fact displayed on the phage surface. Although this is experimentally impractical, it was possible to examine this issue using recently developed neural networks (Nielsen *et al.* 1997) that consider the N-terminal mature region in predicting signal peptidase cleavage sites.

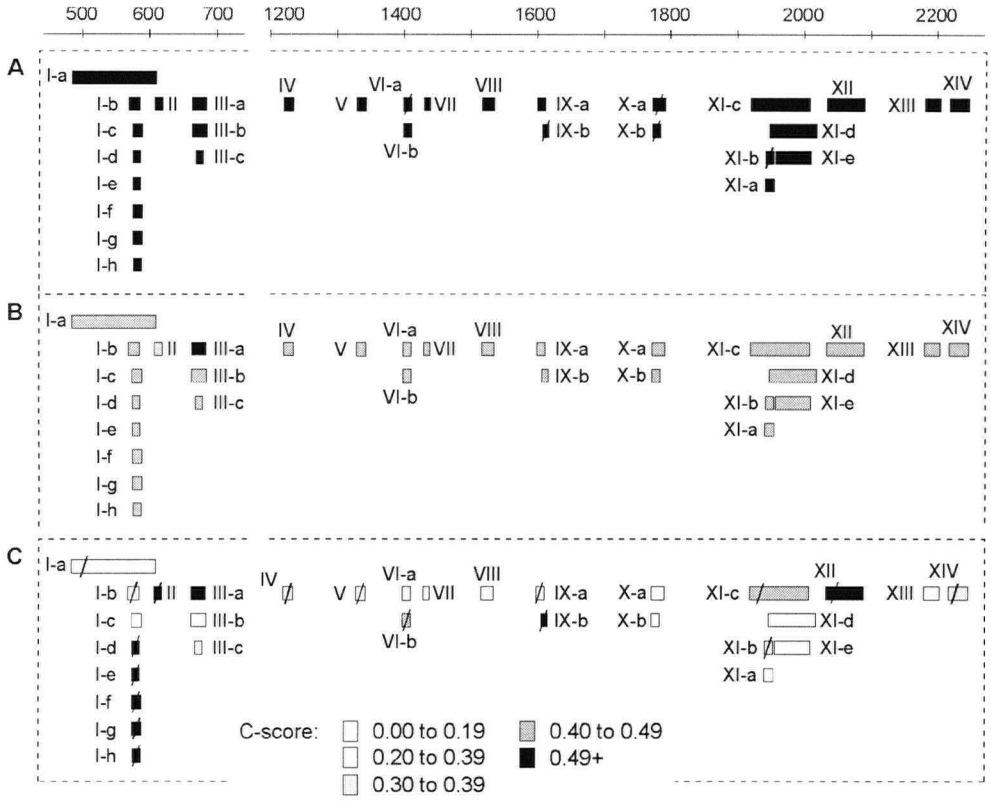


Fig. 6-9. Predicted signal peptide cleavage of antibody-reactive FHA-70 and -80 clones. Each of the peptides encoded by antibody-reactive clones identified in Figures 6-1 and 6-2 was examined for signal peptide cleavage using the neural networks of Nielsen *et al.* (1997). Figure shows the C-score (cleavage-site score), one of three scores provided by these networks. A default cutoff value of 0.49 predicts a likely cleavage site. Scores are shown (see legend) for each peptide displayed with three alternative vectors. The symbol "/" identifies predicted cleavage at a site other than that of wild-type Ff phage. **A.** Predictions for peptides displayed with fDRW70, which separates displayed peptides from signal peptides with an ADGAGA linker. **B.** Predictions for fDRW8*nn* vectors, which employ ADGP linkers. **C.** Predictions for a vector in which the displayed peptide is not separated from the signal peptide. Raw data are shown in Table 2-VI (Materials and Methods).

Accordingly, predictions from these networks were used (i) to provide some assurance that the clones characterized in this study were correctly processed, and (ii) to examine the issue raised earlier (Chapter 1, §1.2.5 and §1.3.1) concerning the putative benefit of retaining Ala-Asp or Ala-Glu residues at positions +1 and +2 of the mature protein. Specifically, signal peptide cleavage predictions were examined for each antibody-reactive peptide displayed (i) in fDRW70 which separates a displayed peptide from the signal peptide with an ADGAGA linker peptide, (ii) in an fDRW8*nn* vector which encodes an ADGP linker, and (iii) in a vector (such as fAFF1; Cwirla *et al.* 1990) which displays a peptide at the N-terminus of mature pIII, *viz.*,

without a linker. The predictions (summarized in Figure 6-9) suggest that each of the encoded peptides would be cleaved at the wild-type cleavage site when displayed in fDRW70 or fDRW8*nn*, but that few would be cleaved correctly or at all when displayed at the mature pIII N-terminus. Thus, all fDRW70 constructs yielded high C-scores (cleavage-site scores, one of three scores reported by the networks; Figure 6-9A), and although altered cleavage sites were predicted for five fDRW70 constructs, relatively high C-scores (≥0.427, cutoff value = 0.49) were also reported for the wild-type cleavage position in these constructs. As well, and although C-scores for fDRW8*nn* constructs were lower than those of fDRW70 and most were lower than the cutoff value (0.49), they were only marginally so (≥0.427; Figure 6-9B). In contrast, constructs in the third (fAFF1-like) vector yielded scores as low as 0.145 and half were less than 0.3; moreover, more than half of the constructs were predicted to have altered cleavage sites (Figure 6-9C).

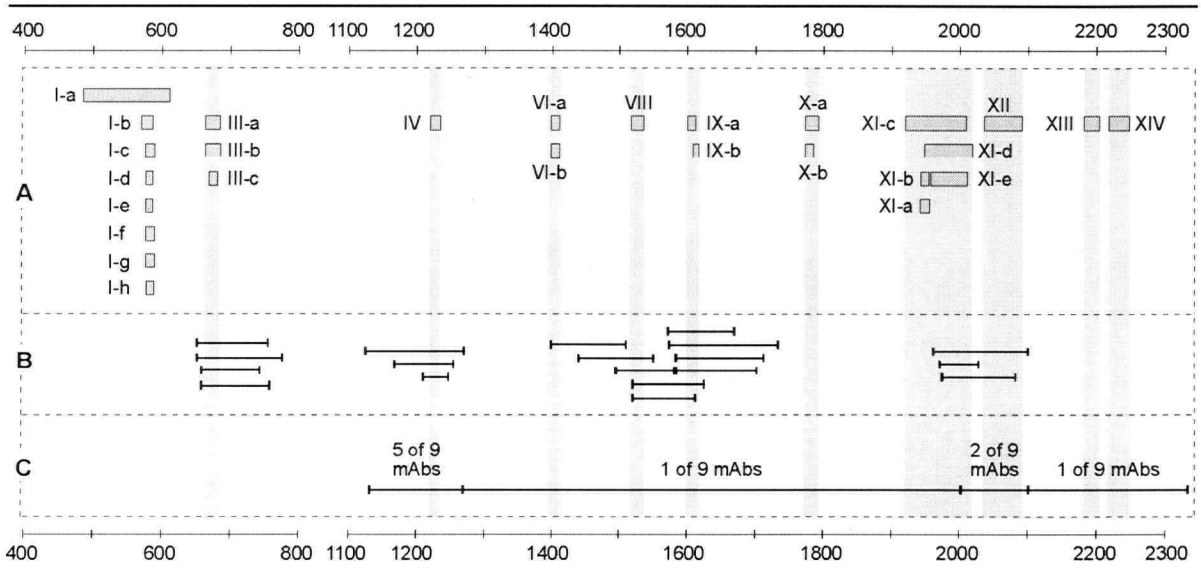


Fig. 6-10. Positional comparison of antibody-reactive clones with results of other studies. **A.** Phage-displayed antibody-reactive clones identified in the present study, mapped according to their position within the primary amino acid sequence of FhaB. **B.** Antibody-reactive clones identified in a similar study (A. Siebers and B. Finlay, unpublished data) that employed a *Pseudomonas aeruginosa* OprF expression system and the same rabbit α-FHA polyclonal sera as employed in the present study. **C.** Recognition of FHA-derived recombinant proteins by nine α-FHA mAbs allowed the approximate mapping of their epitopes to the indicated positions within a 1200-residue immunoreactive domain (Delisse-Gathoye *et al.*, 1990).

6.4. DISCUSSION

6.4.1. Comparison with other studies

As reviewed earlier, experiments to assess the sequence diversity and redundancy of the FHA-70 libraries were inconclusive and no attempt was made to assess the FHA-80 libraries. Accordingly, it was important to establish that possible limitations in these libraries had not seriously flawed this study. Results of two earlier investigations served to do this. One of these -- carried out in our laboratory -- employed a *Pseudomonas aeruginosa* OprF expression system and the same α-FHA sera employed here to identify 19 antibody-reactive OprF-FHA fusion proteins that mapped to four domains.

As indicated in Figure 6-10 (panel A *versus* B), the present study identified antibody-reactive clones within each of these domains and additionally identified clones in regions I, X, XIII and XIV. In a more limited study, Delisse-Gathoye *et al.* (1990) employed several FHA-derived recombinant proteins and nine α-FHA mAbs to map the approximate locations of at least four epitopes within a 1200-residue immunoreactive domain. As indicated in Figure 6-10 (panel A *versus* C), the present study identified 10 or more epitopes within the same domain, and additionally identified epitopes of regions I and III. Thus, limitations such as may have existed in the FHA-70 and -80 libraries nevertheless allowed for an antigenic analysis of FHA that substantially improved our understanding of previously identified antigenic domains and identified a number of additional domains.

6.4.2. Candidate immunogens for eliciting protective antibodies

Four groupings (Table 6-IV) of clones were suggested by considering patterns of reactivity in terms of (i) recognition by antibodies elicited by SDS-denatured FHA (FS1/4) *versus* "native" FHA (FN2/4 and FN1/4), and (ii) the way in which preincubation of antibodies with FHA influenced this recognition. Although the analysis is limited by the small number of sera employed and thus necessarily conjectural, it provides useful insight.

Table 6-IV. FHA regions grouped by patterns of recognition by antibody

Group	Regions of FHA	Recognized by FS1/4	Recognized by FN2/4	Effect of preincubating antibodies with purified FHA
A	VIII, IX, X	yes	no	little or no effect
B	IV, VI	yes	poorly	reduced recognition
C	I, XIV	no	yes	inhibited recognition
D	III, XI, XII, XIII	yes	yes	inhibited recognition

(a) Group A. Clones of regions VIII, IX and X were recognized by FS1/4 but not by FN2/4 or FN1/4 (Figures 6-6 and 6-8). Preincubation of FS1/4 with FHA had only a marginal effect on recognition (Figure 6-8). These findings are consistent with the idea that the phage-displayed peptides and corresponding regions of SDS-denatured FHA adopt similar, non-native conformations that fail to elicit antibodies that recognize native FHA. Accordingly, the sequences appear to have little potential value as immunogens.

(b) Group B. Clones of regions IV and VI reacted strongly with FS1/4 but weakly with FN2/4 and FN1/4 (Figures 6-6 and 6-8). Preincubating FS1/4 and FN2/4 with FHA did not diminish recognition of these clones to the same degree as occurred with the more strongly recognized clones of regions I and XI (Figure 6-8). This may indicate that phage-displayed region IV and VI peptides more closely mimic partially unfolded sequences of SDS-denatured FHA than those of native FHA, or that SDS serves to make these regions more accessible to antibody. This may limit the value of these peptides as immunogens when incorporated into recombinant protein. However, because region IV and VI sequences may be important in *B. pertussis* pathogenesis, their immunogenicity may warrant further study.

Thus, the single region IV clone encodes a peptide ($FHA_{1229-1244}$) that is almost entirely contained within a sequence ($FHA_{1224-1242}$) identified as a possible lectin-like binding domain (Sandros & Tuomanen 1993). Subject to a caution implied by the earlier-described findings of

Tuomanen *et al.* (1993) that monoclonal antibodies against an epitope within $FHA_{1141-1279}$ may mediate undesirable host physiological effects, antibodies able to neutralize this binding domain may be of protective value.

As well, the overlapping clones VI-a and VI-b encode all (VI-a) or most of (VI-b) a sequence ($FHA_{1407-1417}$) homologous to that of C3bi (Sandros & Tuomanen 1993). If the implied mimicry truly exists, then antibodies to this region may be of value.

(c) Group C. While clones of groups I and XIV were not significantly reactive with FS1/4, they reacted with FN2/4 and in some cases with FN1/4 (Figures 6-6 and 6-8), and their recognition could be inhibited by preincubation of antibodies with FHA (Figure 6-8). While it is possible that the lack of recognition by FS1/4 reflects variability in the immune response among animals, the use of SDS-denatured FHA as an immunogen may also account for this. For example, the epitopes that elicited antibodies in FN2/4 may have become masked by other sequences in the SDS-denatured FHA used to elicit FS1/4; alternatively, native folding of these sequences may be required to elicit antibodies capable of subsequently recognizing these epitopes in any form. If this is the case, the epitopes involved will have immunogenic value in recombinant proteins only if, when presented to the immune system, they adopt the conformation found in native FHA.

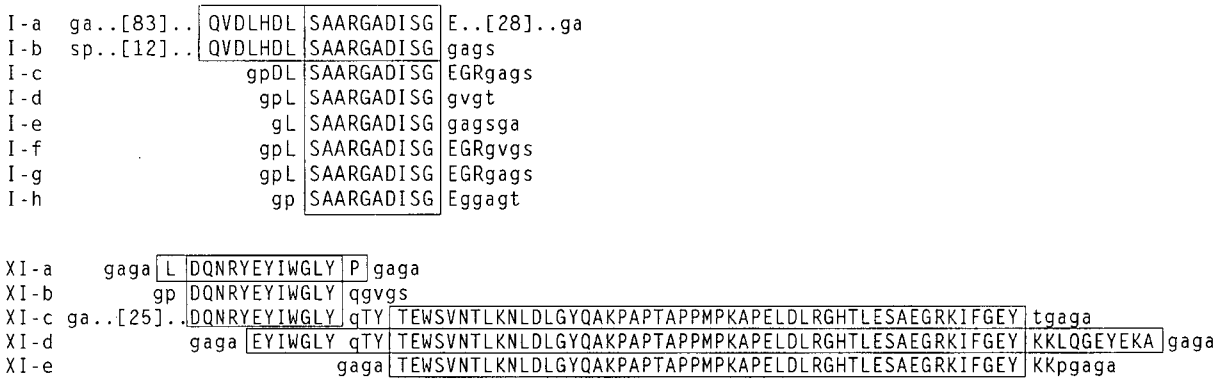


Fig. 6-11. Sequence overlaps in region I and XI clones. Sequences correspond to those of Figure 6-1, and have been abbreviated for convenience; numbers enclosed in square parentheses "{*n*}" indicate the numbers of residues omitted. Boxed regions of overlap are discussed in the text.

Region I clones accounted for more than half of the 109 sequenced clones and their siblings (Figure 6-1) and antibody-reactivity of some clones was relatively strong (Figures 6-6 and 6-8). All share a common encoded sequence ($_{586}$SAARGADISG$_{595}$; Figure 6-11), suggesting that it contains at least a portion of a common epitope. However, clones containing only these and a few other residues were recognized only by FN2/4, while much stronger reactivity with both FN2/4 and FN1/4 was seen in two clones that share a sequence extended N-terminally, ($_{579}$QVDLHDLSAARGADISG$_{595}$; Figure 6-11). Intriguingly, these sequences map within a 422-residue region (FHA$_{442\text{-}863}$) known to contain the FHA heparin-binding domain (Hannah *et al.* 1994). Moreover, they are flanked by sequences relatively rich in arginine, suggesting that the overall sequence ($_{573}$**R**V**R**G**R**GQVDLHDLSAA**R**GADISG EG**R**VNIG**RAR**SDSDVK$_{610}$, included in clone I-a) may form part of the heparin-binding domain (for analyses of heparin-binding domains, see Cardin *et al.* 1991; Margalit *et al.* 1993). The possibility that neutralizing antibodies could be raised to the heparin-binding domain merits further investigation.

Region XIV is less likely to be useful as an immunogen for only a single region XIV clone (FHA$_{2226\text{-}2255}$) was identified, and this was only moderately reactive with FN2/4 and weakly reactive with FN1/4. Curiously, and although α-FHA sera were raised against purified FHA, region XIV maps outside the estimated C-terminal boundary (residue 2165) of mature FHA (Makhov *et al.* 1994).

(d) Group D. Clones of regions III and XI-XIII were recognized by each of FS1/4 and FN2/4; region XI-XIII clones were also recognized by FN1/4 (Figures 6-6 and 6-8). Recognition by FS1/4 and FN2/4 could be inhibited by preincubation with FHA, although this was marginally less effective with region III clones than with those of regions XI-XIII (Figure 6-8). Considering that (i) SDS-denatured and possibly unfolded FHA is capable of giving rise to antibodies that recognize FHA, and (ii) phage-displayed peptides are recognized by

antibodies raised against both SDS-denatured and "native" FHA, it follows that the peptides encoded by clones of these regions adopt native FHA conformation sufficiently well that they may be able to elicit a response against native FHA.

The three region III clones were equally and strongly recognized by FS1/4 and FN2/4. Thus, the common peptide sequence ($_{678}$RDVAAAADLALQ$_{689}$; Figure 6-1) appears to contain those residues that are important for recognition. Although this sequence maps within the 422-residue sequence (FHA$_{442-863}$) that contains the FHA heparin-binding domain (Hannah *et al.* 1994), it seems unlikely that this charge-poor sequence plays a direct role in heparin binding.

Region XI includes five clones spanning 77 residues (FHA$_{1951-2027}$) and their sequences overlap (Figure 6-11) in a way that suggests there are at least two epitopes. Clones XI-a and XI-b include only the first of these, and were moderately well recognized by FS1/4 and FN1/4 but only weakly by FN2/4. The more reactive of the two, clone XI-a (FHA$_{1951-1964}$), is longer than XI-b by two residues. Although clone XI-d includes only a portion (FHA$_{1957-1963}$) of the sequence common to XI-a and XI-b, it was much more reactive, expectedly because it also includes the second and larger immunogenic region. The clones that include this region -- XI-c, XI-d and XI-e -- are the most strongly recognized of all clones analysed. Of these, the 71-residue XI-d sequence is recognized most strongly. Since XI-d lacks N-terminal residues (FHA$_{1952-1956}$) found in the more weakly-recognized XI-c and contains additional C-terminal residues (FHA$_{2018-2027}$), some of these C-terminal residues may account for the stronger recognition. Interestingly, the additional C-terminal sequence ($_{2018}$KKLQGEYEKA$_{2027}$) is homologous to a preceding, partially overlapping sequence ($_{2011}$RKIFGEYKKL$_{2020}$), raising the possibility that each of these sequences are antibody-reactive. Importantly, XI-c, XI-d and XI-e include a Factor X homolog ($_{1979}$LGYQAK$_{1984}$; Sandros & Tuomanen 1993) and, as reviewed in the chapter Introduction, it is not certain that eliciting antibodies to this sequence would be beneficial.

The single region XII clone (encodes a 66-residue sequence, $FHA_{2044-2109}$) and the single region XIII clone (encodes a 23-residue sequence, $FHA_{2190-2212}$) were reactive to varying degrees with FS1/4, FN2/4 and FN1/4 (Figures 6-6 and 6-8). The region XII clone contains a second Factor X homolog ($_{2062}$ETKEVDG$_{2068}$; Sandros & Tuomanen 1993); as with the region XI Factor X homolog, the benefit of antibodies to this sequence are uncertain. As with the region XIV clone described earlier, it is curious that the region XIII clone maps beyond the estimated C-terminus of mature FHA.

6.4.3. Issues related to phage display technology

(a) Library bias. As reviewed in Chapter 1, host and phage biology have been shown to interfere with phage display of some peptides. As well, work reviewed in Chapter 7 provided evidence that virion production varied at least 100-fold -- possibly as high as 10^7-fold -- among stable clones of PDLs constructed with restriction fragments of pAS100 (the 10 kbp *fhaB Eco*RI fragment cloned into pTZ18R), and that some clones were unstable and prone to frameshift mutations or deletions. Jacobsson and Frykberg (1995; see also Chapter 1) similarly found that 47 of 50 affinity-selected clones identified from several gene fragment libraries contained +1 or -1 frameshifts.

Although no survey was conducted to examine the related issues of insert stability, virion production and clonal bias in the FHA-70 and -80 libraries, it is noteworthy that (i) none of antibody-reactive clones possessed frameshifts or deletions, (ii) base substitutions were found in few clones, and (ii) antibody-reactive clones were found both within each antigenic domain identified by previous studies (reviewed in §6.4.1 above) and within previously unidentified antigenic regions. Superficially at least, then, bias appears to have not been a significant problem.

Vector design may have played a beneficial role in this by obviating problems in membrane insertion and signal peptide processing related to the N-terminal mature region of

the pIII fusion proteins. The fDRW70 and fDRW8*nn* vectors, like their fUSE5 parent, produce pIII fusion proteins in which the first two residues following the signal peptidase cleavage site are retained (as Ala-Asp *versus* Ala-Glu). fDRW70 additionally insulates a displayed peptide from the N-terminus with a GAGA linker, while the fDRW8*nn* vectors employ Gly-Pro. As suggested by neural network predictions (Figure 6-9), recognition and processing of signal peptide cleavage sites of constructs in fDRW70 and fDRW8*nn,* in contrast to the same constructs in a vector such as fAFF1, is expected to be normal.

On the other hand, since the FHA-70 and -80 libraries were not assessed experimentally for evidence of bias, it is also possible that biopanning merely selected the most stable, easily produced (and thus most numerous) recombinants. Thus, it cannot be said that this study provided a comprehensive antigenic analysis of FHA, but rather that "certain" (*versus* immunodominant) antigenic regions were identified.

(b) Outstanding issues: polyphage, proteases and mutants. Delays in supply of pIII to the virion assembly/export apparatus can lead to increased numbers of polyphage (reviewed in Chapter 1), each displaying 4-5 copies of pIII per two or more unit-length genomes; as might be expected, some recombinant pIII molecules have this effect (*e.g.*, Smith 1985). Since virions employed in ELISA and immunoblots were quantified on the basis of total virion DNA and coat protein (of which recombinant pIII is a minor component) and were not examined for polyphage, it is possible that increased polyphage production may have resulted in up to two-fold over-estimations of some virion numbers (and thus of recombinant pIII molecules) and to correspondingly diminished signals in immunoblots and ELISA for the affected clones.

Proteolytic degradation of pIII recombinants has been reported (reviewed in Chapter 1). Although none of the displayed peptides were examined for proteolysis, (i) the relatively high virion titers obtained for antibody-reactive clones (data not shown), (ii) the internal consistency of results for clones within a given region, and (iii) the general agreement between these results

and those of previous studies, together suggest that proteolysis was not an important issue. Nevertheless, it is possible that antibodies reacted with truncated forms of the encoded peptides and that some results reflect artefacts of proteolysis.

Although genetic stability of FHA-70 and -80 library clones was not specifically addressed, (i) that it was possible to propagate sufficient numbers of virions to obtain sequencing data of good quality and (ii) that no evidence of insertions or deletions was found suggest that the clones are stable. However, in recent experiments that required a further propagation of selected clones, a significant fraction of virions in preparations of some clones were found to produce large, clear plaques that contrast noticeably with the normally small, turbid plaques. Although these were antibody-reactive and differences in plaque morphology likely reflect alteration of genes other than recombinant gIII, it remains possible that these apparent mutants -- if they existed in the samples used for ELISA and immunoblots -- introduced artefacts.

6.4.4. Related studies and concluding remarks

The utility of the present study would be greatly extended by demonstrating that potentially protective epitopes do offer some degree of protection, or that they have identifiable functional roles and are thus potentially protective. Clearly, further experimentation is required.

Efforts to identify the heparin-binding or other adhesin domains by affinity selecting target clones using erythrocytes (Menozzi *et al.* 1991) and CHO cell monolayers (Menozzi *et al.* 1994) have so far been unsuccessful, and similar efforts with immobilized heparin have provided only a hint of success. Allowing that non-specific interactions may make these efforts inherently more difficult than those employing antibodies, biopanning with antibodies may be more successful. For example, biopanning the FHA-70 and -80 libraries with mAb X3C (Leininger *et al.* 1993) -- which blocks FHA-mediated attachment of *B. pertussis* to epithelial cells, inhibits hemagglutination (Leininger *et al.* 1993) and maps within the region of FHA

containing the heparin-binding domain (Hannah *et al.* 1994) -- might accomplish the same goals as biopanning with heparin. Similarly, biopanning with mAbs 12.5A9 and 13.6E2 (Prasad *et al.* 1993), which block interaction between ciliated cells and the lectin-like domain located within $FHA_{1141-1279}$, might yield target clones that validate or refute the idea derived from sequence comparisons that $FHA_{1224-1242}$ mediates this interaction (Prasad *et al.* 1993; Sandros & Tuomanen 1993).

By providing a useful antigenic analysis of an important *B. pertussis* virulence factor, the current study has demonstrated that, notwithstanding possible inherent limitations of gene fragment PDLs, their employment can yield a useful analysis. Thus, continued efforts to identify adhesin domains may yet yield important insights.

Chapter 7

Issues in vector stability and library construction

7.1. ABSTRACT

By the use of model systems that involved the construction of phage libraries displaying variants of the *Plasmodium falciparum* circumsporozoite protein immunodominant repeats and peptides derived from *Bordetella pertussis* filamentous hemagglutinin, various issues of phage display related to methodology and phage/host biology were explored. Principal findings indicate that vector design is important to library construction and employment, and that attempts to display some foreign peptides on the phage surface can so profoundly affect phage/host biology that in some cases incompatible peptides will be "corrected" so as to enable their display. These and other issues of phage/host biology and their impact on phage display are discussed. Although construction of gene fragment libraries is technically achievable, important questions concerning their "completeness" or diversity remain unanswered, and the utility of these libraries is thus diminished.

7.2. INTRODUCTION

As outlined in greater detail in Chapter 1, achieving my project goals involved working through two model systems of phage display intended to explore basic methodology, illustrate application potential and identify limitations imposed by phage and host biology. The first, relatively simple model involved constructing and employing a small library of variants of the *P. falciparum* CSP immunodominant repeats (Chapter 4), while the second, more complex model involved constructing and employing a series of libraries with fragments of the *B. pertussis fhaB* gene (Chapter 6).

7.2.1. The expected utility of fUSE*n* vectors

In beginning this work, a set of Type 3 vectors, the fUSE*n* derivatives of fd-tet (Figure

7-1), were chosen for their expected utility. In contrast to other then-available vectors, these had been developed over several years (Parmley & Smith 1988; Scott & Smith 1990) and were expected to be reliable. Also, their design incorporated features deemed important to several aspects of the project. Frameshift vectors fUSE1, fUSE3 and fUSE5 (Figure 7-1) appeared particularly useful, for without an appropriate foreign insert these vectors produce no virions and could thus be used to construct libraries free of non-recombinants (Parmley & Smith 1988; Scott & Smith 1990). The importance of this reflected two issues.

First, applications of PDLs characteristically involve multiple rounds of affinity selection (or biopanning) for target clones in which the enriched fraction from one round becomes input to the next. Routinely, between each round of affinity selection, an enriched fraction is amplified by infecting *E. coli* host cells and harvesting virions produced after overnight culture. A voiced concern has been that during this step, competition between non-recombinants and recombinants or even among recombinants might negate the effects of enrichment. This was expected to prove particularly problematic in gene fragment PDLs, for there is abundant evidence (reviewed in Chapter 1) that infectivity and virion production can be markedly reduced in some phage-display recombinants; thus, clonal competition was expected to be at an extreme in these compositionally diverse libraries. By eliminating non-recombinants, frameshift vectors would eliminate a potentially large source of clonal competition.

The second issue related to my goal of revealing something about the influence of a foreign peptide's composition on its ability to be displayed on phage. This goal was important, for there is evidence (reviewed in Chapter 1) that a number of host processes might interfere with successful display of some peptides, yet there have been few systematic studies of this issue. Importantly, achieving this goal did not require that difficult-to-display peptides actually be displayed. Rather, given that the libraries had been constructed with an unbiased set of DNA fragments, characterization (*e.g.*, by virion production and infectivity) of virion-

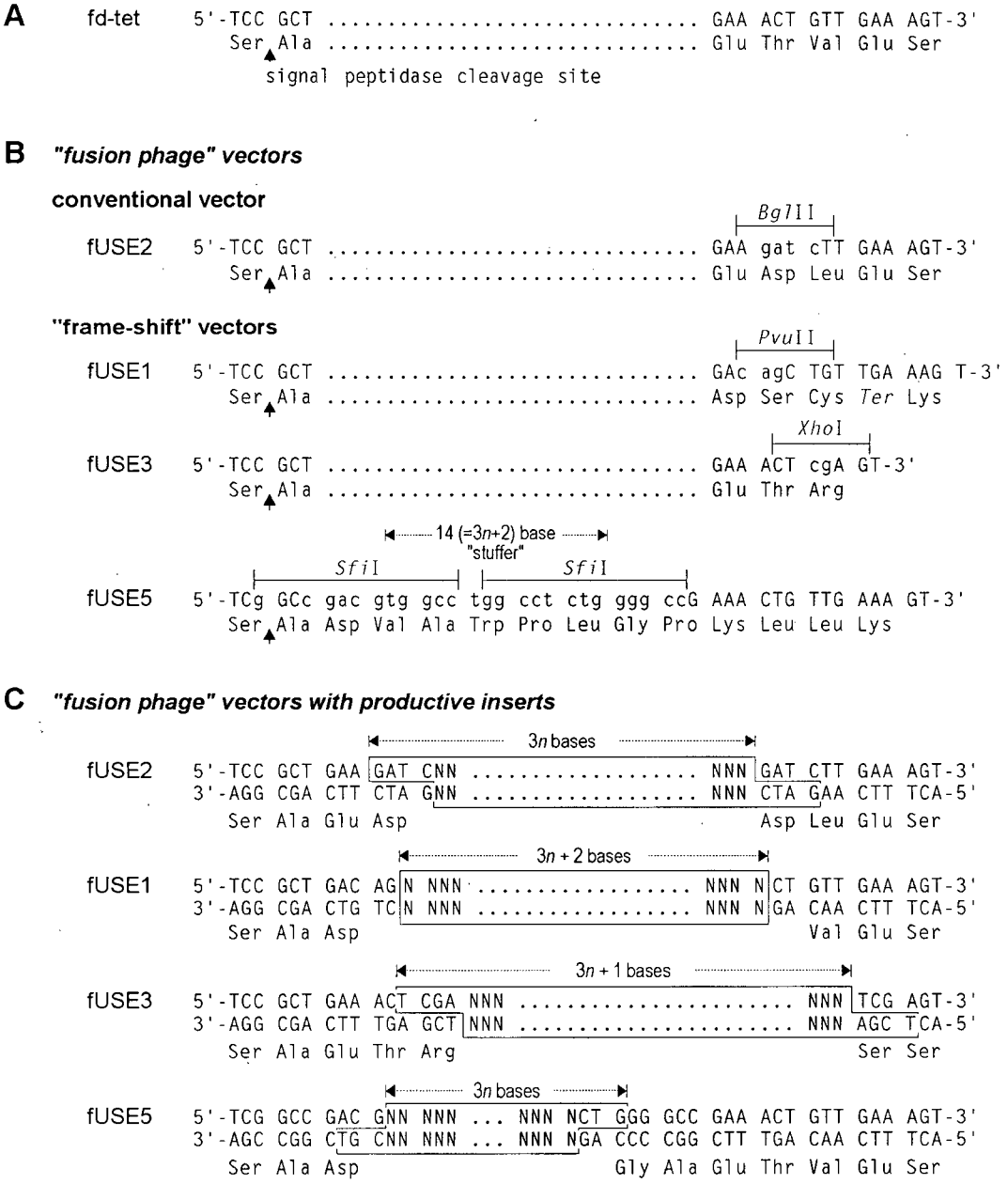


Fig. 7-1. Phage display vectors provided by G. P. Smith. Cloning with these vectors involves inserting DNA of appropriate length and 5' or 3' extensions into phage gIII (see Figure 1-1) at a position corresponding roughly to the N-terminus of pIII after signal peptide excision. **A.** Wild-type gIII and pIII sequences in fd-tet (Zacher *et al.* 1980), the parent of the fUSEn vectors. **B, C.** Conventional vector fUSE2 is contrasted with "frameshift" vectors fUSE1, fUSE3 and fUSE5. fUSE2 possesses a near-wild-type sequence (**B**) and thus produces virions in the absence of an insert into its *Bgl*II cloning site. Inserts into fUSE2 must contain $3n$ bases (where n is an integer) in order to maintain the gIII reading frame (**C**). In contrast, fUSE1, fUSE3 and fUSE5 possess frameshifts in gIII (**B**), with the result that no pIII is produced and -- since phage morphogenesis requires pIII -- no virions are produced. Blunt-end inserts of length $3n$+2 restore the gIII reading frame in fUSE1, while *Xho*I-compatible inserts of length $3n$+1 restore the reading frame in fUSE3 (**C**). Design of fUSE5 is more complex. Cloning into fUSE5 involves removing a $3n$+2 base *Sfi*I-excisable "stuffer" fragment (**B**) -- thereby producing a linear fragment with non-complementary ends -- and inserting a fragment of $3n$ bases (**C**).

producing recombinants randomly selected from FHA-peptide libraries was expected to provide an understanding of the influence of all possible FHA-derived peptides on successful display. Here, the greatest anticipated difficulty was identifying individual recombinants against a potentially large background of non-recombinants. Frameshift vectors were expected to make this step technically trivial, for these would allow for a simple screen (based on the idea that only recombinants produce virions) that allowed sensitive identification of recombinants, even those of low infectivity or productivity[10].

7.2.2. The outcome

Thus, because of fUSE*n* vector design features, technical aspects of the project were expected to be simple, involving routine methods of molecular cloning and the adaptation to phage display of common affinity selection techniques. Surprisingly, from the outset, difficulties with fUSE*n* vectors and their fDRW*n* derivatives made it clear that phage and host biology would interfere in unexpected ways with certain aspects of the project. This chapter reviews these difficulties and various aspects of the evolution of the project, with a view in mind of illustrating certain limitations of phage display technology. Included are issues related to constructing the model libraries that -- as described in Chapters 4 and 6 -- provided insight into the structure of the *P. falciparum* CSP immunodominant repeats, the epitope specificities of α-CSP mAbs, and the antigenic makeup of *B. pertussis* filamentous hemagglutinin. While the focus of this chapter is often necessarily technical, the various issues have where possible been related to phage and host cell biology and their relevance to phage display discussed.

7.3. RESULTS

7.3.1. Early experimentation with frameshift vectors

---

[10]Extremely low numbers of infectious particles can be detected by transduction of a selectable marker (tetracycline resistance in the case of fUSE*n* recombinants or other fd-tet derivatives) into an F-piliated *E. coli* host cell and subsequent colony formation. Since, in contrast to plaque formation, transduction requires only a single infection event, even recombinants possessing only marginal infectivity can be detected.

The vector of choice for the model library of CSP variant epitopes was fUSE5 (Figure 7-1), a frameshift vector with an *Sfi*I-excisable stuffer fragment that upon removal creates incompatible cohesive ends that allow directional cloning of synthetic oligonucleotides. For this and other fUSE*n* vectors, Smith had conveniently provided fUSE5 as infective virions (s.s. fUSE5 DNA packaged by means of a pIII-encoding helper plasmid). Accordingly, fUSE5 RF DNA required for library construction was propagated by infecting the F-piliated *E. coli* strain K91-Kan with these virions, and harvesting RF DNA from large-scale (total = 1.5 L) overnight cultures.

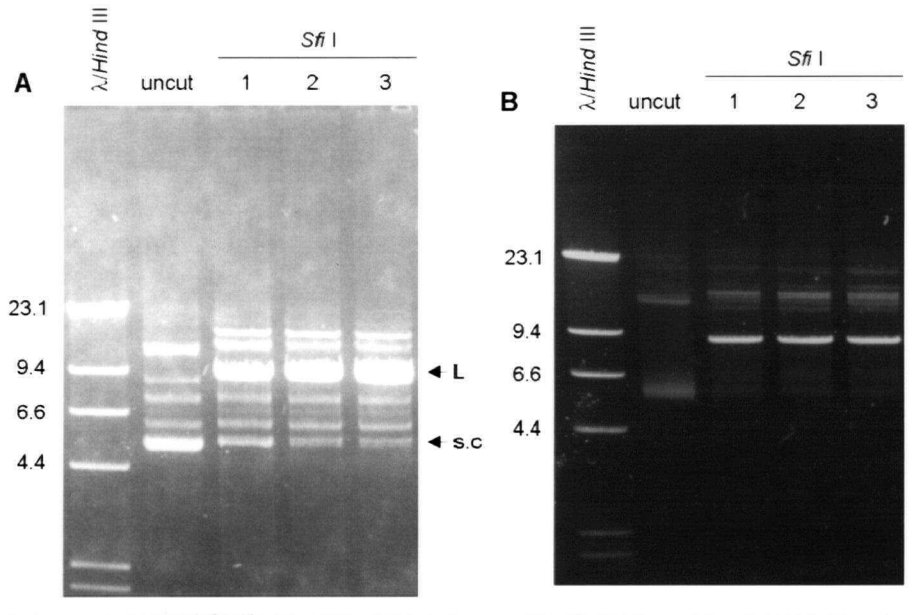


Fig. 7-2. Early efforts to prepare linear fragments of fUSE5 RF DNA. Agarose gels **A** and **B** are of RF DNA propagated in *E. coli* K91-Kan (F-piliated) that had been infected with single-stranded fUSE5 DNA packaged into virions (by Smith, using a pIII-encoding helper plasmid). **A**. Samples from digests in which ~90 µg fUSE5 RF was incubated 1 h with 100 U GIBCO/BRL *Sfi*I (lane 1), an additional 1 h with an additional 100 U *Sfi*I (lane 2), and an additional 1 h with an additional 100 U *Sfi*I (lane 3). **B.** Smaller-scale digests of ~1 µg fUSE5 RF DNA for 2 h with 5 U (lane 1), 10 U (lane 2) and 20 U (lane 3) *Sfi*I.

Indicated sizes of λ/*Hind*III and ϕX174 molecular weight markers are in kilobase pairs. Arrows indicate putative supercoiled (s.c.) and linear (L) DNA species.

(a) Unexpected DNA species in fUSEn RF DNA preparations. In marked contrast to expectations, a routine diagnostic agarose gel of the harvested fUSE5 RF DNA digested with *Sfi*I showed a bewildering variety of DNA species, including a large fraction that resisted

digestion. Persistence was of no avail. For example, in a first digest of one fUSE5 DNA preparation (Figure 7-2, panel A), increasing the quantity of *Sfi*I from 1 U μg^{-1} DNA to 3 U μg^{-1} DNA while concurrently increasing the digestion time from 1 to 3 hours demonstrated that neither increased *Sfi*I nor increased incubation time yielded discernible changes in the recovered bands. A followup 2 h digest of the same fUSE5 DNA preparation with as much as 20 U *Sfi*I μg^{-1} DNA (panel B) yielded similar results. Trials with a second preparation of RF DNA prepared from a smaller-scale culture (~180 mL) and digested with *Sfi*I purchased from an alternative supplier similarly produced species that resisted digestion (not shown). RF DNA of another frameshift vector, fUSE1 (also harvested from a medium-scale [180 mL] culture of F-piliated *E. coli* infected with virion-encapsidated vector DNA), produced multiple DNA species similar to those of fUSE5 (Figure 7-3). In contrast, RF DNA of conventional vector fUSE2 and frameshift vector fUSE3, prepared in a similar manner, each produced a single dominant DNA species that could be digested with restriction enzymes appropriate to their cloning sites (Figure 7-3). Later routine preparations have confirmed that conventional vector fUSE2, unlike fUSE1 and fUSE5 frameshift vectors, almost invariably produces readily-digestible RF DNA.

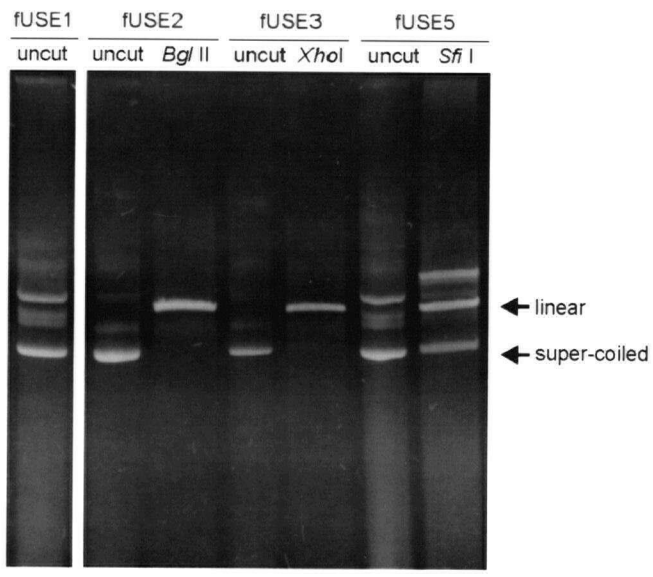


Fig. 7-3. Early preparations of fUSEn RF DNA propagated in *E. coli* K91-Kan infected with Smith-supplied virions. For frameshift vectors fUSE1, fUSE3 and fUSE5, these virions had been packaged by means of a pIII-encoding helper plasmid. Each sample derives from ~1 µg RF DNA digested with an apparent excess of *Bgl*II, *Xho*I or *Sfi*I. fUSE1 DNA digested with *Pvu*II is not shown; this digest showed considerable evidence of "star" activity. Arrows indicate putative linear and super-coiled DNA species.

(b) Frameshift vectors revert to pIII production. Insight into the problem was provided in supplementary material in Smith's "Cloning in fUSE vectors" handbook (Smith 1992). In response to reports of unexpected production of infectious virions by frameshift vector fUSE5, Smith reasoned that mutants in which the gIII reading frame had been restored ("pseudo-revertants" to wild-type pIII, and thus producing infectious virions) could come to dominate a population propagated in an F-piliated host. Indeed, since pIII mediates resistance to superinfection (Boeke *et al.* 1982), cells infected with a frameshift vector would be expected to remain sensitive to superinfection by pseudorevertants. It seemed possible that mutations giving rise to pseudorevertants might also explain the unexpected DNA species or the inability of cloning site restriction enzymes to digest all vector RF DNA.

To minimize pseudorevertants, Smith recommended that frameshift vectors be propagated in a non-piliated (F^-) host strain and that pseudorevertant production be examined for each culture from which RF DNA was extracted. Considering this, and to further examine the degree to which fUSE5 instability could be a problem, K91-Kan-passaged fUSE5 RF DNA was transformed into *E. coli* MC1061 (F^-) and pseudorevertant production[11] by *E.coli* K91-Kan/fUSE5 was examined. This was accomplished by titering infectious virions recovered from overnight culture supernatant on *E. coli* K91-Kan. Most (14 of 20) fUSE5 isolates produced a detectable number of infectious virions, or "pseudorevertants" (Figure 7-4A), and some (2 of 20) produced numbers indistinguishable (given a rather low upper-limit of detection) from virion production by conventional vector fUSE2 employed as a positive control. That the level of production varied over the range of detection limits (thus at least 100-fold) is consistent with the idea that relatively rare mutations arose earlier in some cultures than in others. Similar experiments with fUSE1, fUSE3 and fUSE5 propagated in *E. coli* K802

---

[11]Pseudorevertants of frameshift vectors can be measured as total plaque-forming units (pfu) on an F-piliated *E. coli* host strain; pseudorevertants of amber vectors (described later), as pfu on non-amber-suppressing F-piliated hosts.

(F^-) after transforming this strain with s.s. DNA extracted from Smith-supplied virions, showed that each of these frameshift vectors produced pseudorevertants, fUSE5 producing the greatest number and thus, arguably, being the most unstable. Thus, a minority of fUSE1 and fUSE3 isolates produced relatively few infectious virions (Figure 7-4B and C), while all or most (panels B and C) fUSE5 isolates produced greater numbers of infectious virions.

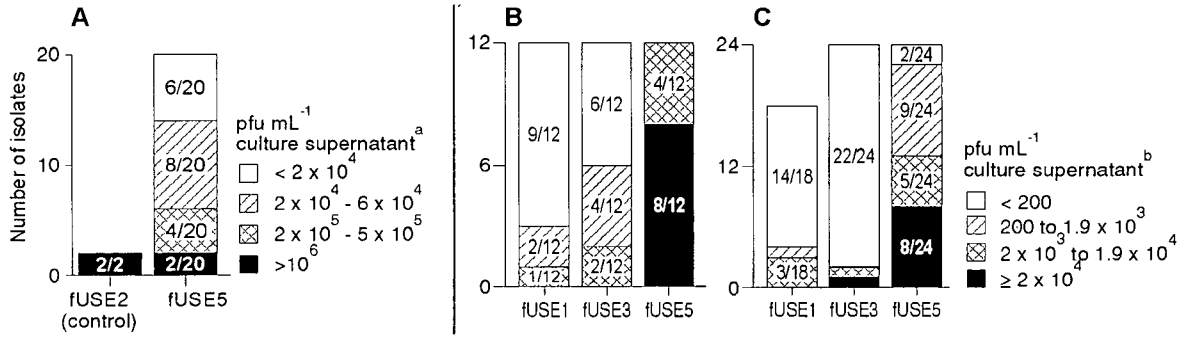


Fig. 7-4. Pseudorevertant production by fUSEn frameshift vectors. **A**. Twenty isolates of *E. coli* MC1061 (F-minus) transformed with fUSE5 and fUSE2 RF DNA that had been propagated in *E. coli* K91-Kan were cultured overnight before assaying virion production as plaque-forming units (pfu). Chart shows the numbers of isolates producing different levels of infectious virions. fUSE2 was expected to produce large numbers of infectious virions, while frameshift vectors fUSE1, fUSE3 and fUSE5 were expected to produce none. **B** and **C**. Twelve (B) and 18-24 (C) isolates of *E. coli* K802 transformed with single-stranded fUSEn DNA extracted from Smith-supplied virions were similarly cultured overnight before assaying virion production. Charts show the numbers of isolates producing different levels of virions.

[a]In assay **A**, lower and upper limits of detection are 2×10^4 and 10^6 pfu mL^{-1} culture supernatant. [b]In **B** and **C**, lower and upper limits of detection are, respectively, 2×10^2 and 2×10^4 pfu mL^{-1}.

(c) Unexpected DNA species include both s.s. and d.s. forms. The identities of the multiple DNA species were determined in three experiments. In the first, agarose gels of DNA extracted from *E. coli* K802/fUSE5 transformants cultured in the presence or absence of chloramphenicol revealed that two of four predominant DNA bands were single-stranded (Figure 7-5), for they were present in DNA harvested from the culture propagated without chloramphenicol, but lacking in the culture propagated with chloramphenicol, which prevents accumulation of s.s. (+) strands (Smith 1988). That the s.s. and d.s. bands appeared paired suggested that they were derived from two vectors, each represented by a d.s. species and its corresponding s.s. species.

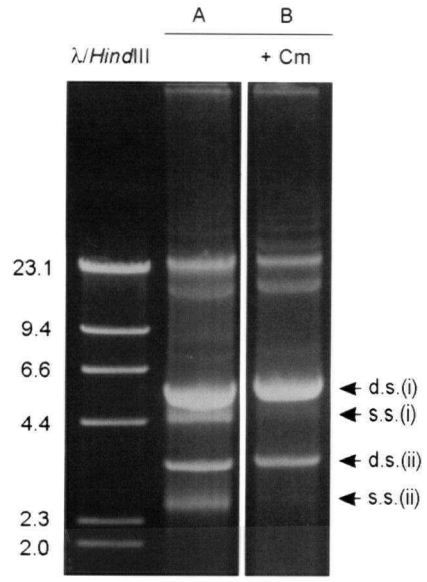


Fig. 7-5. Use of chloramphenicol-containing medium to reduce s.s. DNA species. A pool of four isolates of fUSE5 in *E.coli* K802, screened for low virion production (Figure 7-4B), were cultured ~4 hours to OD_{600} ~0.7. After adding chloramphenicol to culture B, cultures were incubated an additional 1 h (culture A, final OD_{600} =1.2) or 1.75 h (culture B, final OD_{600} = 0.8) before extracting vector DNA.

A, DNA from culture A; **B**, from culture B; neither were treated with a restriction endonuclease. Arrows indicate putative double-stranded (d.s.) and single-stranded (s.s.) DNA species (i) and (ii).

Indicated sizes of λ/*Hind*III molecular weight markers are in kilobase pairs.

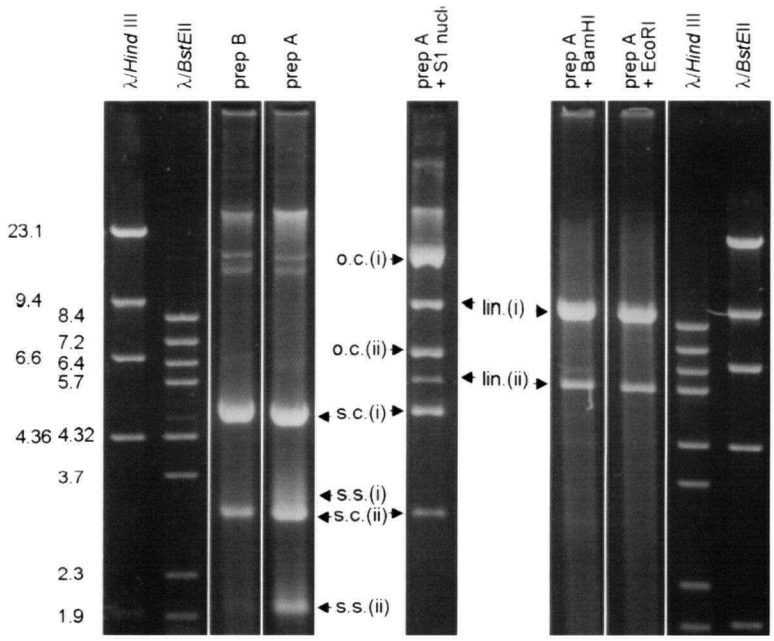


Fig. 7-6. Effects of S1 nuclease and restriction endonucleases on fUSE5 DNA preparations. Agarose gel shows undigested DNA (lanes labelled "prep B" and "prep A" of each of the two DNA preparations described in Figure 7-5, as well as preparation A treated with S1 nuclease, *Bam*HI and *Eco*RI. S1 nuclease degrades single-stranded (s.s.) DNA, including locally single-stranded regions in super-coiled (s.c.) DNA. S1 nuclease thus eliminates putative s.s. species (i) and (ii) and the smear of high- to low-molecular-weight (HMW, LMW) putative s.s. species, would appear to create new open circular (o.c.) forms (i) and (ii) from the corresponding s.c. species, and introduces linear (lin.) species (i) and (ii), presumably from certain o.c. forms. Each of *Bam*HI and *Eco*RI produces two putative linear species, digests the HMW band migrating at >23.1 kbp (characteristic of host cell DNA), and appears to have little effect on the smear of HMW to LMW species. Indicated sizes of molecular weight markers λ/*Hind*III and λ/*Bst*EII are in kilobase pairs.

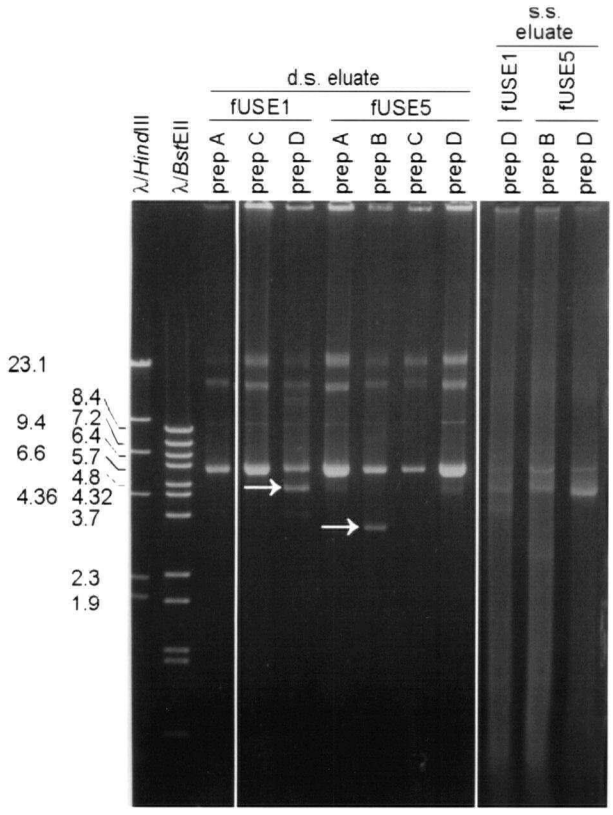


Fig. 7-7. Elution of s.s. DNA with 1.3M NaCl in Qiagen column wash. DNA was harvested from overnight cultures of fUSE1 and fUSE5 isolates (previously screened for low virion production) using a modified Qiagen protocol in which an additional wash with Qiagen reagent QC containing 1.35 M (*versus* 1.0 M) NaCl (QC-1.35) was performed to elute s.s. DNA before eluting d.s. species.

"d.s. eluate" identifies the final eluate, using the standard Qiagen elution reagent for preparations of fUSE*n* DNA. The predominant bands migrated at the expected location for super-coiled RF DNA. Arrows identify other putative super-coiled species.

"s.s. eluate" identifies the eluate from the QC-1.35 wash, concentrated with Centricon 30 microconcentrators. Species included not only putative s.s. and d.s. (super-coiled) DNA, but a smear of high- to low-molecular weight putative s.s. DNA.

Sizes of λ/*Hind*III and λ/*Bst*EII molecular weight markers are in kilobase pairs.

In the second experiment, these DNA preparations were treated with S1 nuclease and restriction enzymes that cut once within the fUSEn genome (Figure 7-6). S1 nuclease, which degrades s.s. DNA and introduces nicks in super-coiled DNA, served to clearly identify the s.s. bands, while restriction enzymes served to identify linear species. These being identified, the migration patterns of other species allowed their identification as super-coiled and open-circular forms. Importantly, the relative sizes and pairing of the s.s., super-coiled d.s., linear and putative open-circular d.s. species again suggested that the preparations contained two vectors, one conceivably being a deletion mutant of the other. This experiment also identified a characteristic "smear" of high- to low-molecular weight DNA -- common in agarose gels of fUSE*n* DNA -- as being of s.s. origin, since it was digested with S1 nuclease.

In the third experiment, the s.s. nature of the "smear" was confirmed by demonstrating that much of it could be eluted (Figure 7-7) with an elevated concentration of NaCl (chosen so

as to elute s.s. but not d.s. DNA; Qiagen 1991) in the washing step of the Qiagen procedure routinely used to prepare vector DNA. This experiment also showed that, as seen earlier, one of three fUSE1 and one of four fUSE5 isolates contained two putative vector species (Figure 7-7, fUSE1, prep D; fUSE5, prep B). Subsequent restriction analysis (not shown) suggested that the faster-migrating forms were deletion mutants of the former.

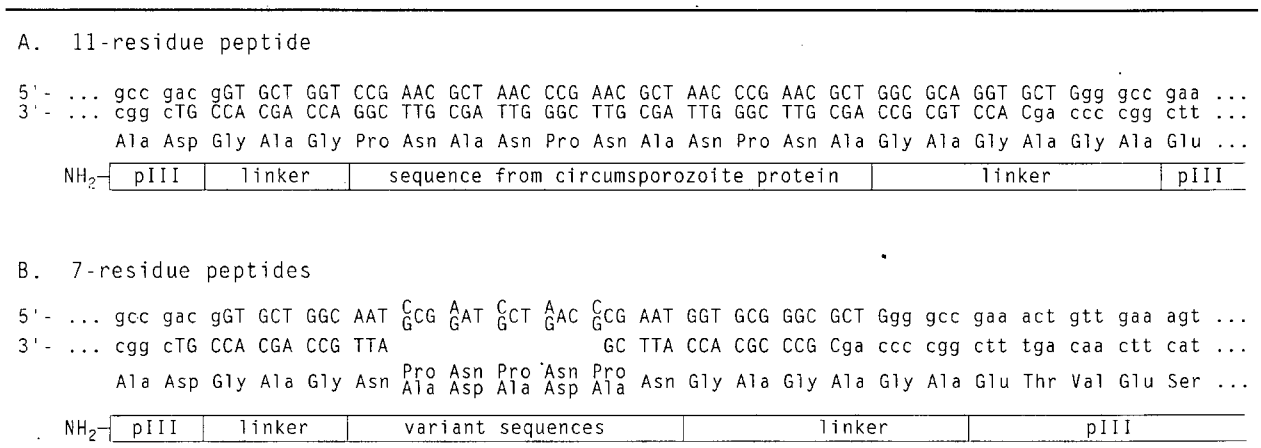


Fig. 7-8. Construction of phage displaying CSP-related peptides. **A**. Oligonucleotides employed to construct (in fUSE5) a clone displaying PNANPNANPNA. **B.** Semi-degenerate oligonucleotides used to construct (in fDRW5) a library of 32 CSP-related peptides. After ligation of the annealed oligonucleotides into fDRW5 and electroporation of ligation products into *E. coli* MC1061, the gap in the complementary (bottom) strand was filled in by *E. coli*. Each peptide is separated from pIII residues by Gly-Ala-Gly and $(Gly\text{-}Ala)_3$ linkers. Lower case letters are vector sequences; upper case, inserted oligonucleotides. Sequences reflect codon usage in *E. coli* and Ff gene III.

7.3.2. A library of variants of the *P. falciparum* CSP immunodominant repeats, and development of amber vector fDRW5

(a) A first phage-display clone constructed with fUSE5. Although disappointing in terms of its stability and the quality of RF DNA produced, fUSE5 had been successfully employed to construct a large and diverse library (Scott & Smith 1990). Indeed, in a first cloning trial in which oligonucleotides encoding an 11-residue CSP-derived peptide (Figure 7-8A) were inserted into fUSE5, careful screening of transformants by virion production, restriction analysis and sequencing readily led to the desired clone. Moreover, retrospective analysis suggested that had screening been omitted and a single virion been arbitrarily chosen

from a large pool produced by all transformants, it would likely have displayed the CSP-derived peptide.

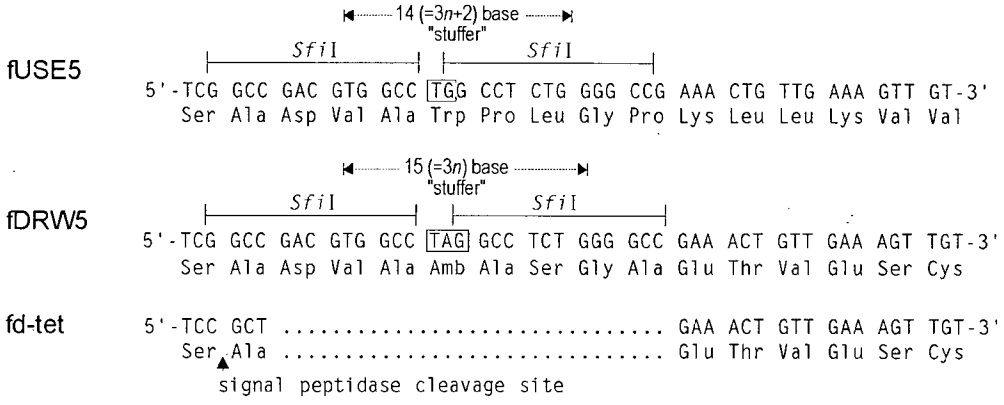


Fig. 7-9. Conversion of frameshift vector fUSE5 to amber vector fDRW5. Replacing the 14-base "stuffer" fragment of fUSE5 with a 15-base stuffer containing a single additional base created an amber stop codon in fDRW5 and restored the gIII reading frame, as illustrated by comparison with fd-tet.

(b) Conversion of fUSE5 to fDRW5. Notwithstanding this success, routine use of fUSE5 required that multiple cultures be screened for pseudorevertant production before RF DNA could reliably be used. Moreover, it was of some concern that the relatively poor quality of DNA preparations might interfere with construction of more complex libraries. A more effective way to accomplish the design objective of the fUSE5 frameshift -- eliminating non-recombinants from a library -- appeared to require only the insertion of a single base into the *Sfi*I-excisable stuffer fragment, thereby restoring the gIII reading frame and converting frameshift vector fUSE5 to amber vector fDRW5 (Figure 7-9). Provided fDRW5 was propagated in an amber-suppressing host and libraries were constructed in a non-suppressing host, fDRW5 would accomplish the fUSE5 design goals. The perceived advantage of this approach was that, in contrast to a putative selective advantage enjoyed by fUSE5 pseudo-revertants over their frameshifted parents, mutations arising during propagation of fDRW5 in an amber-suppressing host would be neutral (base substitutions) or disadvantageous (frameshifts). Possibly also, considering the effects of gIII mutants on cell health (reviewed

in Chapter 1), improved host cell health might improve the quality of RF DNA preparations.

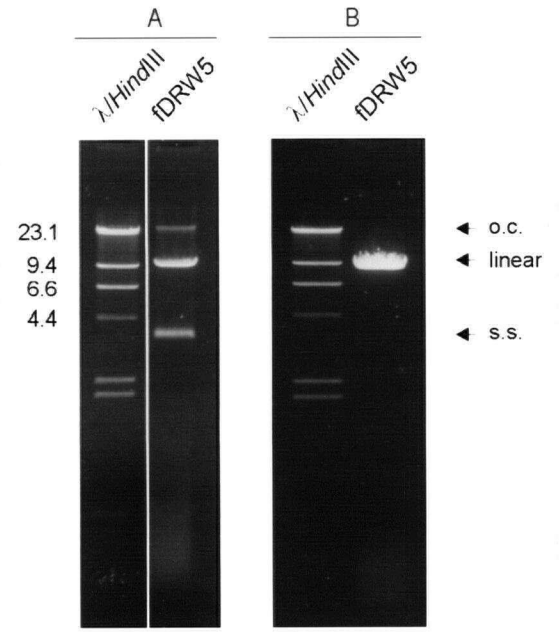


Fig. 7-10. fDRW5 RF DNA preparations for constructing CSP-peptide libraries. Sample in **A** (first library) is from a routine Qiagen preparation of fDRW5 DNA propagated in *E. coli* K802 (SupE, F-minus), digested with *Sfi*I (~10 U μg^{-1} DNA, ≥1 h); preparations of this kind routinely contain single-stranded (s.s.) and other (o.c., possibly open circular) species. Sample in **B** (second library) is from a Qiagen preparation of fDRW5 DNA propagated in *E. coli* K37 (SupD, HfrC), further purified by CsCl density gradient centrifugation to isolate super-coiled DNA, and digested with *Sfi*I (~5 U μg^{-1} DNA, 2 h).

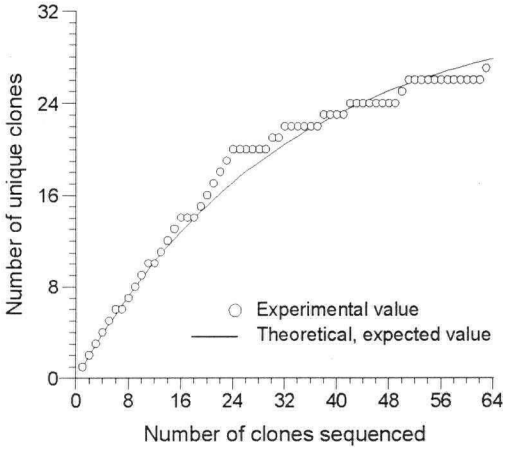


Fig. 7-11. Assessment of clonal bias in first CSP-library. Figure shows the number of unique clones identified as a function of the number of clones sequenced. All sequenced clones contained inserts. Twenty-seven unique clones were identified from 63 clones sequenced. The theoretical or expected value (—, shown for convenience as a continuous variable) is based on drawing at random from an infinitely large population of 32 unique clones.

(c) A first CSP-library yielded only recombinants. When used to construct a library of variant peptides derived from the CSP protein (CSP-library; Figure 7-8B), fDRW5 functioned as expected. Indeed, in spite of employment of an RF DNA preparation that had not been digested to completion and which contained substantial quantities of transformable vector s.s. DNA (Figure 7-10A), only recombinant virions were identified. Thus, of the 63 virion-producing clones sequenced, all displayed a target CSP-peptide (62 clones) or a peptide with an unexpected but CSP-related sequence (1 clone). Bias against specific residues or sequences

appeared minimal, for sequencing of 63 clones had yielded roughly the expected number of unique CSP-peptide variants, *i.e.,* 26 of the 32 possible (Figure 7-11). However, similar to the experiences of others (reviewed in Chapter 1), some limited bias was found against proline-rich peptides among the collection of recovered CSP-peptide sequences.

(d) A second, proline-biased library yielded only pseudorevertants. To identify missing proline-rich clones, a second CSP-library was constructed using the same oligonucleotides as before (Figure 7-8B), but with the ratio of bases encoding Pro *versus* Ala chosen to favor Pro. Unlike fDRW5 RF DNA employed for the first library (Figure 7-10A), RF DNA for this library had been purified by a CsCl gradient and, upon digestion with *Sfi*I, produced only the single expected linear species (Figure 7-10B). Surprisingly, no CSP-peptide clones were recovered from virion-producing clones of this second library, but rather only contaminants (presumably arising from trace amounts of undigestible DNA poorly detectable on an agarose gel; Figure 7-10B) and fDRW5 pseudorevertants. In these, base substitution mutations had converted the amber codon (TAG) to codons encoding Trp (TGG) or Tyr (TAT). Interestingly, the TAG→TAT mutation destroyed the second of the two *Sfi*I restriction sites. Although there may be a host of technical reasons (*e.g.,* failed ligations) explaining why no recombinants were recovered, such that repeating the work with the same materials may have yielded a dominant fraction of recombinants, it had become evident that fDRW5 could produce pseudorevertants.

(e) An alternative amber-suppressing host improved fDRW5 stability. Given that RF DNA for the first and second CSP-libraries had been prepared, respectively, in a SupE (*E. coli* K802) and SupD (K37) host, it was possible that efficiency of amber suppression played a role in fDRW5 stability. This proved to be the case, as demonstrated by titers of pseudorevertants recovered after overnight propagation of fDRW5 in *E. coli* strains K37, K802 and LE392, which carry different amber suppressors (Table 7-I). Thus, and in a manner generally consistent with the reported relative efficiency of amber suppressors (Miller & Albertini 1983;

Engelberg-Kulka & Schoulaker-Schwarz 1996), pseudorevertant numbers were lower in *E. coli* LE392 (SupE, SupF) than in strains K37 (SupD) or K802 (SupE). As well, fDRW5 RF DNA harvested from overnight culture of LE392, in contrast to that harvested from K37 and K802, showed little evidence of s.s. DNA and no evidence of putative deletion mutants (Table 7-I).

Table 7-I. Evaluation of *E. coli* host strains for fDRW5 propagation.

E. coli host strain	Salient phenotype	Pseudorevertant titer (pfu mL^{-1})[a]	Production of single-stranded DNA[b]	Evidence of deletion mutant[d]	Relative quantity of RF DNA[e]
LE392	F-minus, SupE, SupF	6 of 6 cultures: $< 2 \times 10^3$	6 of 6 cultures: no	6 of 6 cultures: no	0.25 to 0.5
K37	F-piliated (HfrC), SupD	3 of 6: 1×10^4 3 of 6: $>1 \times 10^4$	4 of 6: no 1 of 6: yes 1 of 6: note *c*	5 of 6: no 1 of 6: note *c*	1
K802	F -minus, SupE	6 of 6: $>1 \times 10^4$	6 of 6: yes	5 of 6: no 1 of 6: yes	1

Assessment is based on overnight cultures of fDRW5 in each of the indicated strains, 6 cultures per strain.

[a]Titer on non-amber-suppressing host strain, based on duplicate 4-5 µL samples of undiluted culture supernatant applied to lawns of *E. coli* K91-Kan/pACYC184. [b,c,d,e]As judged by the appearance of bands on an agarose gel of fDRW5 DNA. [c,d]Based on agarose gel bands of putative "deletion mutants" (see discussion of Figures 7-5 to 7-7). [c]Super-coiled RF DNA can be difficult to distinguish from bands of s.s. DNA; it was difficult to discern whether one such band (host K37, note *c*) was s.s. or super-coiled RF DNA.

(f) A third, proline-biased CSP-library yielded a small fraction of pseudorevertants.

As expected, four of the remaining six unidentified CSP-peptide clones were successfully isolated from a third library constructed using fDRW5 RF DNA propagated in *E. coli* LE392. In constructing this library, the issue of whether pseudorevertants were of practical consequence was also examined. This was measured in two ways (Table 7-II). First, comparison of virion production by transformants derived from (i) ligations with vector, and (ii) a control ligation involving vector only, provided a crude estimate of pseudorevertant production expected in routine library construction (Table 7-II, ratio A). A second comparison of transformants derived from similar ligations, but with ligation mixtures doped with a two-fold excess of

undigested vector, provided an alternative measure of pseudorevertant production (Table 7-II, ratio B) expected in libraries where certain technical aspects of cloning had been performed inefficiently. These estimates suggested that, commonly, no more than 1 in 300 virion-producing transformants would be pseudorevertants. Clearly, this was not expected to interfere with the need to readily identify recombinants, and the goals of characterizing a diverse collection of virion-producing recombinants thus seemed achievable.

Table 7-II. Construction and assessment of third CSP-library.

Ligation[a]		Electroporation[b]			Trans-formants	Virions produced		
fDRW5 (fmol)	Insert (fmol)	fDRW5 in ligation products (fmol)	Uncut[c] fDRW5 (fmol)	Time constant[d]	Log (cfu)[e]	Log (pfu)[f]	Recombinants: pseudorevertants[g] Ratio A	Ratio B
10	0	1	0	4.4	4.0	10.5	1:1	
10	30	1	0	4.3	4.3	13.0	300:1	
10	90	1	0	4.3	4.1	13.4	700:1	
		0.5	0.1	4.3	5.7	11.4		500:1
		0	0.1	4.3	5.9	8.8		1:1

[a]Each ligation was carried out with the indicated 0:1, 3:1 and 9:1 ratios of insert to fDRW5 RF DNA. [b,c,d]One-tenth, 1/20 or none of the ligation products, together with no or 0.1 fmol fDRW5, not digested with *Sfi*I, were electroporated into *E. coli* MC1061. [d]Time constants (Miller, 1994) provide an index of uniformity of electroporations. [e]Estimates of total numbers of transformants recovered (cfu, colony-forming units) are based on a single 10 µL sample spread on a standard petri plate.

[f]Estimates of virions (PFU, plaque-forming units) produced by transformants are based on duplicate 4-5 µL samples of a single dilution series applied to lawns of *E. coli* K91-Kan.

[g]Ratios of recombinants to pseudorevertants were estimated by computing ratios of pfu, and are discussed in the text.

7.3.3. A *B. pertussis fhaB Sau*3AI fragment library constructed with conventional vectors

Concurrently with work with the CSP-library and development of amber vector fDRW5, work had begun to explore the feasibility of constructing *B. pertussis fhaB* gene fragment libraries using conventional vectors. Motivation for considering conventional *versus* frameshift vectors reflected several ideas. First, RF DNA preparations of the conventional vector fUSE2

were generally qualitatively superior to those of frameshift vectors. Conceivably, this reflected the idea that cells harboring fUSE2 were healthier. Secondly, it was not yet clear that eliminating non-recombinants from a library was imperative to achieving the goal of assessing the influence of a peptide sequence on its ability to be displayed. Accordingly, a *Sau*3AI partial digest of pAS100 (a plasmid harboring ~90% of *B. pertussis fhaB* within a 10 kbp fragment) and three conventional vectors (differing in reading frame at the cloning site; Figure 7-12) were used to construct a set of libraries (FHA-S libraries) displaying FHA-derived peptides of 400-800 residues (Table 7-III).

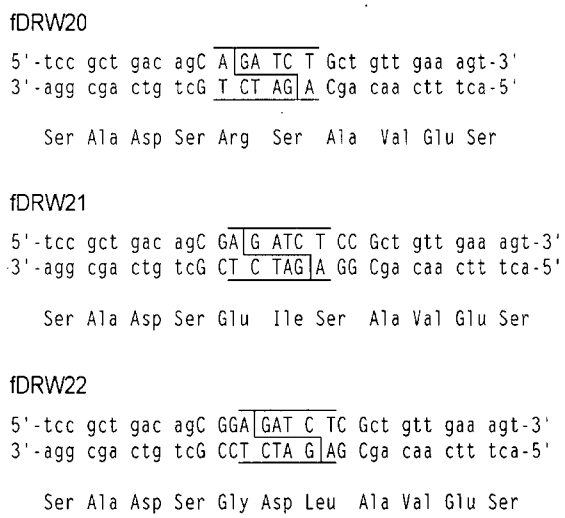


Fig. 7-12. Conventional vectors fDRW20, fDRW21 and fDRW22. Each vector possesses a unique *Bgl*II restriction site in a different reading frame, allowing insertion of DNA bearing *Sau*3AI-compatible cohesive ends in all three reading frames. Lower (parental fUSE1) and upper case (synthetic oligonucleotide) letters illustrate method of vector construction. A single self-complementary oligonucleotide was used to construct fDRW20. A pair of complementary oligonucleotides were used to construct fDRW21 and fDRW22; insertion of this oligonucleotide pair into fUSE1 in one orientation yielded fDRW21, while insertion in the opposite orientation yielded fDRW22.

(a) How can recombinants be distinguished? Not unsurprisingly, the principal problem from the outset was that of identifying FHA-S library recombinants against a large background of non-recombinants. Cloning efficiency appeared to be poor since transformants producing no virions, expected to account for two-thirds of recombinants, accounted for only 7% of total transformants (Table 7-III); this suggested that a smaller number of transformants possessed productive inserts. Indeed, markedly reduced virion production ($<2 \times 10^4$ mL^{-1} culture supernatant) by a similarly small number of transformants (Table 7-III) suggested that this phenotype identified recombinants, and restriction analysis confirmed that most transformants producing no or few virions were recombinants.

Table 7-III. Characteristics of FHA-S libraries.

Library	Molar ratio insert:vector[a]	Number of transformants[b]	Number of clones producing no virions[c]	Number of clones with diminished virion production[d]
I	1:1	1.4×10^5	3 of 96	2 of 96
II	5:1	1.7×10^5	4 of 96	3 of 96
III	25:1	1.8×10^5	8 of 96	9 of 96
IV	125:1	1.4×10^5	11 of 96	11 of 96
Overall			26 of 384 (7%)	25 of 384 (7%)

[a]Each library was constructed with 16 fmol of an equimolar mixture of fDRW20, fDRW21 and fDRW22, with the indicated molar ratio of fragments derived from a *Sau*3AI partial digest of pAS100.

[b]Estimates are based on a single 10 µL sample (1/200th of the products of transformation) spread on LB-Tet and incubated overnight (37°C).

[c,d]After overnight culture (37°C) of 96 transformants from each library in microtiter plates (200 µL LB-Tet per culture), 5 µL culture supernatant was applied to lawns of *E. coli*/pACYC184.

[c]"Clones producing no virions" are those producing <500 pfu mL^{-1} culture supernatant.

[d]"Clones with diminished virion production" are those producing $< 2 \times 10^4$ pfu mL^{-1} culture supernatant.

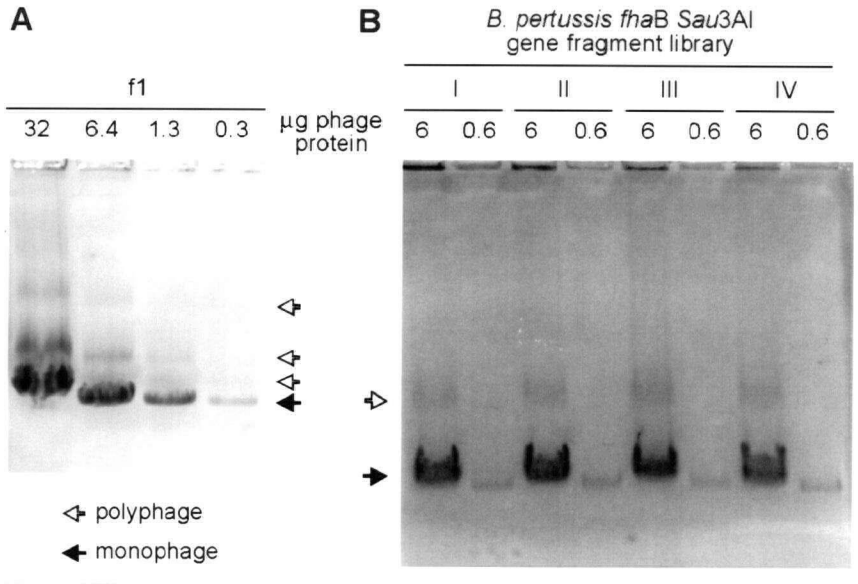


Fig. 7-13. Assessment by agarose gel electrophoresis of FHA-S library virions.

A. Coomassie-stained gels (§2.7.3) of the indicated quantities of CsCl-purified wild-type f1 virions.

B. PEG-precipitated virions harvested from each of four FHA-S libraries. Multiple bands are believed to represent virions possessing single (⬅, monophage) or multiple (⇦, polyphage) copies of the phage genome.

Nevertheless, it remained possible that the remaining large fraction (86%) of transformants included some recombinants that could produce virions at relatively high levels, perhaps approaching those of non-recombinants. This possibility was examined using Coomassie-stained agarose gels of electrophoresed virions (Griess *et al.* 1990), with the

expectation that because the cloned inserts were large (~10-25% of the vector genome) and virion length increases proportionately with the size of a cloned insert, differences in the migration of recombinants and non-recombinants would be readily detectable. However, for each of the FHA-S libraries, these gels revealed only a single dominant band as well as minor bands that likely represented, respectively, monophage and polyphage (Figure 7-13), indicating that recombinant virions were too few in number to be detected in this manner.

(b) Low titers reflect low numbers of physical particles. To confirm that the low titers shown in Table 3-III reflected low numbers of virions rather than low infectivity of a relatively larger number of virions, 100-fold concentrated virions produced by seven virion-producing FHA-S clones, together with unconcentrated virions produced by two CSP-library clones under similar culture conditions, were analysed by Coomassie-stained agarose gels (not shown). That CSP-library virions stained well while the 100-fold concentrated FHA-S virions did not stain detectably suggested that few virion particles had been produced.

(c) Can recombinants be enriched for? It remained possible that recombinants capable of producing large numbers of virions had gone undetected in the gel shown in Figure 7-13. Considering that Ff phage tolerate heating to 70°C (Salivar *et al.* 1964), it seemed possible to enrich for these by excising, from low melting point agarose gels, regions of the gel expected to contain recombinant virions, recovering the virions by melting the excised sections, and amplifying virions in *E. coli*. However, no apparent enrichment was evident in Coomassie-stained gels (not shown) of virions enriched and amplified in this manner, and subsequently concentrated (Materials and Methods, §2.11.1.f).

(d) Can a useful recombinant be identified? An expected minimum achievement of this work was the identification -- by sequencing -- of one or more recombinants encoding peptides of known properties (*i.e.*, known from published studies of FHA). These could be used, for example, as a source of DNA for sub-cloning to construct positive controls for later studies.

Earlier sequencing of CSP-library clones had been an almost trivial undertaking, made so by the high quality of s.s. template extracted from CSP-library virions. However, FHA-S clones produced so few virions (roughly $1/10^6$ to $1/10^7$ of the CSP-peptide clones) that sequencing required the use of d.s. RF template. Disappointingly, the quality of d.s. DNA extracted from overnight cultures of these clones was, like that of vector RF DNA (Figures 7-2 and following), of relatively poor quality and yielded correspondingly poor sequencing results. That this would become a recurring theme in later studies suggested that the low copy number (one-tenth that of wild-type phage fd; Smith 1988) and other properties of *ori(-)*-defective fd-tet -- and presumably also of its fDRW*n* derivatives -- seemed to ensure that RF DNA preparations would be routinely contaminated by a variety of host and phage-derived d.s. and s.s. DNA species.

(e) Implications. This brief study served to illustrate two points. The first was that only clones capable of producing reasonably large quantities of virions would be of practical value, not only to facilitate sequencing but presumably also in ELISA and functional assays. Efforts to display fragments as large as 400 to 800 amino acids had thus been unrealistic if not naive. Also evident was the value of vectors capable of producing libraries free of non-recombinants. Gene fragment or genomic PDLs constructed with *Sau*3AI partial digests might remain achievable goals, provided smaller fragments were targeted and a suitable vector employed.

7.3.4. Amber vector fDRW613, for *Sau*3AI fragment libraries

(a) Design strategy. The fDRW6*nn* vectors (Figure 7-14; also see Materials and Methods, Figure 2-2) were intended for construction of gene fragment and genomic *Sau*3AI fragment PDLs and incorporated several design principles. As with fDRW5, an amber-containing excisable stuffer was included. Unlike fDRW5, excision of the fDRW613 stuffer yielded a linear vector fragment possessing complementary cohesive ends and thus capable of self-ligation. Thus, the length of the stuffer was chosen to ensure that self-ligation would yield

a gIII frameshift.

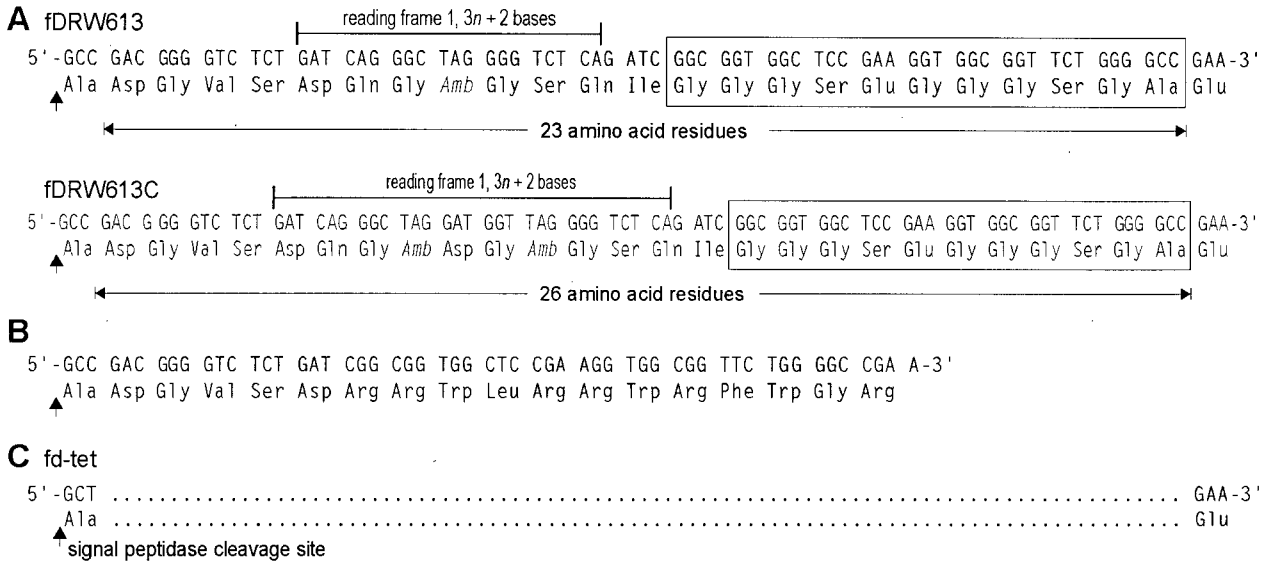


Fig. 7-14. Design of amber vector fDRW613 and two-amber vector fDRW613C. As outlined in Figure 2-2 (Materials and Methods), a series of amber-containing vectors were planned that would allow cloning, in all three reading frames, of gene fragments possessing *Sau*3AI-compatible ends. **A.** fDRW613, one of 6 planned vectors, provides for inserts of length $3n + 2$ (where n is an integer) in reading frame 1. Like fDRW5 (Figure 7-9), fDRW613 contains an excisable stuffer fragment with a single amber codon. Two amber codons are included in the fDRW613C stuffer fragment. Stuffer fragments are excised with *Bsa*I, which recognizes GGTCTC and cleaves as follows, leaving 5'-GATC overhangs at each end of the resulting fragment:

```
5′-GGTCTCN....-3′
3′-CCAGAGNNNNN-5′
```

B. Self-ligation of fDRW613 or fDRW613C (without the stuffer fragment) yields a gIII frameshift; since virion morphogenesis requires pIII production, no virions are produced. **C.** Wild-type gIII sequence of fd-tet serves to highlight the extent of gIII modifications in fDRW613 and fDRW613C.

A linker was added to separate a displayed peptide from pIII. A variety of generally short linkers have been employed in phage display (some are reviewed in Burritt *et al.* 1996). Examples include linkers with homorepeats of strikingly different properties, such as Gly_4 (*e.g.*, Hammer *et al.* 1992; Abrol *et al.* 1994; Kushwaha *et al.* 1994) and Pro_n sequences (*e.g.*, Devlin *et al.* 1990; Gram *et al.* 1993). The sequence GGGSEGGGSA was chosen for the principal reason that, as argued in Chapter 1, similar but longer GGGSE repeats appear to act as a linker between the pIII N-terminal knob and C-terminal stem. Notably, the related $(Gly_4Ser)_3$ sequence is commonly used to separate V_H and V_L domains of single-chain (scFv) antibodies

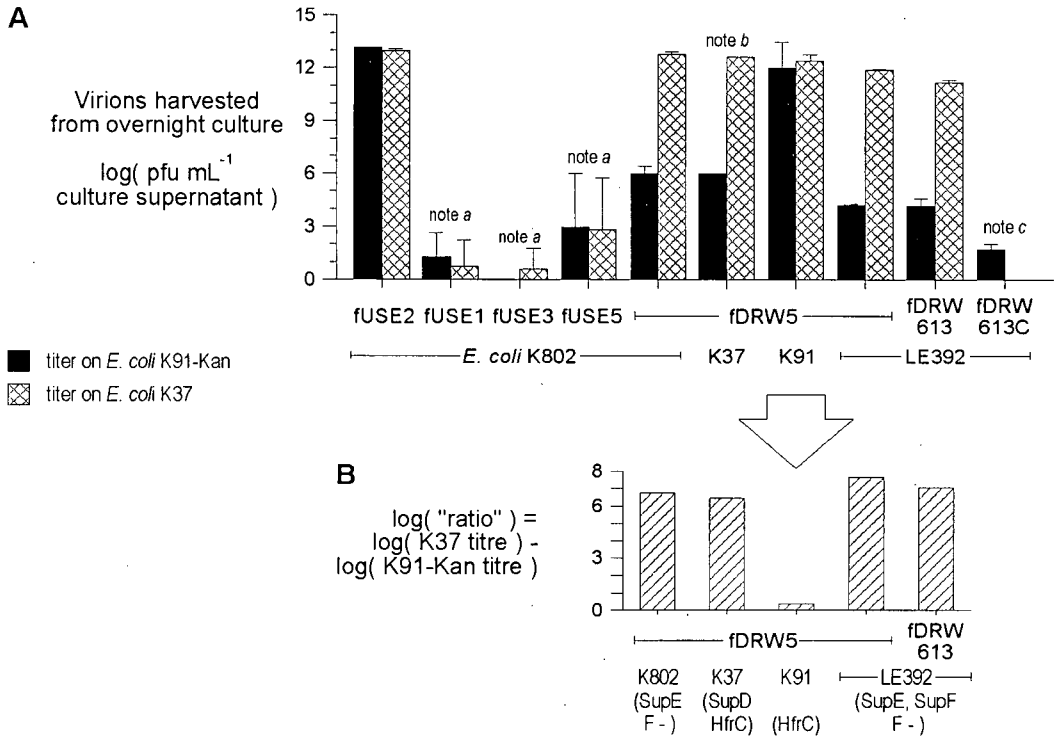
(*e.g.*, McCafferty *et al.* 1990) and has been shown, in a study that examined alternative linkers experimentally, to provide near-maximal flexibility between V_H and V_L domains with minimal proteolysis (Alfthan *et al.* 1995).

(b) A modified design with two amber codons. After finding that as anticipated fDRW613 produced pseudorevertants (data not shown; a related assay of fDRW613 stability is shown later), the idea of improving its ability to select against non-recombinants by incorporating two amber codons within the stuffer fragment was explored. As shown in the resulting design (Figure 7-14), this increased the number of vector residues foreign to native pIII from 23 in fDRW613 to 26 in fDRW613C. By comparison with fDRW5 which possesses only 8 foreign residues, fDRW613 and fDRW613C incorporate more extensively re-engineered pIII molecules. That this mattered was suggested in a comparative assay of pseudorevertant production by frameshift and amber vectors, described below.

7.3.5. Comparative assay of frameshift and amber vector pseudorevertant production

The comparative assay of pseudorevertant production summarized in Figure 7-15 was performed by propagating fUSE*n* and fDRW*n* vectors in a number of amber-suppressing and non-suppressing *E. coli* host strains, and titering pseudorevertants and total virions recovered from supernatants of overnight cultures. The two objectives of this assay were (i) to provide, in contrast to"troubleshooting" studies carried out earlier (*e.g.*, Figure 7-4, Table 7-I), a more formal, comparative assessment of pseudorevertant production by frameshift and amber vectors in various host strains; and (ii) to specifically assess pseudorevertant production by fDRW613C.

Several points emerged concerning the relative advantages of frameshift and amber vectors. The first was that frameshift vectors fUSE1, fUSE3 and fUSE5 *on average* (discussed later) produced fewer pseudorevertants than (i) amber vector fDRW5 propagated in the same host (*E. coli* K802) and (ii) fDRW*n* vectors generally, propagated in other *E. coli* strains (Figure 7-15A). The significance of this increased when yields of vector RF DNA were



C

Vector	fUSE 2	fUSE 1	fUSE 3	fUSE 5	fDRW5			fDRW 613	fDRW 613C
Host strain	*E. coli* K802				K37	K91	LE392		
Relative quantity of RF DNA recovered	1.2	1.5	1.0	4.0	0.9	1.8	1.0	1.0	1.0	1.0

Fig. 7-15. Frameshift and amber vector stability as determined by pseudorevertant production. **A.** Titers of virions harvested from overnight cultures of ten host/vector/plasmid combinations. Titers on *E. coli* K37 and K91-Kan are expected to be similar for each of the fUSE*n* vectors, since amber-suppression (*E. coli* K37) confers no obvious advantage to these vectors. Virion production by fUSE2 is a control for relative plaquing efficiency of *E. coli* K37 and K91-Kan, as well as a control for wild-type (with respect to in-frame expression of gIII) virion production levels. Titers for amber vectors fDRW5, fDRW613 and fDRW613C on the amber-suppressing *E. coli* host strain K37 reflect total virion production, while titers on *E. coli* K91-Kan reflect production of pseudorevertants, ostensibly resulting from loss of the amber codon within these vectors. Values shown are the mean log(titer) ± 2 standard errors, for three independently cultured and processed isolates. **B.** Values are derived from those shown in **A** and represent the log of the ratio of total virions to pseudorevertants. **C.** Relative quantities of RF DNA recovered from cultures assayed in **A** for virion production. Values shown are relative to the quantities recovered for fDRW5 in *E. coli* LE392; they represent a mean for three cultures and are a based on a cursory examination of the relative intensity of bands after agarose gel electrophoresis of undigested RF DNA.

[a]In repeated experiments, standard errors for pseudorevertant production by frameshift vectors fUSE1, fUSE3 and fUSE5 were characteristically greater than those for amber vectors such as fDRW5 and fDRW613 as well as for conventional vector fUSE2. [b]For reasons explained in Materials and Methods (§2.9.3.b), values for fDRW5/K37 are means of duplicate cultures; other values are means of triplicates ± 2 standard errors. [c]fDRW613 plaques poorly or not at all on *E. coli* K37 (Table 2-IV). In later experiments *E. coli* K91/pNK1759 was employed to plaque virions produced by amber vectors.

considered. For example, after overnight culture in *E. coli* K802, fUSE5 produced 10^3-fold fewer pseudorevertants than its amber offspring fDRW5 (Figure 7-15A) but yielded four-fold more RF DNA (Figure 7-15C). In practical terms, 4 x 10^3-fold fewer pseudorevertants would be expected in fUSE5 *versus* fDRW5 libraries.

A second and related point, illustrated by the magnitude of the error bars (Figure 7-15A) and confirmed in repeated experiments, is that pseudorevertant production was more variable in frameshift than in amber vectors. Thus the fUSE5 error bars reflect three independent cultures that produced ~10^5, 10^4 or no detectable pseudorevertants (per mL of culture supernatant). One explanation for this difference in variability is that frameshift mutations leading to fUSE5 pseudorevertants would occur less frequently than base substitutions leading to fDRW5 pseudorevertants (generally, frameshift mutations are less common than base substitutions; Hutchinson 1996), but that they would confer a greater selective advantage by restoring cell health. From this, and the idea that the observed variability reflected the times at which mutations arose, it seemed possible that early mutations would be more likely to yield "jackpot" levels of pseudorevertants in fUSE5 than in fDRW5. In view of this, and considering earlier experiences with frameshift vectors such as the occurrence of apparent deletion mutants, amber vectors appeared more promising for future work.

An influence of the host strain chosen for propagating amber vector fDRW5 was also apparent. As in the earlier cursory study (Table 7-I), the SupE/SupF strain LE392 produced 10-fold fewer pseudorevertants (Figure 7-15A and B) than did K802 (SupE) and K37 (SupD), even after adjusting for quantities of RF DNA recovered (Figure 7-15C). Particularly striking were the effects of propagating fDRW5 in the *non*-amber-suppressing F-piliated *E. coli* host strain K91, where pseudorevertant titers were indistinguishable from total yields. This validated the earlier idea that pseudorevertants might come to dominate a population when the frame-shift vector fUSE5 is propagated in F-piliated hosts (§7.3.1.b).

A final point concerns a possible role of the inserts required to construct the cloning site and linkers of fDRW613 and fDRW613C. Total virion production by fDRW613 (possesses 23 foreign residues near the N-terminus of mature pIII) was reduced 5-fold compared to fDRW5 (8 foreign residues) propagated in the same strain (LE392). More dramatically, no plaques could be detected for fDRW613C (three foreign residues more than fDRW613). Although as described later, fDRW613C plaques can be detected on an alternative plaquing strain, the idea that relatively small inserts could produce such noticeable effects caused me to question the feasibility of constructing phage display libraries with *Sau*3AI fragments[12].

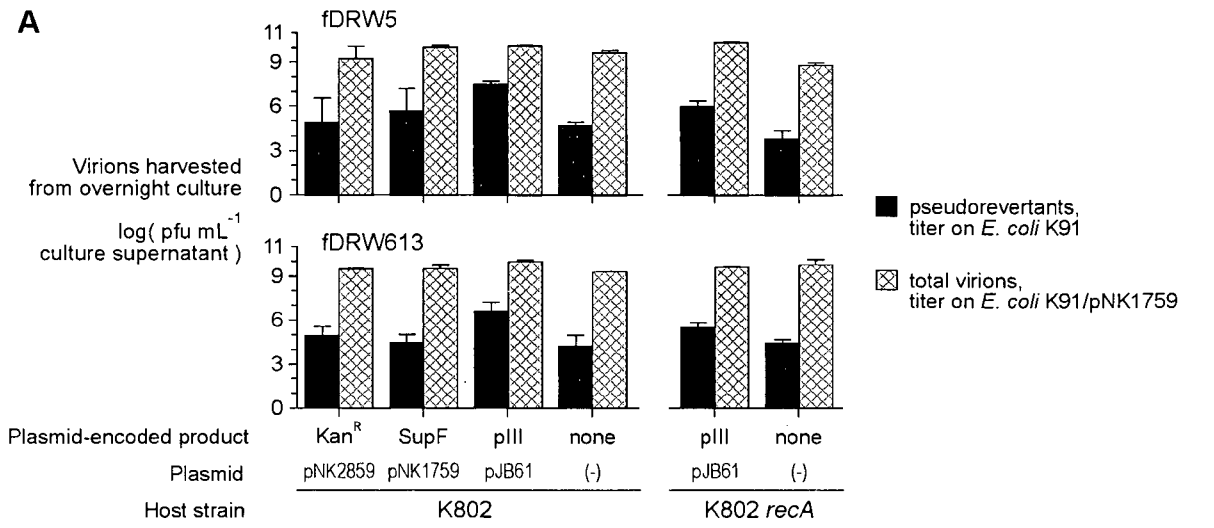
(a) Improving vector stability. On the assumption that the magnitudes of the pseudo-revertant titers reflected a competitive advantage enjoyed by these mutants (this was not demonstrated experimentally), two alternatives for improving vector stability by obviating this advantage were considered. These included propagating amber vectors (i) in a host harboring a pIII-encoding plasmid, and (ii) in a host carrying additional chromosomal copies of *supF* (as mini-transposon insertions). A preliminary study exploring these alternatives employed host strains carrying a pBR322 derivative encoding pIII (pJB61) and two pBR322-derived transposon mutagenesis vectors encoding SupF (pNK1759) and kanamycin resistance (pNK2859; a control).

Notwithstanding possible confounding effects of transposition events, assays of pseudo-revertant production by fDRW5 and fDRW613 propagated in hosts carrying these plasmids were disappointing (Figure 7-16). Thus, for both fDRW5 and fDRW613, propagation in *E. coli* K802 strains harboring pNK1759 (SupF), expected to reduce pseudorevertant numbers, instead yielded numbers similar to those for vector propagation in K802 harboring the control plasmid pNK2859 (Kan^R). Strikingly, propagation in K802 harboring pJB61 (pIII), also expected to reduce pseudorevertant numbers, instead yielded a 60- (fDRW613) to 400-fold

---

[12]Average theoretical size of a single *Sau*3AI fragment is 256 bp in an organism of random sequence and %G+C=50.

(fDRW5) increase in pseudorevertants compared to K802 control strains harboring pNK2859 or no plasmid. The ~10-fold lower production of pseudorevertants by a K802 *recA* strain harboring pJB61 suggests that much of the increase in (pseudo-)revertant numbers was due to homologous recombination.



B

	Plaque formation on *E. coli* ± pJB61			
	K37	K37/pJB61	K91-Kan	K91-Kan/pJB61
fUSE2	yes	no	yes	no
fDRW613	yes	no	n.d.[a]	n.d.

Fig. 7-16. Effects of SupF and wild-type pIII on pseudorevertant production. A. Titers of virions harvested from overnight cultures of amber vectors fDRW5 and fDRW613 propagated in *E. coli* K802 and K802 *recA* harboring (i) pJB61 (encodes wild-type pIII), (ii) pNK1759 (SupF), (iii) pNK2859 (Kan^R, control) or (iv) no plasmid. Titers on *E. coli* K91/pNK1759 reflect total virion production, while titers on *E. coli* K91 reflect production of pseudorevertants, *i.e.*, virions capable of plaquing on this non-amber-suppressing host, ostensibly because of a loss of the amber codon within these vectors. Values shown are the mean log(titer) ± 2 standard errors for three independently cultured and processed isolates. **B.** Control to confirm pJB61 gIII expression. Ca. 10^1-10^3 fUSE2 and fDRW613 pfu were plaqued on the indicated host strains. Since pIII prevents super-infection, no plaques were expected on strains harboring pJB61.

[a]n.d., not determined: except for pseudorevertants, fDRW613 plaques only on amber-suppressing hosts.

7.3.6. Is a gene fragment library possible? *B. pertussis fhaB* restriction fragment libraries constructed with frameshift vector fUSE1

As discussed earlier (§7.3.5), diminished plaque production for fDRW613 and lack of

fDRW613C plaques prompted me to question whether display of even moderately sized gene fragments would be feasible. The issue was addressed directly by constructing a set of libraries (FHA-H/H/H libraries) displaying a wide assortment of different-sized fragments derived from pAS100, a plasmid carrying 10 kbp of the *B. pertussis fhaB* gene. This was done by cloning *Hha*I, *Hin*P1I and *Hpa*II restriction fragments (Table 7-IV) into frameshift vector fUSE1 and taking certain measures (Materials and Methods, §2.11.2) to ensure that virtually all transformants were recombinants.

Table 7-IV. FHA-H/H/H libraries: distribution[a] of *Hha*I/*Hin*P1I/*Hpa*II fragment sizes.

	Restriction fragment size (base pairs)				
Restriction endonuclease[b]	<20	20-49	50-99	100-249	250-484
*Hha*I and *Hin*P1I	89	64	42	24	2
*Hpa*II	16	13	22	31	8

[a]Body of table shows, for each range of sizes, the number of fragments expected for a complete digest of pAS100 with the indicated restriction enzymes. pAS100 is comprised of the vector pTZ18R and the 10 kbp *Eco*RI fragment of *B. pertussis fhaB*. [b]Each of these recognize G and C-containing tetranucleotides and were thus expected to cut frequently within *B. pertussis* DNA (%G+C=66-70%). HhaI and *Hin*P1I recognize GCGC, *Hha*I producing a 5' overhang, *Hin*P1I, a 3' overhang. *Hpa*II recognizes CCGG.

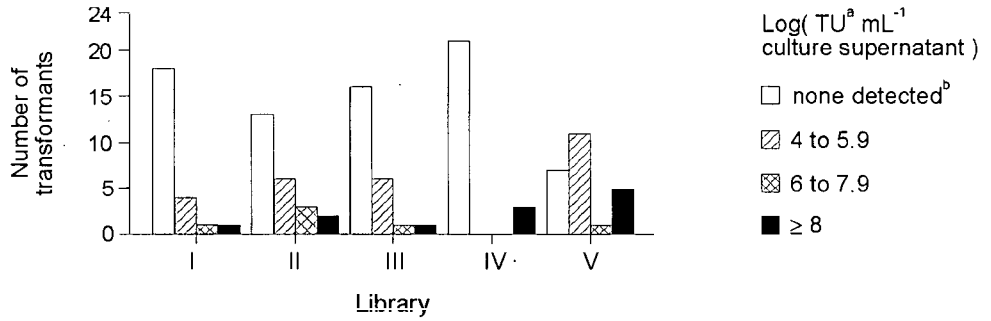


Fig. 7-17. Screen for virion-producing transformants in FHA-H/H/H libraries. Chart summarizes virion production by twenty-four transformants from each of five libraries (I, II, ..., V), as measured by the ability of progeny virions to transduce tetracycline resistance into *E. coli* host cells (Materials and Methods, §2.11.2.e). Overall, 38 per cent (45 of 120) of transformants produced a detectable number of virions. All transformants were expected to be recombinants (see accompanying text and Materials and Methods, §2.11.2). [a]Lower limit of detection is 10^4 TU mL^{-1}. [b]TU, transducing unit.

(a) Virion production varies >10^4-fold among clones. Results of a first survey of library transformants (Figure 7-17) were consistent with expectations that (i) all would be

recombinants, and (ii) that virion production would vary significantly among clones. Thus, an expected fraction (62% *versus* a predicted 67%) of transformants produced no virions (arguably because they possessed inserts that did not restore the fUSE1 gIII reading frame), while the remaining (gIII-restoring) transformants produced virions, at levels that varied over a $>10^4$-fold range.

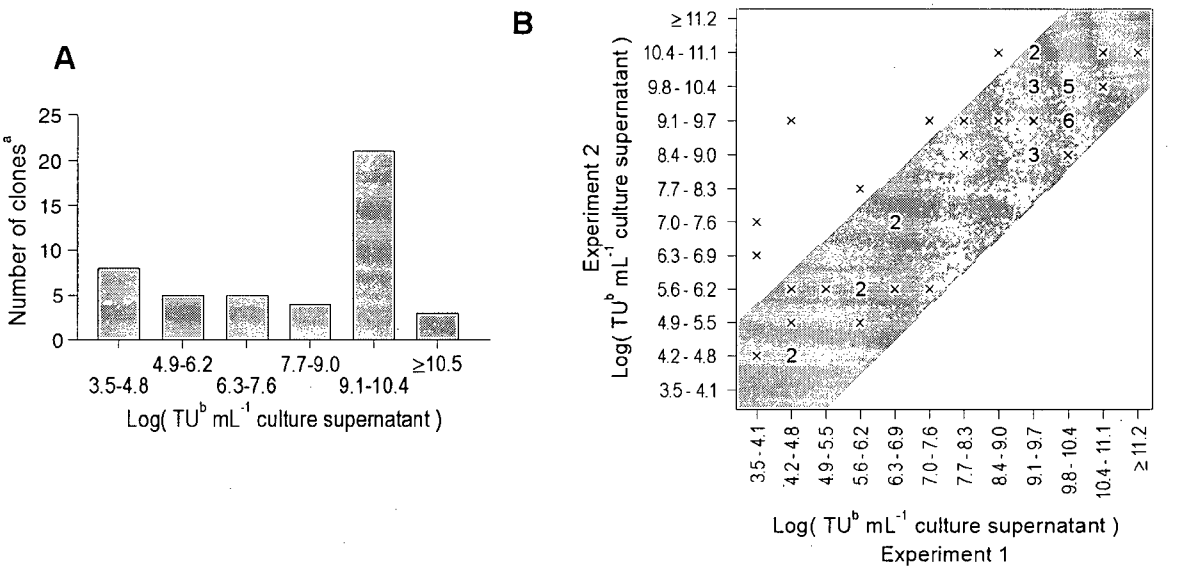


Fig. 7-18. Virion production by selected FHA-H/H/H library clones. Transformants producing a detectable number of virions (Figure 7-17) were assayed for virion production measured as transducing units. Chart **A** categorizes clones by level of virion production in the first (Experiment 1) of two repeated assays. In **B**, the results of the first (Experiment 1) and second (Experiment 2) assays are compared for each clone. Titers are plotted as discrete points for the indicated ranges. Integers 2, 3, 5 and 6 indicate the number of data points that plotted to the indicated positions. The shaded diagonal band in **B** represents a "reasonable" range of expected values for the two experiments. [a]A total of 45 clones were assayed. [b]TU, transducing unit.

(b) Some clones may be unstable. Followup assays of titers of virion-producing clones were performed with the goals of (i) more precisely defining the range of variability in virion production among these clones, and (ii) identifying clones that would be further characterized, by restriction analysis and sequencing, in order to assess the influence of insert size and peptide sequence on virion production. The first assay (Figure 7-18A) showed that while more than half of the clones produced virions at "reasonable" levels ($\geq 10^9$ mL^{-1} culture supernatant,

arbitrarily defined), a significant fraction produced fewer virions such that there was a $>10^7$-fold range in titers among all clones. Curiously, these results were not repeatable for some clones. Thus, while most clones produced similar numbers of virions in a second assay, others produced as many as 10^3-fold more (Figure 7-18B). Also, clones showing the greatest variability tended to have low titers in at least one (generally the second) assay. A tentative conclusion was that these represented unstable clones, where deletions or other mutations, arising at different times during culture, resulted in the variable titers.

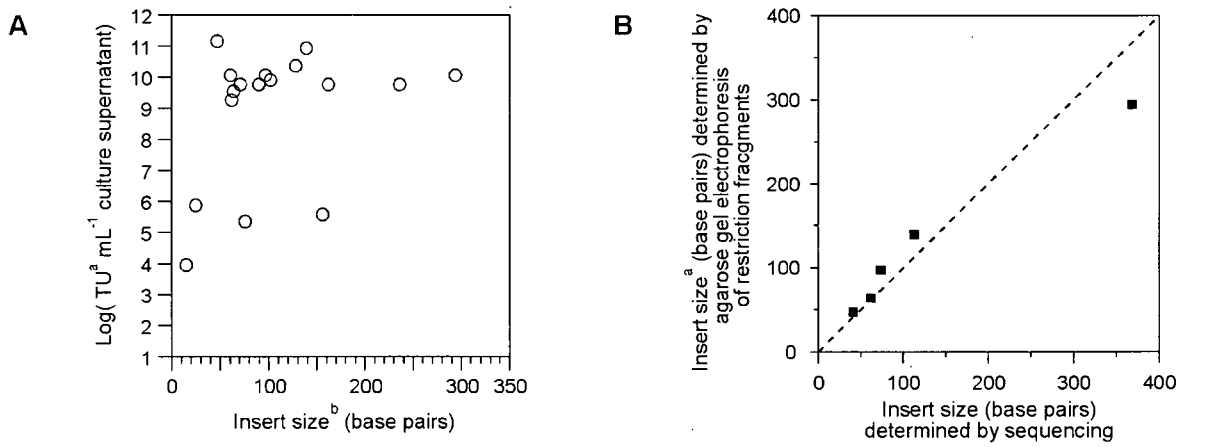


Fig. 7-19. Assessment of influence of insert size on virion production by selected FHA-H/H/H library clones. Chart **A** represents 17 virion-producing clones whose insert sizes could be determined by agarose gel electrophoresis of *Hae*III restriction fragments of RF DNA. Transducing units reported for these clones correspond to those in Figure 7-18 (Experiment 1). Chart **B** compares, for five clones producing high levels of virions ($\geq 10^9$ mL^{-1} culture supernatant), insert sizes shown in **A** with those determined by sequencing. [a]TU, transducing unit. [b]Insert sizes determined by agarose gel electrophoresis represent the sum of visible fragments and accordingly represent a "minimum" insert size; sizes of inserts cleaved more than once by *Hae*III to produce fragments smaller than 236 base pairs may be underestimated.

(c) Virion production does not correlate with insert size. Restriction analysis of RF DNA extracted from each of the 45 virion-producing clones, performed to provide an estimate of insert sizes, proved unexpectedly difficult. For many clones, and in particular for clones producing few virions, d.s. RF DNA preparations were of poor quality (*e.g.*, a background "smear" of high- to low-molecular weight species was pronounced on agarose gels, unexpected species were present) or of low yield. Nevertheless, reasonable estimates of insert sizes were

obtained for 17 clones, and as suggested by a comparison of these estimates with previously determined virion titers, virion production does not correlate with insert size (Figure 7-19A). Repeated efforts to sequence selected clones -- performed in part with the goal of assessing the influence of a peptide sequence on virion titers -- were largely unsuccessful. In many cases, virion titers were too low to allow preparation of adequate quantities of s.s. template, and as before the quality of d.s. sequencing template was low. However, sequencing results for five clones were of sufficient quality to confirm that restriction analysis-derived estimates of insert size were reasonable (Figure 7-19B).

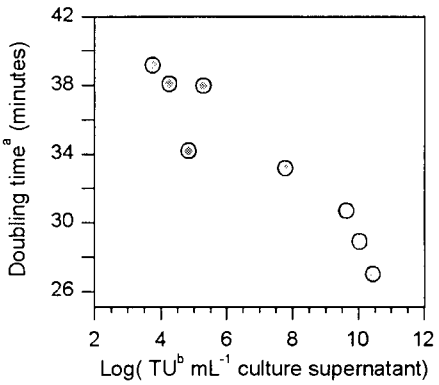


Fig. 7-20. Correlation of virion production with host cell doubling time, selected FHA-H/H/H library clones. Virion production in overnight cultures and host cell doubling times were estimated, in non-concurrent assays, for 12 selected clones. Values shown are means of duplicate cultures.
[a]Based on a least squares linear regression of log(OD_{600}) *versus* time for OD_{600} values in the range 0.04-0.3.
[b]TU, transducing unit.

(d) Host cell health correlates with virion production. Given difficulties in restriction analysis and sequencing, few straightforward alternatives existed for continued assessment of the influence of an insert on virion titers. An indirect view was provided by the experiment summarized in Figure 7-20, which shows a strong correlation between host cell doubling time (T_d) and virion production. Briefly, virion titers were highest in infected host cells with the shortest doubling times (T_d). The 1.5-fold range of T_d values (27 to 39 minutes) suggests that the selective advantage of mutation could be large in clones producing few virions, and is consistent with the earlier idea (§7.3.6.b) that insert instability accounted for unexpectedly high experiment-to-experiment differences in titers of such clones.

(e) Selection pressures define a minimum threshold of virion production. Insert instability was examined in a conceptually simple experiment. The results (Figure 7-21) reflect

figure shown on following page

Fig. 7-21. Virion production by 1° and 2° clones of FHA-H/H/H library members. For each of 40 library members, virion production by 1° clones is compared with that of 2° clones. For convenience (see also footnote *a*), 1° clones were cultured in *E. coli* K802 and 2° clones in *E. coli* K91. In **A**, data are plotted as ranges (indicated by horizontal and vertical lines, ——) and as means (indicated by ■ at the point of intersection of horizontal and vertical lines). In **B**, only the means are shown. The shaded diagonal band represents an arbitrary "reasonable" range of expected values for 1° and 2° clones possessing identical inserts in phage gIII. Grid lines in **A** and **B** correspond to those in panels **C** to **H**; summary data in panels **C** to **H** are discussed in the text.

[a]As explained in Materials and Methods (§2.11.3), RF DNA was extracted from early cultures of virion-producing FHA-H/H/H library transformants. Later, 40 library members were recovered by transforming *E. coli* K802 with this RF DNA. For each library member, three transformants were selected as 1° clones. As well, three 2° clones were derived by infecting *E. coli* K91 with virions harvested from the same initial cultures from which the RF DNA was extracted.

[b]Maximum, [c]Minimum, [d]Mean value, for a given library member, of its three 1°(X-axis) or 2° clones (Y-axis).

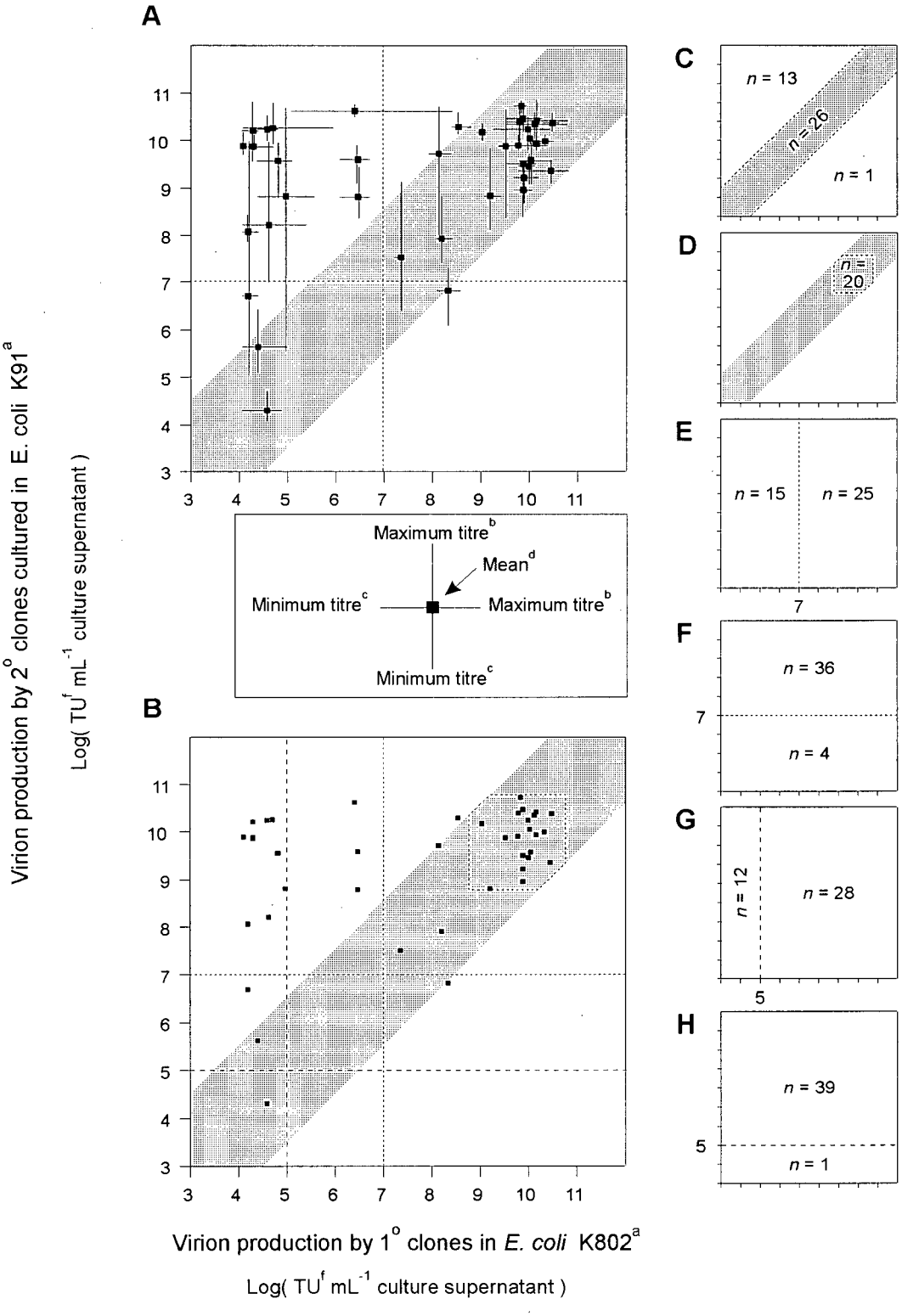

A
B
C
D
E
F
G
H
Maximum titre[b]
Mean[d]
Minimum titre[c]
Maximum titre[b]
Minimum titre[c]
n = 13
n = 26
n = 1
n = 20
n = 15
n = 25
n = 36
n = 4
n = 12
n = 28
n = 39
n = 1
Virion production by 2° clones cultured in E. coli K91[a]
Log(TU[f] mL[-1] culture supernatant)
Virion production by 1° clones in E. coli K802[a]
Log(TU[f] mL[-1] culture supernatant)

Fig. 7-21.

two ideas. First, a 2° clone derived by infecting an *E. coli* host with a virion produced by a parental (1°) clone will yield titers similar to those of the 1° clone if the 1° and 2° clones encode the same sequence (the 1° clone is stable), or higher titers otherwise (the 2° clone is a mutant). Secondly, titers of a 1° clone and its 2° clones will vary little from culture to culture if the 1° clone is stable, but may vary substantially otherwise. Thus, for each of 40 FHA-H/H/H library members, Figure 7-21 compares virion titers of three 1° clone cultures with those of three 2° clones. The data are presented with horizontal and vertical bars (panel A) to illustrate variability, without bars (panel B) for clarity, and in "snapshot" views (panels C to H) for discussion. Two points emerged from the data.

First, for more than half of the 40 clones (the 26 of panel C) titers of 1° and 2° clones were similar (panel B), suggesting that these clones were stable, although variability (panel A) in titers of some of these suggested otherwise. Most of the apparently stable clones (the 20 of panel D) produced relatively high numbers of virions ($\sim 10^{10}$ mL^{-1}), although other apparently stable clones produced as few as $\sim 10^{5}$ virions mL^{-1} (see the shaded diagonals in A or B). Secondly, changes in titer between 1° and 2° clones were greatest for 1° clones producing the lowest titers, as illustrated by arbitrary "low-titer" boundaries in panels E *versus* F, and in G *versus* H. The general trend of the data thus seemed to indicate that a threshold titer (say, of $\sim 10^{7}$ mL^{-1}) determined whether a clone was stable. This trend, in context of the correlation of titer with host cell doubling times (Figure 7-20), suggested that insert stability reflected selection pressures related to host cell viability.

(f) Clones producing few virions possess frameshifts in gIII. As before, attempts to characterize some of these clones by sequencing were frustrated by the poor quality of d.s. sequencing template, in spite of attempts (partially successful) to improve template quality (Materials and Methods, §2.11.3.c). Nevertheless, the limited sequencing data allowed partial characterization of five clones (Figure 7-22). Surprisingly, both of the two sequenced low-titer

1° clones were found to possess inserts such that gIII was out of frame. For these library members, the corresponding 2° clones possessed smaller inserts, and gIII had been shifted into the correct reading frame (Figure 7-22, *a* and *b*). In the remaining three, higher-titer 1° clones, gIII was found to be in the correct reading frame (Figure 7-22, *c* to *e*). Finding clones with gIII out-of-frame was unexpected, for the initial screen (Figure 7-17) had assumed that all virion-producing transformants possessed inserts such that gIII was in frame. It was now evident that some, or perhaps many, low virion titers reflected mutations arising during culture of recombinants in which gIII was initially out of frame.

A

Library member	1° clones		2° clones	
	gIII in frame?	Insert size[a] (bp)	gIII in frame?	Insert size[a] (bp)
ⓐ	no	88, 88	yes	62, 62
ⓑ	no	132	yes	131, 97
ⓒ	yes	110	yes	110
ⓓ	yes	128	n.d.[b]	n.d.
ⓔ	yes	113	n.d.	n.d.

B

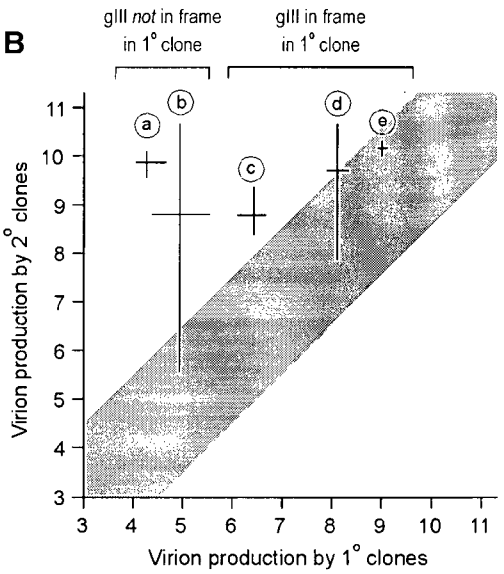


Fig. 7-22. Characterization of selected FHA-H/H/H library 1° and 2° clones. For each of five library members (*a* to *e*), table in **A** shows results of sequencing 1° and 2° clones. The graph in **B** is a snapshot of Figure 7-21A in which only clones *a* to *e* have been included. Sizes of cloned inserts, derived by sequencing, of 1° and 2° clones of five library members (*a* to *e*) are shown in **A**. [a]Where two values are shown, these indicate that two clones possessed inserts of the indicated sizes. [b]n.d., not determined.

The earlier-interpreted data of Figure 7-21 could now be re-interpreted as suggesting that (i) 1° clones in which pIII was in-frame (the 26 of Figure 7-21C) generally produced relatively high virion titers, and (ii) the remaining 1° clones, producing no pIII, gave rise to pIII-producing 2° clones (mutants) that by restoring cell health enjoyed considerable selective advantage. The correct interpretation is likely a combination of both sets of interpretations, for

the reason that for some library members (*e.g.*, Figure 7-22B, library member *d*), 1° clones yielded high titers of low variability, suggesting that gIII was initially in frame, while their 2° clones yielded higher titers of greater variability, suggesting a mutation.

The encouraging finding in this was that at least half of the library members (*viz.*, the 20 of Figure 7-21D) appeared stable and produced virions that varied only over a ~100-fold range. Subjectively, the practical consequence of such variability was expected to be limited to making it necessary to harvest virions for sequencing, ELISA or other assays from large cultures (*e.g.*, ≥2 L) of the affected clones. More important consequences -- such as in affinity-selection, library amplification and immunodetection via plaque lifts -- were expected for other seemingly stable clones that yielded consistently low titers ($\sim 10^4$ mL^{-1}). Conceptually, the instability of clones harboring gIII frameshifts could be seen as a benefit that increased library diversity, for a certain fraction of gIII frameshift "corrections" would be expected to lead to display of peptides including native FHA sequences. Indeed, instability of other clones might be seen in a similar light.

7.3.7. A strategy for gene fragment and genomic libraries

These findings and other experience to date suggested the following strategy for constructing gene fragment and genomic PDLs.

(a) Maximize redundancy. Including as many overlapping sequences as possible in a library, such as by using random DNA fragments produced by DNase I digestion (Anderson 1981), would allow a given target peptide to be expressed in a variety of sequence contexts, hopefully including some favorable for both expression and affinity selection.

(b) Minimize clonal competition. Possible approaches included (i) use of vectors able to reduce the fraction of non-recombinants, (ii) construction and separate processing of multiple libraries displaying different ranges of sizes of displayed peptides, and (iii) favoring screening over affinity selection for identifying target clones, thereby obviating the need for amplification

of affinity-enriched fractions.

(c) Keep vectors simple. Poor or no plaque formation by fDRW613 and fDRW613C suggested that apparent improvements can actually interfere. Thus, the design of fDRW70 described below incorporated only those pIII modifications that were required for creating a cloning site or previously shown to be effective or at a minimum innocuous.

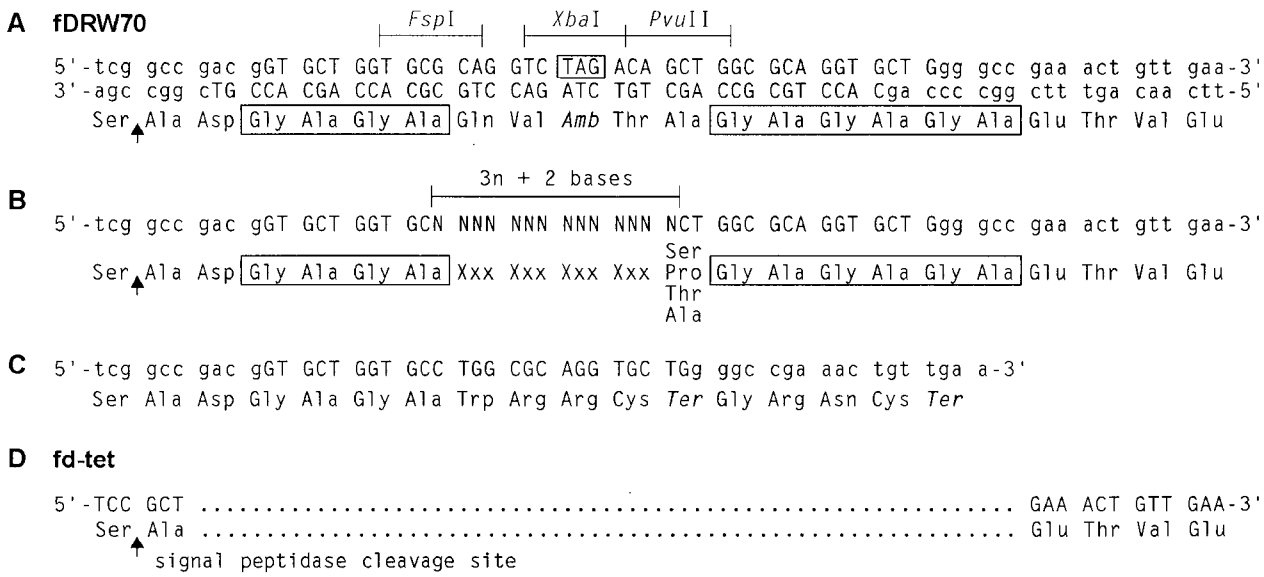


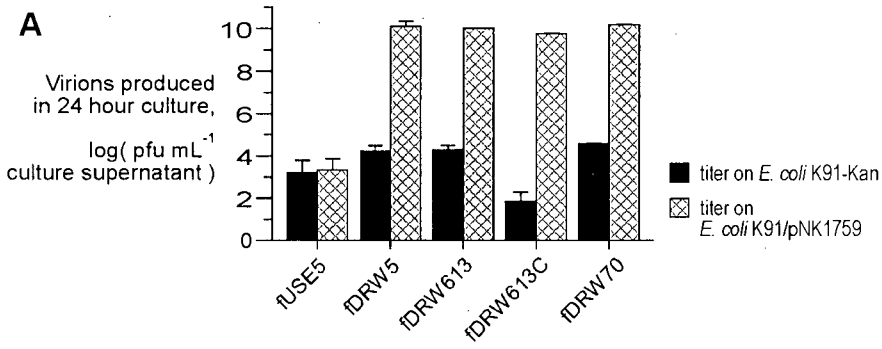
Fig. 7-23. Design of amber vector fDRW70. **A.** *Fsp*I and *Pvu*II restriction sites allow removal of a "stuffer" fragment containing an amber codon (TAG), thereby creating a linear fragment suitable for receiving blunt-end inserts. The *Xba*I site can be employed to ensure that the amber codon is intact in individual fDRW70 isolates. In non-amber-suppressing host strains, production of pIII (required for virion propagation) is prevented in undigested vector possessing the amber codon. Linker sequences Gly-Ala-Gly-Ala and Gly-Ala-Gly-Ala-Gly-Ala are incorporated into the vector. Upper case (parental fDRW5) and lower case (synthetic oligonucleotide) sequences illustrate method of constructing fDRW70. **B.** Productive inserts (those in which gIII is in the correct reading frame) must contain $3n + 2$ bases (where n is an integer). **C.** In the event of vector self-ligation, gIII will possess a frameshift, preventing pIII production. **D.** Wild-type sequence of fd-tet, provided for comparison with fDRW70.

7.3.8. Design of amber vector fDRW70, and continued assessment of other vectors

(a) Vector design. fDRW70 (Figure 7-23) was developed for cloning DNase I-generated gene fragments of a single length ($3n$+2 bp, where n is an integer). Although multiple vectors with different reading frames, such as the fDRW2*n* and fDRW6*nn* series, are required to maximize redundancy in restriction fragment libraries, they serve little purpose with DNase I-

derived fragments. Like fDRW613, fDRW70 includes an excisable amber-containing stuffer which upon removal and subsequent vector re-ligation yields a gIII frameshift. The stuffer includes an *Xba*I restriction site that can be employed diagnostically to ensure that the amber codon is intact in a given RF DNA preparation. fDRW70 also encodes Gly-Ala linkers similar to those encoded by the oligonucleotides used to construct the successful CSP-libraries (Figure 7-8).

(b) Pseudorevertant production by fDRW70 and other vectors. Assessment of fDRW70 pseudorevertant production provided a convenient time to examine unresolved issues with other vectors. The common results, derived from titering pseudorevertants and total virions recovered from overnight cultures of fDRW70 and other vectors in *E. coli* LE392, are shown in Figure 7-24.



B

Vector	Relative quantity of RF DNA recovered
fUSE5	3.2
fDRW5	1

Fig. 7-24. Stability of fUSE5 and amber vectors, as determined by pseudorevertant production. **A**. Titers of 24-hour virion production by frameshift vector fUSE5 and amber vectors fDRW5, fDRW613, fDRW613C and fDRW70 propagated in *E. coli* LE392. For amber vectors, titers on amber-suppressing *E. coli* K91/pNK1759 reflect total virion production, while titers on *E. coli* K91-Kan reflect pseudo-revertant production. For fUSE5, titers on *E. coli* K91/pNK1759 and K91-Kan are expected to be similar since amber-suppression confers no obvious advantage to this vector. Values shown are the mean log(titer) ± 2 standard errors for four (fUSE5, fDRW5) or three (fDRW613, fDRW613C, fDRW70) independently cultured and processed isolates. **B**. For fUSE5 and fDRW5, the mean (n=4) relative yields of RF DNA extracted from the same cultures from which virion titers were derived, as determined by intensity of fluorescence of samples on agarose gels.

fDRW70 produced similar numbers of pseudorevertants as fDRW5 and fDRW613 (Figure 7-24A). As before (Figure 7-15), culture-to-culture variability of these amber vectors was low. The implied predictability of this suggested that an inherent, unavoidable mechanism

accounted for the mutations required for pseudorevertant production.

Earlier attempts to examine the effect of two amber codons in fDRW613C had been frustrated by lack of plaque formation of fDRW613 on *E. coli* K37 (Figure 7-15). New data for fDRW613C now showed that use of two amber codons reduced pseudorevertant production ~100-fold (Figure 7-24A), suggesting that this may be a useful strategy to consider in future vector designs. The new data were obtained by using *E. coli* K91(F-piliated)/pNK1759(SupF) instead of K37 (F-piliated, SupD) as a plaquing strain to estimate total virion production. As shown elsewhere (Materials and Methods, Table 2-IV), strain K91/pNK1759 not only produces plaques for fDRW613C, but also produces almost 2-fold more plaques for fDRW70 and 5-fold more for fDRW613 than does strain K37. As well, and in contrast to the variability seen earlier with strain K37 (Figure 7-15), total virion numbers estimated using strain K91/pNK1759 as a plaquing strain varied little among fDRW5, fDRW613 and fDRW613C (Figure 7-24A). Since both K37 and K91 are derivatives of K38 (Smith 1988), the difference may lie in differences in amber-suppression, since K37 is SupD (chromosomal) and K91/pNK1759 is SupF (within a minitransposon on a transposon mutagenesis plasmid). Notwithstanding this possibility, it was both curious and bothersome that choice of plaquing strain could so readily influence (possibly even confound) simple assays.

(c) Comparison of frameshift vector fUSE5 with fDRW5. As before (Figure 7-15), frameshift vector fUSE5 produced fewer pseudorevertants than its amber counterpart fDRW5 (Figure 7-24A), although the 10-fold difference seen here using strain LE392 for vector propagation was less striking than the 10^3-fold difference seen earlier with strain K802 (Figure 7-15A). Both fUSE5 and fDRW5 produced RF DNA of a single species that could be digested to apparent completion with *Sfi*I (Figure 7-25). The general quality of fUSE5 RF DNA appeared to have improved over earlier preparations. It is possible that use of *E. coli* host strain LE392 (*versus* K802 and K91 in earlier preparations) accounted for this difference, for

the quality of fDRW5 RF DNA derived from LE392 had earlier been found (Table 7-I) to be better than that derived from strains K802 and K37. These findings, and the 3.2-fold greater yields of fUSE5 *versus* fDRW5 RF DNA (Figure 7-24B) now suggested that, apart from the relative unpredictability of frameshift vector pseudorevertant production, frameshift vectors may be of greater utility than amber vectors when propagated in an appropriate host strain. As with choice of plaquing strains, it was again curious and bothersome that choice of strains for vector propagation might have such noticeable consequences.

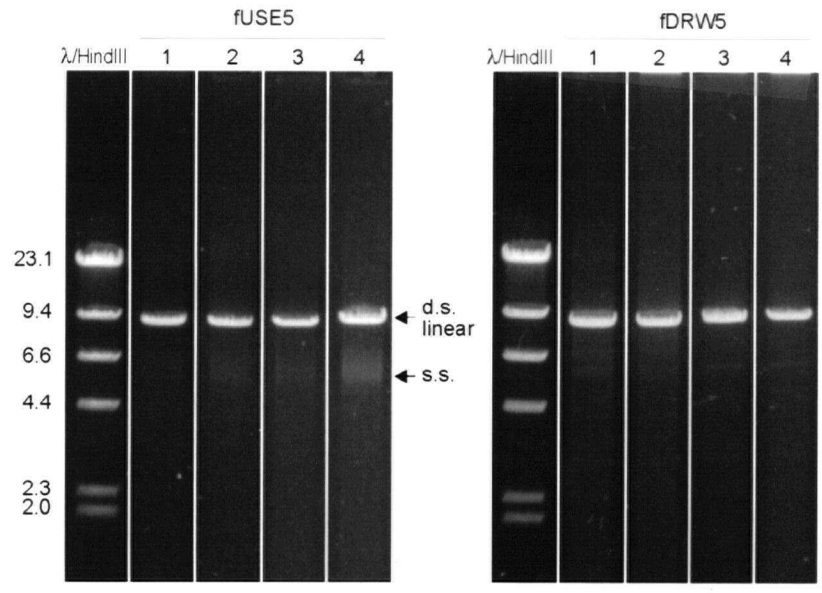


Fig. 7-25. RF DNA of fUSE5 and fDRW5 propagated in *E. coli* LE392. Agarose gels of samples of RF DNA for each of four isolates of fUSE5 and fDRW5, harvested from the same cultures from which virion titers in Figure 7-24 were derived, treated with ATP-dependent DNase to reduce or eliminate s.s. DNA, and digested with *Bam*HI.

(d) Mutations giving rise to pseudorevertants. Sequencing of pseudorevertants derived from these and earlier studies revealed the nature of mutations giving rise to pseudorevertants (Figure 7-26). In frameshift vector fUSE5, single base insertions near the cloning site (Figure 7-26, *a* to *c*) served to restore the gIII reading frame and in two cases destroyed the second of the two *Sfi*I recognition sites, making stuffer excision impossible. In a pseudorevertant of frameshift vector fUSE1, a single base *deletion* similarly restored the gIII reading frame and destroyed the *Pvu*II cloning site (not shown). Pseudorevertants of fDRW5 and fDRW613 consistently arose from a common mechanism, in which single base substitutions converted the

Figure shown on following page

Fig. 7-26. Pseudorevertants of frameshift and amber vectors. In **(a)**, insertion of a single G destroyed the second *Sfi*I site in fUSE5 and converted this frameshift vector to a (wild-type) virion-producing phenotype or "pseudorevertant"; two of the pseudorevertants sequenced from virions produced during the experiment reported in Figure 7-24 shared this sequence, as did one pseudorevertant identified in earlier experiments. In **(b)**, insertion of a single C similarly destroyed the second *Sfi*I site in a single pseudorevertant from the experiment of Figure 7-24. The *Sfi*I site was not lost in the fUSE5 pseudorevertant shown in **(c)**, derived from an earlier experiment.

All pseudorevertants identified for the amber vector fDRW5 involve base substitutions within the amber stop codon. In **(d)**, an A→G substitution converted the amber codon to one encoding tryptophan; the three pseudorevertants identified during the experiment shown in Figure 7-24 were of this kind, as were two of the pseudorevertants identified during an unsuccessful attempt to construct a CSP-peptide library (§7.3.2.d). In **(e)**, a G→T substitution converted the amber codon to a tyrosine codon and destroyed the second *Sfi*I site; four of the pseudorevertants identified during an unsuccessful attempt to construct a CSP-peptide library were of this kind.

Three different substitutions leading to loss of the amber stop codon in fDRW613 are shown in **(f)**; each of these pseudorevertants was identified from the experiment of Figure 7-24.

In **(g)**, **(h)** and **(i)**, large deletions leading to loss of both amber codons in fDRW613C led to virion production for three pseudorevertants identified from the experiment of Figure 7-24.

In **(j)**, a G→T substitution destroyed the amber codon and *Xba*I site (used diagnostically to confirm that the amber site is intact) in fDRW70; two of the pseudorevertants identified from the experiment of Figure 7-24 were of this kind. In a third pseudorevertant, shown in **(k)**, a 33 base pair deletion eliminated the amber codon and cloning site entirely.

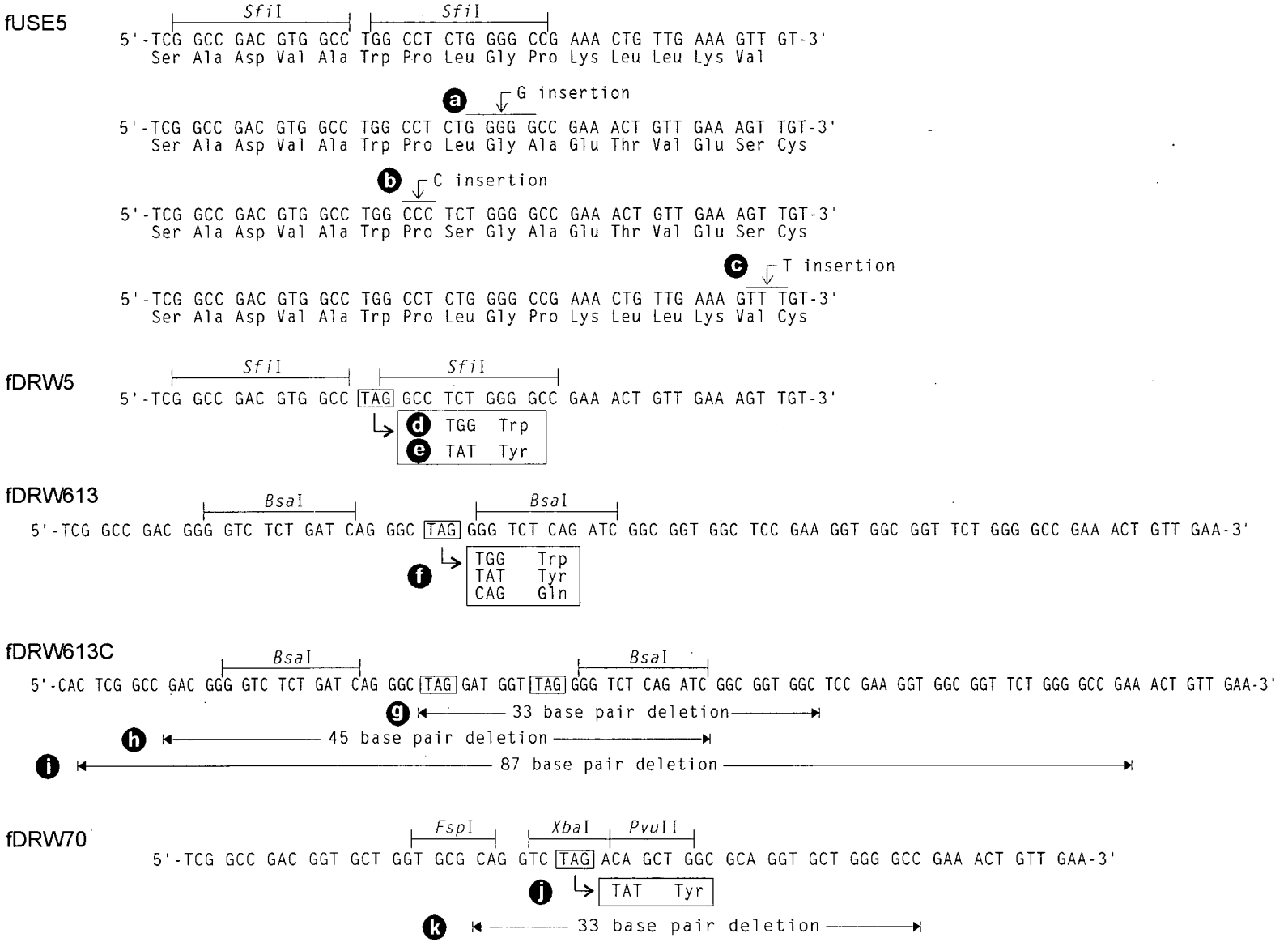

fUSE5
SfiI
SfiI
5'-TCG GCC GAC GTG GCC TGG CCT CTG GGG CCG AAA CTG TTG AAA GTT GT-3'
Ser Ala Asp Val Ala Trp Pro Leu Gly Pro Lys Leu Leu Lys Val
a
G insertion
5'-TCG GCC GAC GTG GCC TGG CCT CTG GGG GCC GAA ACT GTT GAA AGT TGT-3'
Ser Ala Asp Val Ala Trp Pro Leu Gly Ala Glu Thr Val Glu Ser Cys
b
C insertion
5'-TCG GCC GAC GTG GCC TGG CCC TCT GGG GCC GAA ACT GTT GAA AGT TGT-3'
Ser Ala Asp Val Ala Trp Pro Ser Gly Ala Glu Thr Val Glu Ser Cys
c
T insertion
5'-TCG GCC GAC GTG GCC TGG CCT CTG GGG CCG AAA CTG TTG AAA GTT TGT-3'
Ser Ala Asp Val Ala Trp Pro Leu Gly Pro Lys Leu Leu Lys Val Cys
fDRW5
SfiI
SfiI
5'-TCG GCC GAC GTG GCC TAG GCC TCT GGG GCC GAA ACT GTT GAA AGT TGT-3'
d TGG Trp
e TAT Tyr
fDRW613
BsaI
BsaI
5'-TCG GCC GAC GGG GTC TCT GAT CAG GGC TAG GGG TCT CAG ATC GGC GGT GGC TCC GAA GGT GGC GGT TCT GGG GCC GAA ACT GTT GAA-3'
f
TGG Trp
TAT Tyr
CAG Gln
fDRW613C
BsaI
BsaI
5'-CAC TCG GCC GAC GGG GTC TCT GAT CAG GGC TAG GAT GGT TAG GGG TCT CAG ATC GGC GGT GGC TCC GAA GGT GGC GGT TCT GGG GCC GAA ACT GTT GAA-3'
g
33 base pair deletion
h
45 base pair deletion
i
87 base pair deletion
fDRW70
FspI
XbaI
PvuII
5'-TCG GCC GAC GGT GCT GGT GCG CAG GTC TAG ACA GCT GGC GCA GGT GCT GGG GCC GAA ACT GTT GAA-3'
j
TAT Tyr
k
33 base pair deletion

Fig. 7-26.

amber codon to one encoding Trp, Tyr or Gln (Figure 7-26, *d* to *f*). The same mechanism occurred in fDRW70 (see *j*). A 33 bp deletion occurred in one fDRW70 pseudorevertant (see *k*). Deletions encompassing both amber stop codons yielded pseudorevertants of fDRW613C (see *g*, *h* and *i*).

7.3.9. *B. pertussis fhaB* DNase I fragment libraries constructed with fDRW70

(a) Library construction. The earlier-mentioned strategy (§7.3.7) of creating multiple, maximally diverse gene fragment libraries was implemented using fDRW70 and size-fractioned DNase I-generated fragments of the *B. pertussis fhaB* 10 kbp *Eco*RI restriction fragment, as summarized in Table 7-V.

Table 7-V. Construction of *B. pertussis fhaB* DNase I fragment (FHA-70) libraries.

Library[a]	Insert size (base pairs)	Electroporation time constant[b]	Log(transformants)[c]	Log(virions recovered)[d]
70-X	no insert	8.4, 8.6	5.2, 5.1	9.5
70-A	30-75	8.3, 8.7	5.5, 5.0	9.6
70-B	75-150	7.9, 8.6	4.9, 4.9	9.5
70-C	150-300	8.5, 8.7	4.7, 5.0	9.1
70-D	300-600	8.3, 8.7	4.7, 5.1	8.7

[a]Each library was constructed with 30 fmol fDRW70 RF DNA and 90 fmol of the indicated sizes of *B. pertussis fhaB* DNase I fragments. Two ligations and electroporations were performed for each library. Thus the values shown for 'Electroporation time constant' and 'Log(transformants)' are values for these duplicates, while the 'Log(virions recovered)' represents the total yield for each library. [b]Time constants (Miller 1994) provide an index of sample-to-sample uniformity of electroporations. [c]Each estimate of the number of transformants recovered (colony-forming units) is based on a single 10 µL sample spread on a standard petri plate. [d]Total virion yield (as plaque-forming units) was estimated from a full-plate plaque assay; each estimate derives from a single 100 µL sample of a 100-fold serial dilution series.

(b) Preliminary evaluation. Results were immediately disappointing, for a control ligation (library 70-X, Table 7-V) of fDRW70 without a foreign insert, expected to yield few virion-producing transformants, yielded titers similar to those of ligations of fDRW70 with inserts (libraries 70-A to 70-D). It seemed possible that these results were misleading. It might be allowed, for example, that the library 70-X titers reflected efficient virion production

by a small number of pseudorevertants, while library 70-A to-D titers reflected relatively inefficient production by recombinants. Accordingly, 48 transformants from each library were screened using a microtiter plate assay designed to crudely but conveniently identify virion-producing clones. The results (Figure 7-27) strongly suggested that there were a greater number of virion-producing transformants in libraries 70-A to 70-D than in control library 70-X, and that these libraries contained recombinants.

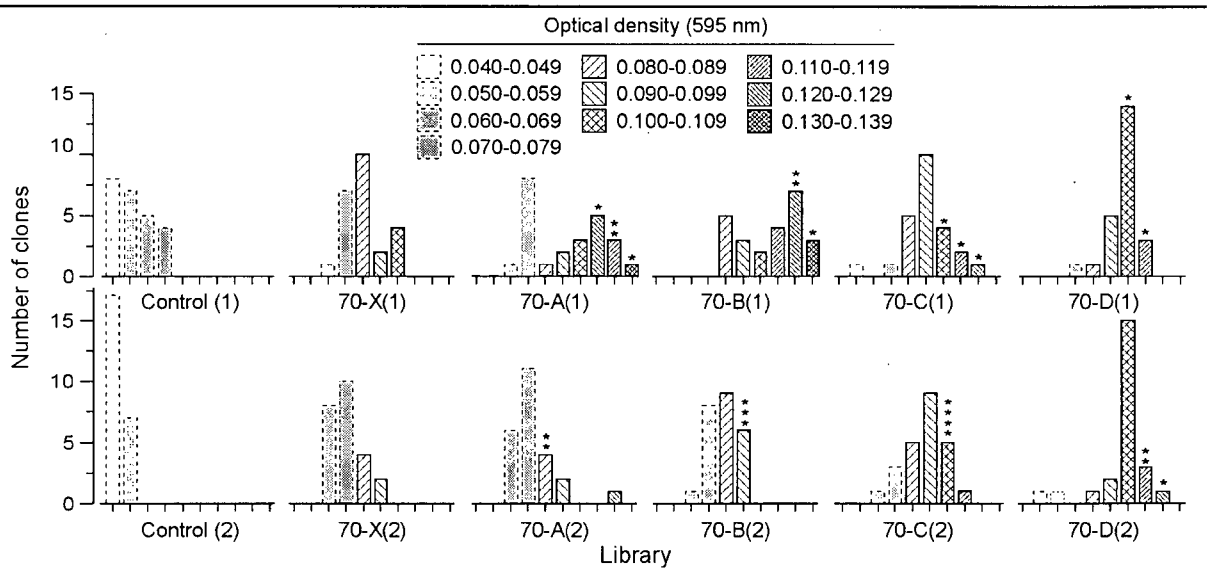


Fig. 7-27. A screen for virion-producing clones in FHA-70 libraries. For each library (70-A, ..., 70-X), two ligations and electroporations [(1) and (2)] were performed (Table 7-V). From each ligation, virions from 24 transformants were used to infect *E. coli* K91-Kan, and infected cells incubated 14 h in 200 μL LB-Tet (37°C) before diluting 20-fold into fresh medium, incubating 25 h (37°C) and assessing culture growth by optical density at 595 nm (OD_{595}). The figure shows the distribution of OD_{595} values. Each of controls (1) and (2) are uninfected cultures. In principle, OD_{595} values greater than those for controls (1) and (2) identify transformants producing virions, while those falling in the same ranges as these controls identify transformants producing no virions. Based on this assay, 24 transformants were selected for sequencing (Figure 7-28). These are identified with asterisks (*); the number of asterisks over each bar represents the number of transformants that were selected.

(c) All 24 clones selected for analysis are of a new class of pseudorevertant. Twenty-four putative virion-producing transformants (identified with "*" in Figure 7-27) were selected from libraries 70-A to -D for sequencing, with the goals of confirming that these were recombinants and of beginning to assess the way a peptide influences its ability to be displayed. Surprisingly, none of the clones were the expected recombinants. Rather, all were of a novel

class of pseudorevertant that arose from a mutation (a single base deletion, Figure 7-28) that occurred not during virion propagation as seen in earlier assays (Figure 7-26) of pseudo-revertant production, but apparently after stuffer fragment removal and subsequent vector self-ligation.

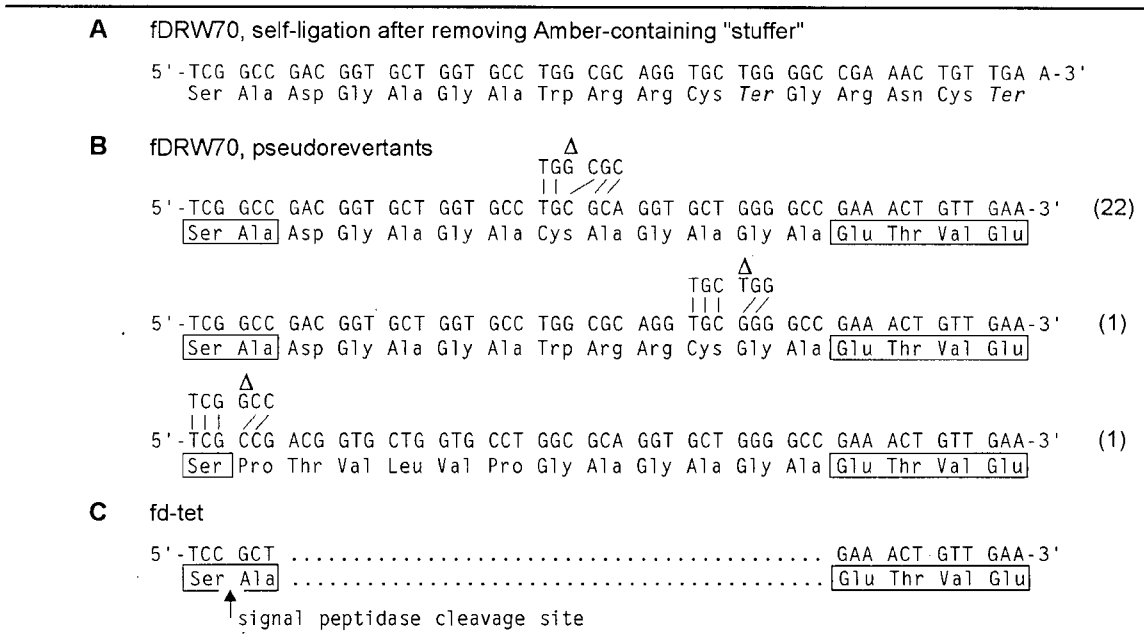


Fig. 7-28. A novel class of fDRW70 pseudorevertant. Of 24 FHA-70 library clones identified as virion-producing and putatively insert-containing via the screening method shown in Figure 7-27, none contained inserts. Rather, all were of a new class of pseudorevertant. **A**. By design (Figure 7-23), vector self-ligation is expected to yield a frameshift in gene III, preventing pIII production and thus virion production. **B**. Surprisingly, single-base deletions (highlighted with Δ) in products of vector self-ligations restored the gIII reading frame in all 24 clones. Numbers in parentheses indicate the number of clones containing the indicated sequence. Boxed sequences identify wild-type amino acid sequences shared with fd-tet (**C**).

7.3.10. *B. pertussis fhaB* DNase I fragment libraries constructed with fDRW8*nn* vectors

In spite of finding only pseudorevertants, it remained possible that the FHA-70 libraries included some recombinants. However, prudence required that another set of libraries be constructed with a new vector before attempting to affinity select target clones from the FHA-70 and new (FHA-80) libraries.

(a) Use of a rare restriction site in fDRW8nn vectors. The amber-containing excisable

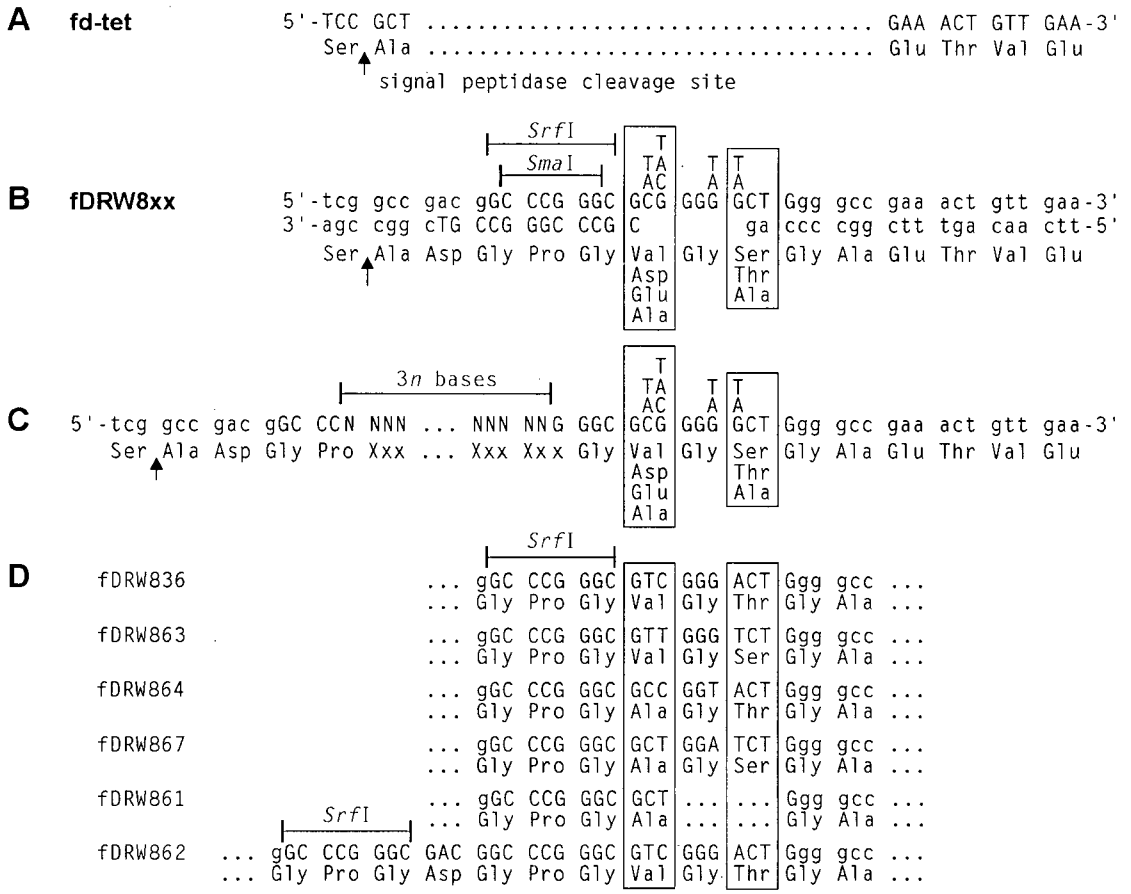


Fig. 7-29. Design and construction of conventional vector series fDRW8*nn*. Figure outlines a series of conventional vectors designed to possess minimal alternations to gIII and pIII, and multiple alternative "linker" peptides between inserted foreign peptides and pIII.

A. Wild-type gIII and pIII sequences of fd-tet, provided for comparison.

B. *Srf*I and *Sma*I restriction sites were included to allow (i) digestion of RF DNA with *Sma*I (*versus* the more costly *Srf*I) to create linear fragments suitable for receiving blunt-end inserts, and (ii) re-digestion of ligation products (*viz.*, vector with foreign DNA inserted into the cloning site) with the "rare-cutter" *Srf*I to eliminate non-recombinants (the products of vector self-ligation). Upper case (parental fDRW5) and lower case (synthetic oligonucleotide) sequences illustrate method of construction this series of vectors. Degeneracy at selected positions was intended to create a set of vectors possessing optimal codons for a variety of linker peptides; boxed sequences identify nucleotide degeneracy intended to encode variations in amino acid sequence. The incompletely complementary oligonucleotide pair was designed with the idea that following transformation of ligation products into *E. coli* LE392, the host cell would fill in the 6-nucleotide gap in the complementary (bottom) strand.

C. The fDRW8*nn* vector series was designed for cloning of a single kind of DNA fragment, *viz.*, those of length $3n$ (where n is an integer) and where the native reading frame of inserted DNA begins at the second base.

D. Only four (vectors fDRW836, fDRW863, fDRW864 and fDRW867) of the 12 expected (as in **B**) amino acid sequences were recovered, in spite of repeated efforts to construct a more diverse set of vectors. Two unexpected sequences (fDRW861 and fDRW862) were also recovered.

stuffer was abandoned in the new fDRW8*nn* vector series. Instead the cloning site was created with a rare, *Srf*I recognition sequence (GCCCGGGC) with the idea that after ligation, non-recombinants could be eliminated by digesting ligation products with *Srf*I before transformation of *E. coli*. As evident from the design (Figure 7-29), a secondary goal was that of optimizing the presentation of peptides with a variable linker between the inserted peptide and the remainder of pIII. Although vectors encoding 12 variant linkers were expected, only four of these and two unexpected sequences were identified (Figure 7-29D) from the small number of transformants screened.

Table 7-VI. Construction of *B. pertussis fhaB* DNase I fragment (FHA-80) libraries.

	Library	Insert size (base pairs)	Electroporation time constant[b]	Log(Trans-formants)[c]	Log(Virions recovered)[d]
Ligation products not re-digested[a] with *Srf*I	80-X	no insert	8.9, 8.8	4.1, 4.2	13.2
	80-Y		9.0, 8.9	3.4, 3.7	12.4
Ligation products re-digested[a] with *Srf*I	80-A	30-75	8.9, 8.9	4.2, 4.2	12.9
	80-B	75-150	8.9, 8.4	4.5, 4.3	12.9
	80-C	150-300	8.9, 8.9	4.5, 4.4	12.8
	80-D	300-600	8.9, 8.9	4.6, 4.6	12.3

[a]Following ligation of the indicated groups of *B. pertussis fhaB* DNase I fragments into fDRW8*nn*, and to reduce the numbers of non-recombinants (*i.e.*, the products of vector self-ligation), ligation products for libraries 80-Y and 80-A, -B, -C and -D, but not 80-X, were re-digested with *Srf*I before electroporation into *E. coli* MC1061.

[b]Time constants (Miller 1994) provide an index of sample-to-sample uniformity of electroporations; two electroporations were performed for each library.

[c]Each estimate (one per electroporation) of the number of transformants recovered (colony-forming units) is based on a single 10 µL sample spread on a standard petri plate.

[d]Total virion yield (physical particles) as determined by UV spectroscopy after pooling the harvests from the two electroporations performed for each library.

(b) Library construction. The FHA-80 libraries (Table 7-VI) were constructed in a same manner as FHA-70 libraries, except that following ligation the ligation products for libraries 80-A to -D and for one (library 80-Y) of two (80-X and 80-Y) control ligations (lacking insert) were digested with *Srf*I before electroporating these products into *E. coli*. The *Srf*I digestion strategy worked moderately well, for almost 10-fold fewer virions were recovered from library

80-Y than from library 80-X (Table 7-VI). Virion yields varied little among libraries 80-A to -D. These, and counts of transformants revealed little concerning the recombinant *versus* non-recombinant fraction, and the libraries were employed without further characterization.

7.3.11. Assessment of FHA-70 and -80 libraries with α-FHA polyclonal antibodies

Among the least demanding applications of a PDL is a search for immunoreactive clones with polyclonal antibodies: finding at least one reactive clone would seem a simple task. Thus, identification of antibody-reactive FHA-peptide clones from FHA-70 and FHA-80 libraries using α-FHA pAbs provided a convenient way of assessing PDL technology, even given apparently "poor" libraries such as the FHA-70 series. Two fundamentally different ways of identifying immunoreactive clones were possible: (i) conventional PDL methods, and (ii) direct screening by plaque lifts or colony blots. Choosing between these was made difficult by the different needs of the two project goals: (i) demonstrating the feasibility of using conventional methods with a genomic PDL (a general model), and (ii) performing a survey of the antigenic makeup of *B. pertussis* FHA (a specific application).

Thus, conventional PDL methods employ multiple rounds of affinity selection with amplification of an enriched library fraction between each round. A concern with gene fragment PDLs in particular is that clonal competition during library amplification might negate the gains from enrichment. For the project's general model, it was important to test whether this occurred. For the specific application, it was important to avoid these problems.

As well, conventional methods (multiple rounds of affinity selection from PDLs) employing a mAb will in principle eventually lead to a single peptide, *viz.*, that with the greatest affinity for the mAb. This suits a general model application, such as identifying a single protein with a role in pathogenesis. With pAbs, the outcome of employing conventional methodology is less predictable. It may, for example, lead to identification of (i) peptides for which there are the greatest number of (possibly low-affinity) cognate pAbs or (ii) peptides that

bind with the greatest avidity to a relatively smaller number of cognate pAbs. Thus, conventional methods are not suited to the needs of a broad survey of the antigenic makeup of FHA. Rather, direct screening by colony blots or plaque lifts would be preferable.

In context of these ideas, a single round of biopanning followed by plaque lift screening provided a workable compromise to meet both sets of objectives in a set of related experiments..

(a) Biopanning with α-FHA polyclonal antibodies. As expected for pAbs, a single round of biopanning FHA-70 and FHA-80 libraries with α-FHA pAbs yielded an immediate apparent enrichment for target clones (Figure 7-30). Thus, for biopannings performed with 3.6 μg and 360 ng of pAbs, up to ~100-fold more virions were recovered from FHA-displaying libraries than from control libraries or from libraries panned with only 36 or 3.6 ng of pAbs.

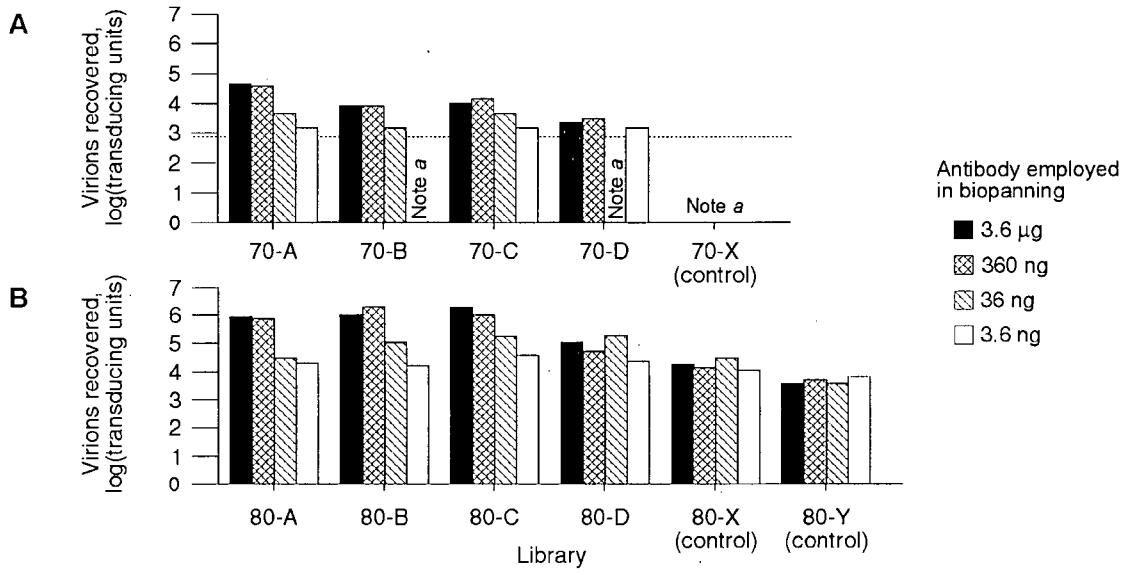


Fig. 7-30. Output from a single round of biopanning FHA-70 and -80 libraries. Each of the indicated libraries (70-A, ..., 80-Y) was biopanned with 3.6 μg, 360 ng, 36 ng and 3.6 ng of an equimolar pool of biotinylated, *E. coli*-absorbed protein-A purified rabbit α-FHA sera FN2/4 and FS1/4. *Ca.* 5×10^8 - 2×10^9 virions were employed in biopanning FHA-70 libraries (**A**); *ca.* 9×10^{10} - 7×10^{11} virions, in biopanning FHA-80 libraries (**B**). Estimates of virions recovered (as transducing units, TU) were derived from single 5 μL samples of 10-fold serial dilutions of biopan output (eluted virions).

[a]No TU were detected for the indicated biopans. The lower limit of detection, 750 TU, is indicated by a dotted horizontal line (·····) in **A**..

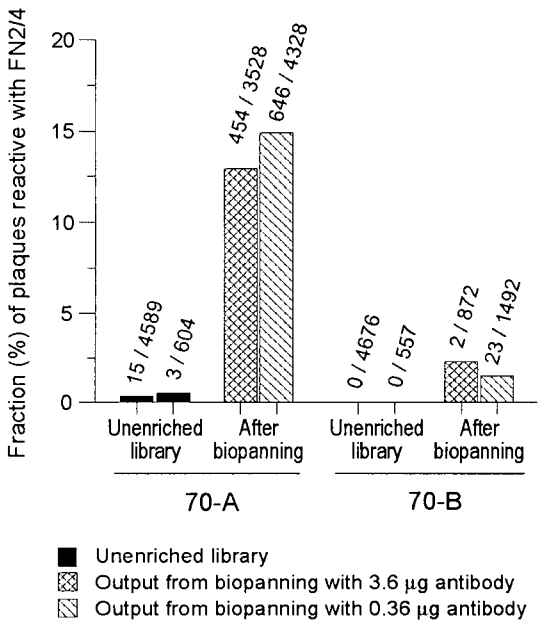


Fig. 7-31. Effect of biopanning on fraction of antibody-reactive clones recovered from FHA-70 libraries: unenriched library *versus* biopanning output. Chart shows the fractions of clones recognized in plaque lifts (with a 1:8,000 dilution of Protein A-purified FN2/4 α-FHA pAbs) of the indicated libraries before ("unenriched library") and after biopanning with 3.6 µg and 360 ng α-FHA pAbs FN2/4 and FS1/4. Each bar represents a single plaque lift. The numbers "*r* / *t*" placed over each bar represent the number of antibody-reactive (*r*) plaques (colored plaques on a nitrocellulose disc) as a fraction of the total (*t*) number of plaques visible on the corresponding lawn; these values are expressed as a percentage on the Y-axis.

(b) Importance of biopanning to identification of antibody-reactive clones. Whether antibody-reactive clones could be readily identified without biopanning was determined by probing plaque lifts of unenriched and biopanned samples of libraries 70-A and -B with α-FHA pAbs. That few (library 70-A) or no (70-B) antibody-reactive clones were identified in the unenriched libraries (Figure 7-31), but that as many as 15% of clones (70-A) were reactive after one round of biopanning suggests that, even for an undemanding application employing pAbs, the ability to affinity-rich for target clones is important and that PDLs offer advantages not found with recombinant expression libraries, such as λgt11, that cannot be affinity-enriched.

(c) Identification of antibody-reactive clones, and effect of library amplification on fraction of antibody-reactive clones. Plaque lifts with α-FHA pAbs were used for several purposes in the experiment summarized in Figure 7-32. First, plaque lifts of biopanning-enriched library fractions (Figure 7-32, solid bars) were used to confirm that as suggested earlier by the data of Figure 7-30, antibody-reactive clones existed in these fractions. Indeed, as many as 70% of virions reacted well with α-FHA pAbs (Figure 7-32, panel B, library 70-C).

Secondly, these plaque lifts were used to identify candidate clones for further

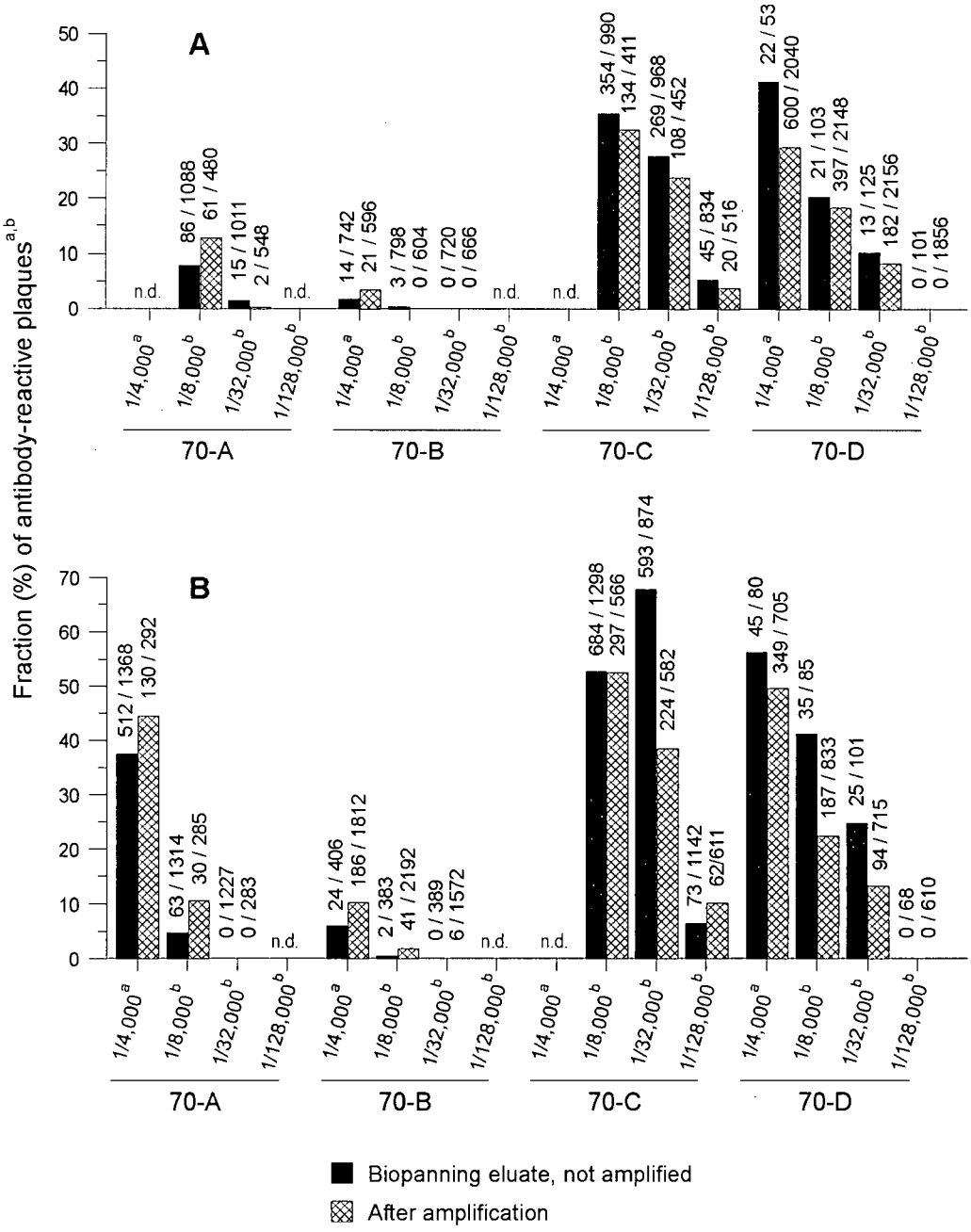


Fig. 7-32. Effect of library amplification on fraction of antibody-reactive clones recovered from FHA-70 libraries. After samples of eluates recovered from biopanning the FHA-70 libraries with 3.6 µg (**A**) and 360 ng (**B**) pooled FN2/4 and FS1/4 α-FHA pAbs were amplified in *E. coli* K91-Kan, plaque lifts were used to determine the fractions of antibody-reactive clones in biopan eluates before and after amplification. The fraction *"r / t"* placed over each bar represent the number of antibody-reactive (*r*) plaques (colored plaques on nitrocellulose) as a fraction of the total (*t*) number of plaques visible on the corresponding lawn; these values are expressed as a percentage on the Y-axis.

[a]Plaque lifts were probed with an equimolar pool of *E. coli*-absorbed Protein A-purified FN2/4 and FS1/4, each employed at a 1 in 8,000 dilution. [b]Plaque lifts were probed with the indicated dilution of *E. coli*-absorbed Protein A-purified FN2/4 alone. n.d., not determined: lawns were thin or of otherwise poor quality and not expected to yield reliable results

characterization. As described in Chapter 6, sequencing of 109 antibody-reactive clones identified from these and other plaque lifts, and subsequent immunocharacterization of 31 unique peptides, formed the basis of a useful antigenic analysis of FHA.

As well, concurrent plaque lifts of biopanning-enriched fractions that had been subsequently amplified in *E. coli* (Figure 7-32, hatched bars) provided a measure of the effects of clonal competition, such as between recombinants and non-recombinants. Curiously, data for libraries displaying small peptides (70-A, 10-25 amino acids; 70-B, 25-50 amino acids) suggested that amplification yielded small gains in the fractions of antibody-reactive clones, while libraries displaying larger peptides (70-C, 50-100 amino acids; 70-D, 100-200 amino acids) showed moderate losses. Since true gains seem unlikely to have occurred, and the plaque lifts were performed concurrently, this is puzzling. That larger-peptide libraries scored greater losses makes sense, for the reasons that a larger peptide would be more likely to possess residues or sequences not tolerated by phage/host biology. Although relative losses in some samples represented more than 40% of antibody-reactive virions (*e.g.*, Figure 7-32, panel B, library 70-C probed with 1/32,000 dilution of pAbs), the data of Figure 7-31 suggests that a subsequent round of affinity-enrichment would readily compensate for these losses.

The effects of amplification did not vary among the different antibodies or antibody dilutions employed in the plaque lifts. Specifically, similar effects were noted for plaque lifts probed with (i) a 1/4,000 dilution of pooled pAbs raised against denatured FHA (FS1/4; see Materials and Methods) and "native" FHA (FN1/4), or (ii) 1/8,000, 1/32,000 and 1/128,000 dilutions of FN2/4. This provides some evidence that the overall composition of high-affinity *versus* low-affinity recombinants may have been unaltered. From this it would follow that clonal competition among recombinants was minimal and that the principal source of amplification effects was competition between non-recombinants and recombinants.

7.4. DISCUSSION

7.4.1. General findings

This project was begun with the goals of illustrating the application potential of phage display and identifying limitations imposed by phage and host biology. The first of these goals was largely successful and is the subject of later chapters. The second goal included evaluating the role of a peptide sequence on its ability to be displayed on the phage surface. This task was delayed by real and perceived difficulties with vectors that had been designed to make the task achievable by virtue of their ability to select against non-recombinants. As discussed earlier in greater detail, this ability was deemed important for two reasons. First, the otherwise difficult task of identifying virion-producing recombinants was to be made simple. Second, it was anticipated that clonal competition between recombinants and non-recombinants would lead to practical difficulties. These concerns have been partly justified.

For example, efforts to identify virion-producing recombinants from the FHA-S library (§7.3.3) constructed with conventional vectors illustrated the futility of these efforts and confirmed the expected value of excluding non-recombinants. Indeed, effective exclusion of non-recombinants from the first and third CSP libraries (§7.3.2.c, §7.3.2.f) and from the FHA-H/H/H libraries (§7.3.6) allowed for studies that would have been more difficult to carry out otherwise. Importantly, although pseudorevertants can be expected in libraries constructed with frameshift or amber vectors, their small relative numbers would be expected to be of little consequence in identifying recombinants in well-constructed libraries (§7.3.2.f). Sub-optimal ligations, however, can be expected to bring pseudorevertants to the fore (§7.3.2.d, §7.3.9.c).

That clonal competition is an issue is suggested by several findings, including (i) the 10^2- to 10^7-fold differences in virion production by apparently stable FHA/H/H/H recombinants (§7.3.6.f), (ii) the diminished ability of vectors fDRW613 and fDRW613C to form plaques, and (iii) the effects of amplification on the antibody-reactive fractions of large-peptide (50 to 200

amino acids) FHA-70 clones (§7.3.11.c). Whether clonal competition has practical consequences will depend on the specific application. Thus, in work with the FHA-70 and FHA-80 libraries, the issue was avoided altogether by screening these libraries via plaque lifts without amplifying the affinity-enriched output from a single round of biopanning. However, this approach is not broadly applicable for most common applications (reviewed in Chapter 1) of phage display involve identification of target clones by multiple rounds of affinity selection with amplification of enriched fractions between each round. Indeed, the principal benefits of phage display would seem to derive from these "conventional" methods.

Several experiments provided examples of the selective advantage enjoyed by mutants of vectors or their recombinants. The identification of pseudorevertants from the second CSP-library (§7.3.2.d) and the FHA-70 libraries (§7.3.9.c) illustrates how pseudorevertants can come to dominate libraries where ligations or other technical aspects are sub-optimal. Propagation of fUSE*n* (§7.3.1) or fDRW*n* (§7.3.5) in F-piliated strains also serves to illustrate the ways in which selection favors initially small subsets of a phage population. The most relevant examples are provided in studies of the FHA-H/H/H libraries (§7.3.6). From these studies, it appeared that cell viability and virion production are correlated, such that cells yielding titers of less than 10^7 mL^{-1} can be expected to "correct" gIII-inserted DNA to enable greater production of virions.

Importantly, successful identification of a diverse collection of recombinants would have allowed systematic analysis of the issue of clonal competition. Regrettably, each effort to identify recombinants within the various restriction fragment and DNase I fragment libraries examined was met with technical difficulties or led to the identification of pseudorevertants. Given the success in eliminating or reducing non-recombinants in CSP-libraries constructed with fDRW5, it came as a surprise to find only pseudorevertants after sequencing 24 virion-producing clones from the FHA-70 libraries (§7.3.9.c), for these had been constructed with a

vector not strikingly different from fDRW5. It seems likely that a renewed attempt to construct FHA-peptide libraries with fDRW70 would yield greater success. The most effective approach to eliminating non-recombinants may have been the simplest. Thus, the FHA-80 libraries were constructed using the simple strategy of re-digesting ligation products with the restriction enzyme used to open up the vector cloning site, and a cursory analysis (Table 7-VI) suggests that this was effective. It is possible that transformants randomly selected from these libraries would have proven to be mostly recombinants.

7.4.2. Vector issues

Clearly, many difficulties with this work derived from vectors. Yields of vector RF DNA were generally low[13], and the quality of both vector and recombinant RF DNA was commonly poor. Low yields derive from use of fd-tet as a parent for fUSE*n* vectors and their fDRW*n* derivatives, for the *ori(-)*-defect in fd-tet reduces the RF DNA copy number about 13-fold (Smith 1988). It is tempting to speculate that problems with quality of RF DNA preparations also derived from use of fd-tet and undefinable effects of the *ori(-)*-defect. As reviewed below, other issues with these vectors appear to reflect unexpected and possibly more definable aspects of phage and host biology.

(a) fUSE5 and fDRW5 RF yields. The 3.2-fold greater yield of fUSE5 *versus* fDRW5 RF DNA propagated under the same conditions (Figure 7-24B) appears to be a phenomenon of phage biology, for increased yields of RF DNA for pIII-defective fd-tet derivatives have been noted elsewhere (Smith 1988). The mechanism is likely a simple consequence of the role of pV in Ff phage biology, whereby pV sequesters (+) strands to prevent their complementation (to form d.s. RF molecules) and enable their incorporation into virions (Model & Russel 1988). Thus the fUSE5 gIII defect could trigger a series of events, in which (i) lack of pIII production

---

[13]To a first approximation, fUSE*n* and fDRW*n* vector yields are roughly 0.1 µg RF DNA per mL of overnight culture in rich medium, roughly 1/50th of yields for pUC or similar contemporary cloning vectors.

leads to (ii) increased numbers of (+) strands sequestered by pV, (iii) a consequent reduced intracellular pool of free pV, (iv) a greater number of unsequestered (+) strands being available for complementation to d.s. RF molecules, (v) increased RF copy number leading to increased gV transcripts, and (vi) increased pV synthesis to establish a new steady state. Accordingly, the (beneficial) increased RF DNA yields of pIII-defective fUSE5 over pIII-producing fDRW5 derives from an accident of design.

(b) fUSE5 pseudorevertants. An accident of design is also suggested by the finding that each of the three insertions giving rise to fUSE5 pseudorevertants (Figure 7-26, *a* to *c*) occurred within a sequence of two or more homonucleotides in the second of two short direct repeats. Thus, inserting G converted the second string of four G's in CTGGCCTCT**GGGG** to a string of five G's, inserting C converted the second CC dinucleotide in GGCCTGG**CC** to CCC, and inserting T converted GTTGAAAG**TTG** to GTTGAAAG**TTT**G. As occurred in fUSE5, frameshift mutations generally occur in runs of two or more of the same base (Hutchinson 1996). Whether the direct repeats played a role in promoting the frameshifts is not known, but the possibility is interesting. Intriguingly, the fUSE5 cloning site (Figure 7-26) contains the rather lengthy imperfect palindrome TCGGCCgacGtGGCCtGGCCtCtggGGCCGA (mismatched bases are in lower case), which in turn contains four direct repeats of GGCC, two of GGCCGA, as well as two partially overlapping TGGCCT repeats. A variety of secondary structures, possibly contributing to replication errors, can be imagined. Perhaps significantly, fUSE1 and fUSE3 lack the kinds of repeats found in fUSE5 and produce ~100-fold fewer pseudorevertants (Figure 7-15); this would suggest that the design of fUSE5 could be changed to improve its stability.

(c) Amber suppression. The differing abilities of host strains to suppress amber codons, as reflected in pseudorevertant numbers (Figure 7-15), is in general agreement with studies of amber suppressor efficiency (Miller & Albertini 1983; other references cited below). Of the three

amber suppressors employed (Figure 7-15), *supD* has generally been found to be the least efficient. Superficially this appears to reflect the low abundance of *supD* suppressor tRNA in cells, for $tRNA_2^{Ser}$, the non-mutant form of *supD* tRNA (Steege 1983) is only ~0.5% of total cellular tRNA under various culture conditions (Dong *et al.* 1996). Although present at higher levels than $tRNA_2^{Ser}$, the non-mutant isoforms of the relatively more efficient *supE* (Inokuchi *et al.* 1979) and *supF* (Goodman *et al.* 1968) are also among the least abundant cellular tRNA species (Dong *et al.* 1996). Thus, increased production of these suppressor tRNAs might be expected to improve amber suppression efficiency. Yet, providing additional copies of *supF* had no obvious effect on pseudorevertant production by amber vectors (Figure 7-16), suggesting that improvements in vector stability cannot derive from this approach.

It has also been shown that amber suppression is context-sensitive; that is, bases flanking a UAG codon can markedly influence suppression efficiency. Up to 35-fold differences in efficiency have been found in studies (Bossi 1983; Miller & Albertini 1983; Edelmann *et al.* 1987; Engelberg-Kulka & Schoulaker-Schwarz 1996) that showed that the greatest single context-related determinant of suppression efficiency was the base following the UAG codon, such that for each of *supD, supE, supF* and *supJ*, UAG-A and UAG-G have been shown to provide the most efficient suppression. Importantly, the amber codons in fDRW*n* vectors are presented in these contexts.

Although it was not shown by specific experiment that amber vector pseudorevertants have a competitive advantage, variations in their relative numbers when propagated in different amber-suppressing host strains (Figure 7-15) suggest that they do. Yet amber vectors had been designed with the expectation (§7.3.2.b) that mutations would be neutral (substitutions) or disadvantageous (frameshifts). Thus, in contrast to frameshift vectors, mutations in amber vectors should confer no selective advantage. This was an over-simplification, for unlike translation of sense codons, amber codon translation involves competition between suppressor

tRNA and Release Factor I. Indeed, many of the context effects noted earlier appear related to this competition (Edelmann *et al.* 1987; Pedersen & Curran 1991; Poole *et al.* 1995; Engelberg-Kulka & Schoulaker-Schwarz 1996). Thus, the selective advantage in a mutation eliminating an amber codon may derive from a reduction in aborted translation products, and this suggests that pseudorevertants of amber vectors cannot be eliminated.

(d) A role for VSP repair? Very short patch (VSP) repair (reviewed in Lieb & Bhagwat 1996) is a DNA repair mechanism that corrects mismatches arising in mutational hot-spots involving spontaneous deamination of the internal cytosine of 5'-CCWGG-3' (W = A or T) sequences methylated by the *dcm* gene product. The mutagenic potential of VSP repair (a tendency to "mis-repair" some sequences to 5'-CCWGG-3') is a topic of some debate and has been the focus of several statistical studies (Bhagwat & McClelland 1992; Merkl *et al.* 1992; Gutierrez *et al.* 1994) and noted in others (*e.g.*, Burge *et al.* 1992; Karlin *et al.* 1992) as a means of accounting for the relative scarcity of CTAG sequences in the genomes of *E. coli* and bacteriophage, including f1 (Blaisdell *et al.* 1996). Here the idea is that a mismatched sequence such as 5'-C<u>C</u>AG-3'/3'-GATC-5' (where <u>C</u> is incorrect and should be T) is mis-repaired to 5'-CCAG-3'/3'-GGTC-5' rather than '5-CTAG-3'/3'-GATC-5'. Curiously, VSP repair thus not only preserves but also creates the sequences (5'-CCWGG-3') that lead to the mutational hot-spots (Berg & Silva 1997).

Intriguingly, two of the three types of base substitutions that led to fDRW*n* pseudo-revertants (Figure 7-26) can arise from VSP-mediated mutation (Gutierrez *et al.* 1994). Thus, C-TAG (in fDRW5, fDRW613 and fDRW70; '-' denotes the boundary between codons) can give rise to C-T<u>G</u>G or C-<u>C</u>AG (Figure 7-26, *d* and *f*), while TAG-G (fDRW5 and fDRW613) can also yield <u>C</u>AG-G (Figure 7-26, *f*). CTAGG (present in fDRW5 and fDRW613) is a preferred VSP substrate when mismatched, while CTAGA (fDRW70) also supports VSP activity (Glasner *et al.* 1995; Lieb & Rehmat 1995). Admittedly, it is highly conjectural to argue

that VSP repair accounts for the mutations that led to pseudorevertants. Indeed, since the A:T→G:C transitions and G:C→T:A transversions required to convert fDRW*n* amber codons to Trp-, Gln- or Tyr-encoding codons (Figure 7-26) are common spontaneous mutations in *E. coli* (reviewed in Hutchinson 1996), VSP repair need not be invoked as a specific mechanism. Nevertheless, the possibility and the mechanism are interesting.

(e) SOS induction. As reviewed and argued in Chapter 1, it seems likely that fUSE*n* and fDRW*n* vectors, as derivatives of the *ori(-)*-defective vector fd-tet, induce the SOS response. This accident of design came about from the observation that pIII-defective mutants were not host-lethal in fd-tet (Smith 1988) and the subsequent exploitation of this by incorporating gIII frameshifts into fUSE*n* frameshift vectors (Parmley & Smith 1988). SOS induction by *ori(-)* mutants is a relatively recent finding (Higashitani *et al.* 1992, 1995), and from phage display literature it appears not to be widely known. Its significance to the present study derives from the following.

First, SOS induction, itself RecA-mediated, leads to further induction of *recA* and other recombination-related genes (Lloyd & Low 1996; Walker 1996), and to increased homologous recombination in some systems (Abbot 1985; Dolzani *et al.* 1991; Mudgett 1991). Thus the unexpectedly high numbers of putative pseudorevertants that resulted from propagating fDRW5 and fDRW613 in *E. coli* harboring the pIII-encoding plasmid pJB61 (Figure 7-16) may be the products of homologous recombination between vector gIII and pJB61 gIII, rather than pseudorevertants arising from base changes in vector gIII.

Second, SOS induction also results in an increased rate of some base substitution mutations, including the previously noted G:C→T:A transversions required to mediate the amber- to Tyr-codon mutations that arise in fDRW*n* pseudorevertants; these are enhanced 2.6-, 7- and 25-fold in various detection systems (Miller & Low 1984; Yatagai *et al.* 1991a; Yatagai *et al.* 1991b), including one that employed M13. Increased rates of (undesirable) base substitutions

would also be expected in recombinants.

SOS induction does not increase the frequency of frame-shift mutation (Miller & Low 1984; Yatagai *et al.* 1991b) and thus does not account for the gIII frameshifts that created fUSE*n* pseudorevertants. Indeed, frameshift mutations are less common than base substitutions in both SOS-induced (Miller & Low 1984; Yatagai *et al.* 1991a; Yatagai *et al.* 1991b) and uninduced cells (reviewed in Hutchinson 1996). That the relative numbers of (Figure 7-15) of fUSE*n* and fDRW*n* pseudorevertants, which arise by frameshift and base substitutions respectively, are consistent with this general finding suggests that frameshift vectors are inherently more stable than amber vectors.

(g) Amber vectors may prove more useful than frame-shift vectors. Notwithstanding that as reviewed above (i) frameshift vectors are inherently more stable than amber vectors, especially in SOS-induced cells, (ii) fUSE5 yields larger quantities of RF DNA than fDRW5, and (iii) pseudorevertants of fDRW5 and other amber vectors cannot be totally eliminated, amber vectors may nevertheless prove to be the more useful. This is because fUSE*n* frameshift vectors are fd-tet derivatives by necessity, while fDRW*n* vectors are fd-tet derivatives because it was convenient to derive them from fUSE5. Thus the fDRW*n* design strategy, but not the frameshift strategy, could be adapted to *ori(-)*-competent vectors. Doing so would be expected to yield benefits such as improved stability, improved yields and possibly better quality of harvested RF DNA, as well as enhanced virion production, including the formation of larger plaques. DNA manipulations would thus be expected to be less troublesome, other tasks such as virion propagation and harvesting more convenient, and some methods such as plaque lifts more sensitive.

7.4.3. Concluding remarks

The long-term goal of these studies, as defined in the Overview of Chapter 1, was to examine the feasibility of constructing genomic PDLs that could be exploited in the study of

pathogenesis. A key component of this goal was determining, in effect, what fraction of peptide-encoding sequences could be successfully displayed on phage, and much of Chapter 1 was concerned with reviewing the ways in which phage and host biology might be expected to reduce this fraction.

While these studies have confirmed that phage/host biology can have surprising, sometimes profound and nearly always troublesome effects, little direct light has been shed on the initial and more relevant question. Obviously, gene fragment PDLs can be constructed and successfully employed, and successful employment of the FHA-peptide libraries (Chapter 6) provides a clear example of this. Yet the unanswered questions regarding the completeness and diversity of such libraries detracts from the utility of the results. Thus, because of possible biases in the library, the antigenic analysis of FHA provided in Chapter 6 was limited to describing "immunoreactive" rather than "immunodominant" regions. While it seems unlikely that any gene fragment library can be perfect, few improvements can be made until we understand the mechanisms that make this so. Accordingly, fundamental studies, such as of the role of a peptide's composition on its ability to be displayed, remain important.

Chapter 8

Summary and General Discussion

My thesis project was begun with the general goal of exploring the feasibility of employing phage display in the study of pathogenesis, and with a specific goal of demonstrating the feasibility of genomic PDLs. The accomplishment of these goals involved working through model systems designed to illustrate the application potential of this new technology and to identify inherent limitations.

The first system, described in Chapters 3 and 4, involved (i) developing a model for folding of the immunodominant tetrapeptide repeats of the *P. falciparum* circumsporozoite protein (CSP), (ii) constructing a small library of variants of these repeats, and (iii) systematically examining the abilities of α-CSP mAbs to recognize these variants. The results of this work extended the experimental work of Dyson *et al.* (1990) and Satterthwait *et al.* (1990), and in particular provided evidence that (i) the tetrapeptide repeats adopt turn-like conformations defined by the sequence Asn-Pro-Asn-Ala and stabilized by hydrogen bonds that derive from a unique peptide geometry imposed by the initial Asn-Pro residues, and (ii) that this conformation is important to the abilities of the α-CSP mAbs to recognize their epitopes. That this study was possible demonstrated two features of phage display that confirmed its utility and technical simplicity and suggested that it could have broad applicability. First, displayed peptides are not conformationally constrained by their carrier pIII molecules. Second, even though the displayed peptides comprise a stoichiometrically tiny fraction of total phage protein, simple methodology such as ELISA can allow discrimination in the recognition of subtly different peptides by mAbs.

The approach adopted in this latter study was unorthodox in the sense that biopanning, commonly employed in phage display applications, was not employed. Biopanning was

explored in Chapter 5. Two α-CSP mAbs were employed to biopan each of two RPLs displaying 6- and 15-residue peptides, with the expectation that a *collection* of peptides containing Asn-Pro or Asx-Pro sequences would be identified. Indeed, a principal attraction of RPLs as illustrated in the pioneering studies of Cwirla *et al.* (1990), Devlin *et al.* (1990) and Scott and Smith (1990) is that biopanning with mAbs can lead to such collections. Such a collection was not found. Rather, for each mAb, only a single sequence similar in sequence to native CSP was identified. As discussed in both Chapters 1 and 5, it is a not uncommon experience to identify only a few or no target peptides when biopanning RPLs with mAbs. As argued in Chapter 5, this may be because -- in some cases at least -- biopanning with a mAb must in principle lead to a single peptide with the highest achievable functional avidity for the mAb employed. Thus, collections of peptide sequences (a desirable outcome) derive from (i) relying on the system to work ineffectively, or (ii) tuning the system by varying a number of system parameters. Considering that the early appeal of RPLs derived from their technical simplicity, this inherent limitation detracts from their utility.

Chapter 6 described successful employment of libraries displaying fragments of *B. pertussis* filamentous hemagglutinin (FHA). Biopanning these libraries with α-FHA pAbs and subsequently characterization of antibody-reactive clones provided an antigenic analysis of FHA that is both confirmed by and greatly extends earlier studies (Delisse-Gathoye *et al.* 1990; A. Siebers & B. Finlay, unpublished). In a general way, this study illustrated that construction of genomic PDLs is technically achievable and that, notwithstanding certain limitations, such libraries can be useful. It seems likely that success in the FHA studies derived in part from employment of a "strategy" described in Chapter 7 (§7.3.7) for constructing gene fragment libraries. This strategy grew out of earlier studies (Chapter 7) that suggested success would require that library redundancy be maximized (by use of multiple libraries constructed with DNase I-generated fragments) and that clonal competition be minimized (by a vector that

reduced the numbers of non-recombinants, and by separate biopanning of libraries displaying different ranges of sizes of peptides). As discussed in Chapters 6 and 7, use of polyclonal antibodies was a relatively undemanding test of the utility of gene fragment and genomic PDLs and I have not yet been able to show that target clones can be selected from these libraries using probes (immobilized heparin and the surfaces of eukaryotic cells) that are more directly relevant to the study of pathogenesis. As suggested in Chapter 6, it is possible that the most straightforward applications of such libraries in pathogenesis will be those that employ mAbs of known biological activity, such as the ability to neutralise bacterial adhesion.

The material of Chapter 7, in reviewing various difficulties encountered with normally routine procedures -- such as vector DNA propagation and characterization of recombinants -- serves to illustrate a theme developed in Chapter 1. More specifically, the experiences of myself (Chapter 7) and others (Chapter 1, §1.5) make it clear that not every application attempted is successful and that phage and host biology (reviewed in Chapter 1, §1.2-1.4) may play a larger role in this than initially believed (*e.g.*, Parmley & Smith 1988; Scott & Smith 1990). It was suggested in Chapter 1 that virtually every host/phage process -- from gene regulation and translation through to membrane-mediated virion assembly and export -- might impose limitations on phage display, and it appears that some of these and other processes (*e.g.*, SOS response induction by fd-tet) account for difficulties described in Chapter 7. Some of these influences of phage and host cell biology are of consequence, while the practical importance of others is questionable or debatable. Thus, a role for VSP repair in influencing fDRW*nn* vector stability (§7.4.2.d) is speculative but interesting. In contrast, SOS induction by *ori(-)*-defective derivatives of fd-tet may have more serious consequences (§7.4.2.e). As suggested in Chapter 7, this is correctable, and adaptation of the fDRW*nn* vector design strategy to an *ori(-)*-competent vector would have a number of benefits.

As suggested in Chapter 1 (§1.6.2), construction of a gene fragment library with Type

3 (multivalent) vectors might be seen as an "extreme test" of phage display. There were two principal reasons for this suggestion. First, multivalent *versus* monovalent display (§1.5.2) is more likely to interfere with phage infectivity, such as by interfering with pIII adsorption to the F-pilus. Second, the peptides of a gene fragment library are expected to be more compositionally heterogenous than those of common PDL applications (§1.6.2). Thus, phage and host biology can be expected to interfere in varying degrees (as shown in Chapter 7) with successful incorporation of different peptides into virions, and the resulting clonal competition during routine procedures can be expected to reduce the odds of successful employment of these libraries.

Clearly, the first issue can be resolved with monovalent display. Regarding the second issue, it has been suggested (*e.g.*, Smith 1993) that monovalent display might obviate problems with display of certain proteins. However, this has not been demonstrated experimentally, and it is not clear how this would help. Similarly, it is not entirely clear whether reduced expression of pIII fusion proteins would similarly facilitate display, as suggested for fd-tet derivatives (Smith and Scott 1993). In simplistic terms, either a pIII fusion protein *can* be synthesized, inserted into the host cell and incorporated into a virion, or it cannot. Conceivably, the best that can be hoped for is (i) that use of monovalent vectors may allow virions to be produced without the troublesome pIII fusions, and (ii) that use of fd-tet derivatives or other vectors that reduce expression of virion genes generally will not be host-lethal when pIII fusion proteins cannot be exported.

Importantly, all of the phage display systems reviewed in Chapter 1 rely on the host cell secretory apparatus, and most documented host-imposed limitations are those imposed by this apparatus. Recently, alternative phage display systems based on lytic phage have been developed, such as those involving bacteriophages λ and T4 (Maruyama *et al.* 1994; Efimov *et al.* 1995; Sternberg & Hoess 1995; Mikawa, 1996 #430; see also Cortese *et al.* 1996). A perceived

advantage of these systems is that incorporation of recombinant proteins into virion coats or tail fibres does not require, as with Ff phage, the host cell's membrane translocation apparatus or other aspects of membrane-mediated assembly and export. Notwithstanding these advantages of lytic phage-based systems, Ff-based phage display remains highly useful. Chapters 1, 4 and 6 provide many examples of successful applications. As suggested in Chapter 7 in context of gene fragment libraries, that while it is unlikely that any system can be made perfect, few improvements can be made until we understand the mechanisms that make this so. Importantly, interpretation of data from imperfect systems must be necessarily qualified. That Chapter 6 describes a study of immunogenic *versus* immunodominant regions of FHA is an example of this. Accordingly, fundamental studies of phage display technology, such as examining the role of a peptide's composition on its ability to be displayed, remain important.

Literature Cited

Abbot P.J. (1985). Stimulation of recombination between homologous sequences on carcinogen-treated plasmid DNA and chromosomal DNA by induction of the SOS response in *Escherichia coli* K12. *Mol.Gen.Genet.* **201**, 129-132.

Abrol S., A. Sampath, K. Arora & V.K. Chaudhary (1994). Construction and characterization of M13 bacteriophages displaying gp120 binding domains of human CD4. *Indian J.Biochem.Biophys.* **31**, 302-309.

Adey N.B., A.H. Mataragnon, J.E. Rider, J.M. Carter & B.K. Kay (1995). Characterization of phage that bind plastic from phage-displayed random peptide libraries. *Gene* **156**, 27-31.

Adzhubei A.A. & M.J.E. Sternberg (1993). Left-handed polyproline II helices commonly occur in globular proteins. *J.Mol.Biol.* **229**, 472-493.

Alfthan K., K. Takkinen, D. Sizmann, H. Soderlund & T.T. Teeri (1995). Properties of a single-chain antibody containing different linker peptides. *Protein Eng.* **8**, 725-731.

Alonso P.L., T. Teuscher & M. Tanner (1994). SPf66: research for development or development of research? *Vaccine* **12**, 99-101.

Anderson S. (1981). Shotgun DNA sequencing using cloned DNase I-generated fragments. *Nucleic Acids Res.* **9**, 3015-3027.

Andersson H. & G. von Heijne (1991). A 30-residue-long "export initiation domain" adjacent to the signal sequence is critical for protein translocation across the inner membrane of *Escherichia coli*. *Proc.Natl.Acad.Sci.USA* **88**, 9751-9754.

Arico B., S. Nuti, V. Scarlato & R. Rappuoli (1993). Adhesion of *Bordetella pertussis* to eukaryotic cells requires a time-dependent export and maturation of filamentous hemagglutinin. *Proc.Natl.Acad.Sci.USA* **90**, 9204-9208.

Armstrong J., R.N. Perham & J.E. Walker (1981). Domain structure of bacteriophage fd adsorption protein. *FEBS Lett.* **135**, 167-172.

Arnot D.E., J.W. Barnwell & M.J. Stewart (1988). Does biased gene conversion influence polymorphism in the circumsporozoite protein-encoding gene of *Plasmodium vivax*? *Proc.Natl.Acad.Sci.USA* **85**, 8102-8106.

Arnot D.E., J.W. Barnwell, J.P. Tam, V. Nussenzweig, R.S. Nussenzweig & V. Enea (1985). Circumsporozoite protein of *Plasmodium vivax*: gene cloning and characterization of the immunodominant epitope. *Science* **230**, 815-818.

Arnot D.E., M.J. Stewart & J.W. Barnwell (1990). Antigenic diversity in Thai *Plasmodium*

vivax circumsporozoite proteins. *MBP* **43**, 147-150.

Ballou W.R., J. Rothbard, R.A. Wirtz, D.M. Gordon, J.S. Williams, R.W. Gore, I. Schneider, M.R. Hollingdale, R.L. Beaudoin, W.L. Maloy, L.H. Miller & W.T. Hockmeyer (1985). Immunogenicity of synthetic peptides from circumsporozoite protein of *Plasmodium falciparum*. *Science* **228**, 996-999.

Barbas C.F. (1993). Recent advances in phage display. *Curr.Opin.Biotechnol.* **4**, 526-530.

Bass S., R. Greene & J.A. Wells (1990). Hormone phage: an enrichment method for variant proteins with altered binding properties. *Proteins* **8**, 309-314.

Beck E. & B. Zink (1981). Nucleotide sequence and genome organisation of filamentous bacteriophages f1 and fd. *Gene* **16**, 35-58.

Berg O.G. & P.J.N. Silva (1997). Codon bias in *Escherichia coli*: the influence of codon context on mutation and selection. *Nucleic Acids Res.* **25**, 1397-1404.

Bernstein F.C., T.F. Koetzle, G.J.B. Williams, E.F.J. Meyer, M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi & M. Tasumi (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J.Mol.Biol.* **112**, 535-542.

Bhagwat A.S. & M. McClelland (1992). DNA mismatch correction by very short patch repair may have altered the abundance of oligonucleotides in the *E. coli* genome. *Nucleic Acids Res.* **20**, 1663-1668.

Bianchi E., A. Folgori, A. Wallace, M. Nicotra, S. Acali, A. Phalipon, G. Barbato, R. Bazzo, R. Cortese, F. Felici & A. Pessi (1995). A conformationally homogenous combinatorial peptide library. *J.Mol.Biol.* **247**, 154-160.

Blaisdell B.E., A.M. Campbell & S. Karlin (1996). Similarities and dissimilarities of phage genomes. *Proc.Natl.Acad.Sci.USA* **93**, 5854-5859.

Bleul C., M. Muller, R. Frank, H. Gausepohl, U. Koldovsky, H.N. Mgaya, J. Luande, M. Pawlita, J. ter Meulen, R. Viscidi & L. Gissmann (1991). Human papillomavirus type 18 E6 and E7 antibodies in human sera: increased anti-E7 prevalence in cervical cancer patients. *J.Clin.Microbiol.* **29**, 1579-1588.

Blond-Elguindi S., S.E. Cwirla, W.J. Dower, R.J. Lipshutz, S.R. Sprang, J.F. Sambrook & M.-J.H. Gething (1993). Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP. *Cell* **75**, 717-728.

Boberg J., T. Salakoski & M. Vihinen (1992). Selection of a representative set of structures from Brookhaven Protein Data Bank. *Proteins* **14**, 265-276.

Boeke J.D., P. Model & N.D. Zinder (1982). Effects of bacteriophage f1 gene III protein on the host cell membrane. *Mol.Gen.Genet.* **186**, 185-192.

Boeke J.D., M. Russel & P. Model (1980). Processing of filamentous phage pre-coat protein. Effect of sequence variations near the signal peptidase cleavage site. *J.Mol.Biol.* **144**, 103-116.

Bossi L. (1983). Context effects: translation of UAG codon by suppressor tRNA is affected by the sequence following UAG in the message. *J.Mol.Biol.* **164**, 73-87.

Böttger V., A. Böttger, E.B. Lane & B.A. Spruce (1995a). Comprehensive epitope analysis of monoclonal anti-proenkephalin antibodies using phage display libraries and synthetic peptides: revelation of antibody fine specificities caused by somatic mutations in the variable region genes. *J.Mol.Biol.* **247**, 932-946.

Böttger V. & E.B. Lane (1994). A monoclonal antibody epitope on keratin 8 identified using a phage peptide library. *J.Mol.Biol.* **235**, 61-67.

Böttger V., P.C. Stasiak, D.L. Harrison, D.M. Mellerick & E.B. Lane (1995b). Epitope mapping of monoclonal antibodies to keratin 19 using keratin fragments, synthetic peptides and phage peptide libraries. *Eur.J.Biochem.* **231**, 475-485.

Boyd D. & J. Beckwith (1990). The role of charged amino acids in the localization of secreted and membrane proteins. *Cell* **62**, 1031-1033.

Bradbury A. & A. Cattaneo (1995). The use of phage display in neurobiology. *Trends Neurosci.* **18**, 243-249.

Brennan M.J., D.L. Burns, B.D. Meade, R.D. Shahin & C.R. Manclark (1992). Recent advances in the development of pertussis vaccines. In *Vaccines: new approaches to immunological problems*, pp. 23-52. Edited by R. Ellis. Shoreham, MA: Butterworth.

Brennan M.J. & R.D. Shahin (1996). Pertussis antigens that abrograte bacterial adherence and elicit immunity. *Am.J.Respir.Crit.Care.Med.* **154**, S145-S149.

Brooks B.R., R.W. Pastor & F.W. Carson (1987). Theoretically determined three-dimensional structure for the repeating tetrapeptide unit of the circumsporozoite coat protein of the malaria parasite *Plasmodium falciparum. Proc.Natl.Acad.Sci.USA* **84**, 4470-4474.

Burge C., A.M. Campbell & S. Karlin (1992). Over- and under-representation of short oligonucleotides in DNA sequences. *Proc.Natl.Acad.Sci.USA* **89**, 1358-1362.

Burkot T.R., Z.W. Da, H.M. Geysen, R.A. Wirtz & A. Saul (1991). Fine specificities of monoclonal antibodies against the *Plasmodium falciparum* circumsporozoite protein: recognition of both repetitive and non-repetitive regions. *Parasite Immunol.* **13**, 161-170.

Burritt J.B., C.W. Bond, K.W. Doss & A.J. Jesaitis (1996). Filamentous phage display of oligopeptide libraries. *Anal.Biochem.* **238**, 1-13.

Campbell J.R. (1989). DNA sequence of the gene encoding a *Plasmodium falciparum* malaria candidate vaccine antigen. *Nucleic Acids Res.* **17**, 5854.

Cao G., A. Kuhn & R.E. Dalbey (1995). The translocation of negatively charged residues across the membrane is driven by the electrochemical potential: evidence for an electrophoresis-like membrane transfer mechanism. *EMBO J.* **14**, 866-875.

Cardin A.D., D.A. Demeter, H.J.R. Weintraub & R.L. Jackson (1991). Molecular design and modelling of protein-heparin interactions. *Methods Enzymol.* **203**, 556-583.

Caspers P., R. Gentz, H. Matile, J.R. Pink & F. Sinigaglia (1989). The circumsporozoite protein gene from NF54, a *Plasmodium falciparum* isolate used in malaria vaccine trials. *MBP* **35**, 185-190.

Chakrabartty A., A.J. Doig & R.L. Baldwin (1993). Helix capping propensities in peptides parallel those in proteins. *Proc.Natl.Acad.Sci.USA* **90**, 11332-11336.

Chang A.C. & S.N. Cohen (1978). Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J.Bacteriol.* **134**, 1141-1156.

Cheadle C., Y. Ivashchenko, V. South, G.H. Searfoss, S. French, R. Howk, G.A. Ricca & M. Jaye (1994). Identification of a Src SH3 domain binding motif by screening a random phage display library. *J.Biol.Chem.* **269**, 24034-24039.

Cherry J.D. (1996). Historical review of pertussis and the classical vaccine. *J.Infect.Dis.* **174(Suppl 3)**, S259-S263.

Choo Y. & A. Klug (1995). Designing DNA-binding proteins on the surface of filamentous phage. *Curr.Opin.Biotechnol.* **6**, 431-436.

Chou P.Y. & G.D. Fasman (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Adv.Enzymol.* **47**, 45-148.

Chou P.Y. & G.D. Fasman (1979). Prediction of β-turns. *Biophys.J.* **26**, 367-384.

Clackson T. & J.A. Wells (1994). *In vitro* selection from protein and peptide libraries. *Trends Biotechnol.* **12**, 173-184.

Colomer-Gould V. & V. Enea (1990). *Plasmodium yoelii nigeriensis* circumsporozoite gene structure and its implication for the evolution of the repeat regions. *MBP* **43**, 51-58.

Conley A.J., J.A.I. Kessler, L.J. Boots, J.-S. Tung, B.A. Arnold, P.M. Keller, A.R. Shaw & E.A. Emini (1994). Neutralization of divergent human immunodeficiency virus type 1 variants and primary isolates by IAM-41-2F5, an anti-gp41 human monoclonal antibody. *Proc.Natl.Acad.Sci.USA* **91**, 3348-3352.

Corey D.R., A.K. Shiau, Q. Yang, B.A. Janowski & C.S. Craik (1993). Trypsin display on the surface of bacteriophage. *Gene* **128**, 129-134.

Cortese R., F. Felici, G. Galfre, A. Luzzago, P. Monaci & A. Nicosia (1994). Epitope

discovery using peptide libraries displayed on phage. *Trends Biotechnol.* **12**, 262-267.

Cortese R., P. Monaci, A. Luzzago, C. Santini, F. Bartoli, I. Cortese, P. Fortugno, G. Galfre, A. Nicosia & F. Felici (1996). Selection of biologically active peptides by phage display of random peptide libraries. *Curr.Opin.Biotechnol.* 7, 616-621.

Cortese R., P. Monaci, A. Nicosia, A. Luzzago, F. Felici, G. Galfre, A. Pessi, A. Tramontano & M. Sollazzo (1995). Identification of biologically active peptides using random libraries displayed on phage. *Curr.Opin.Biotechnol.* **6**, 73-80.

Cox F.E.G. (1991). Malaria vaccines - progress and problems. *Trends Biotechnol.* **9**, 389-394.

Crameri R., R. Jaussi, G. Menz & K. Blaser (1994). Display of expression products of cDNA libraries on phage surfaces. A versatile screening system for selective isolation of genes by specific gene-product/ligand interaction. *Eur.J.Biochem.* **226**, 53-58.

Crameri R. & M. Suter (1993). Display of biologically active proteins on the surface of filamentous phages: a cDNA cloning system for selection of functional gene products linked to the genetic information responsible for their production. *Gene* **137**, 69-75.

Crissman J.W. & G.P. Smith (1984). Gene-III protein of filamentous phages: evidence for a carboxyl-terminal domain with a role in morphogenesis. *Virology* **132**, 445-455.

Cwirla S.E., E.A. Peters, R.W. Barrett & W.J. Dower (1990). Peptides on phage: a vast library for identifying ligands. *Proc.Natl.Acad.Sci.USA* **87**, 6378-6382.

Dame J.B., J.L. Williams, T.F. McCutchan, J.L. Weber, R.A. Wirtz, W.T. Hockmeyer, W.L. Maloy, J.D. Haynes, I. Schneider, D. Roberts, G.S. Sanders, P.E. Reddy, C.L. Diggs & L.H. Miller (1984). Structure of the gene encoding the immunodominant surface antigen on the sporozoite of the human malaria parasite *Plasmodium falciparum. Science* **225**, 593-599.

Dasgupta S. & J.A. Bell (1993). Design of helix ends: amino acid preferences, hydrogen bonding and electrostatic interactions. *Int.J.Pept.Protein Res.* **41**, 499-511.

Davis J.R., J.F. Cortese, D.A. Herrington, J.R. Murphy, D.F. Clyde, A.W. Thomas, S. Baqar, M.A. Cochran, J. Thanassi, M.M. Levine & C.S. Hackett (1992). *Plasmodium falciparum*: in vitro characterization and human infectivity of a clone line. *Exp.Parasitol.* **74**, 159-168.

Decker M.D. & K.M. Edwards (1996). The multicenter acellular pertussis trial: an overview. *J.Infect.Dis.* **174**, S270-S275.

DeGraaf M.E., R.M. Miceli, J.E. Mott & H.D. Fischer (1993). Biochemical diversity in a phage display library of random decapeptides. *Gene* **128**, 13-17.

Del Giudice G., A. Douglas, J.P. Verhave, R.A. Wirtz & F. Zavala (1989). Comparative analysis of ELISAs employing repetitive peptides to detect antibodies to *Plasmodium*

falciparum sporozoites. *Bull.World Health Organ.* **67**, 515-523.

Del Giudice G., H.D. Engers, C. Tougne, S.S. Biro, N. Weiss, A.S. Verdini, A. Pessi, A.A. Degremont, T.A. Freyvogel, P.-H. Lambert & M. Tanner (1987a). Antibodies to the repetitive epitope of *Plasmodium falciparum* circumsporozoite protein in a rural Tanzanian community: a longitudinal study of 132 children. *Am.J.Trop.Med.Hyg.* **36**, 203-212.

Del Giudice G., A.S. Verdini, M. Pinori, A. Pessi, J.-P. Verhave, C. Tougne, B. Ivanoff, P.-H. Lambert & H.D. Engers (1987b). Detection of human antibodies against *Plasmodium falciparum* sporozoites using synthetic peptides. *J.Clin.Microbiol.* **25**, 91-96.

Del Portillo H.A., R.S. Nussenzweig & V. Enea (1987). Circumsporozoite gene of a *Plasmodium falciparum* strain from Thailand. *MBP* **24**, 289-294.

Delisse-Gathoye A.M., C. Locht, F. Jacob, M. Raaschou-Nielsen, I. Heron, J.-L. Ruelle, M. de Wilde & T. Cabezon (1990). Cloning, partial sequence, expression, and antigenic analysis of the filamentous hemagglutinin gene of *Bordetella pertussis*. *Infect.Immun.* **58**, 2895-2905.

Dennis M.S. & R.A. Lazarus (1994a). Kunitz domain inhibitors of tissue factor-factor VIIa. I. potent inhibitors selected from libraries by phage display. *J.Biol.Chem.* **269**, 22129-22136.

Dennis M.S. & R.A. Lazarus (1994b). Kunitz domain inhibitors of tissue factor-factor VIIa. II. potent and specific inhibitors by competitive phage selection. *J.Biol.Chem.* **269**, 22137-22144.

Devlin J.J., L.C. Panganiban & P.E. Devlin (1990). Random peptide libraries: a source of specific protein binding molecules. *Science* **249**, 404-406.

Dolzani L., C. Lagatolla & C. Monti-Bragadin (1991). Prevalence of recombinational versus mutational events in damaged plasmid DNA containing regions of homology with the chromosome. *Mutation Res.* **264**, 127-134.

Domenighini M., D. Relman, C. Capiau, S. Falkow, A. Prugnola, V. Scarlato & R. Rappuoli (1990). Genetic characterization of *Bordetella pertussis* filamentous hemagglutinin: a protein processed from an unusually large precursor. *Mol.Microbiol.* **4**, 787-800.

Dong H., L. Nilsson & C.G. Kurland (1996). Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J.Mol.Biol.* **260**, 649-663.

Duenas M. & C.A. Borrebaeck (1994). Clonal selection and amplification of phage displayed antibodies by linking antigen recognition and phage replication. *Bio/Technology* **12**, 999-1002.

Dunbrack R.L.J. & M. Karplus (1994). Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. *Struct.Biol.* **1**, 334-340.

Dunn I.S. (1996). Phage display of proteins. *Curr.Opin.Biotechnol.* **7**, 547-553.

Dybwad A., B. Bogen, J.B. Natvig, O. Forre & M. Sioud (1995a). Peptide phage libraries can be an efficient tool for identifying antibody ligands for polyclonal antisera. *Clin.Exp.Immunol.* **102**, 438-442.

Dybwad A., O. Forre, J.B. Natvig & M. Sioud (1995b). Structural characterization of peptides that bind synovial fluid antibodies from RA patients: a novel strategy for identification of disease-related epitopes using a random peptide library. *Clin.Immunol.Immunopathol.* **75**, 45-50.

Dyson H.J., A.C. Satterthwait, R.A. Lerner & P.E. Wright (1990). Conformational preferences of synthetic peptides derived from the immunodominant site of the circumsporozoite protein of *Plasmodium falciparum* by ^{1}H NMR. *Biochemistry* **29**, 7828-7837.

Edelmann P., R. Martin & J. Gallant (1987). Nonsense suppression context effects in *Escherichia coli* bacteriophage T4. *Mol.Gen.Genet.* **207**, 517-518.

Edwards K.M. (1993). Acellular pertussis vaccines - a solution to the pertussis problem? *J.Infect.Dis.* **168**, 15-20.

Edwards K.M., B.D. Meade, M.D. Decker, G.F. Reed, M.B. Rennels, M.C. Steinhoff, E.L. Anderson, J.A. Englund, M.E. Pichichero, M.A. Deloria & A. Deforest (1995). Comparison of 13 acellular pertussis vaccines - overview and serological response. *Pediatrics* **96(3 Suppl S)**, 548-557.

Eerola R., P. Saviranta, H. Lilja, K. Pettersson, T. Lovgren & M. Karp (1994). Expression of prostate specific antigen on the surface of a filamentous phage. *Biochem.Biophys.Res.Commun.* **200**, 1346-1352.

Efimov V.P., I.V. Nepluev & V.V. Mesyanzhinov (1995). Bacteriophage T4 as a surface display vector. *Virus Genes* **10**, 173-177.

Egan J.E., J.L. Weber, W.R. Ballou, M.R. Hollingdale, W.R. Majarian, D.M. Gordon, W.L. Maloy, S.L. Hoffman, R.A. Wirtz, I. Schneider, G.R. Woollett, J.F. Young & W.T. Hockmeyer (1987). Efficacy of murine malaria sporozoite vaccines: implications for human vaccine development. *Science* **236**, 453-456.

Eichinger D.J., D.E. Arnot, J.P. Tam, V. Nussenzweig & V. Enea (1986). Circumsporozoite protein of *Plasmodium berghei*: gene cloning and identification of the immunodominant epitopes. *Mol.Cell.Biol.* **6**, 3965-3972.

Endemann H. & P. Model (1995). Location of filamentous phage minor coat proteins in phage and in infected cells. *J.Mol.Biol.* **250**, 496-506.

Engelberg-Kulka H. & R. Schoulaker-Schwarz (1996). Suppression of termination codons. In *Escherichia coli and Salmonella: cellular and molecular biology*, pp. 909-921. Edited by F. C. Neidhardt. Washington DC: ASM Press.

Ewanowich C.A., A.R. Melton, A.A. Weiss, R.K. Sherburne & M.S. Peppler (1989). Invasion of HeLa 229 cells by virulent *Bordetella pertussis*. *Infect.Immun.* **57**, 2698-2704.

Fasman G.D., K. Park & D.H. Schlesinger (1990). Conformational analysis of the immunodominant epitopes of the circumsporozoite protein of *Plasmodium falciparum* and *knowlesi*. *Biopolymers* **29**, 123-130.

Felici F., L. Castagnoli, A. Musacchio, R. Jappelli & G. Cesareni (1991). Selection of antibody ligands from a large library of oligopeptides expressed on a multivalent exposition vector. *J.Mol.Biol.* **222**, 301-310.

Felici F., A. Luzzago, A. Folgori & R. Cortese (1993). Mimicking of discontinuous epitopes by phage-displayed peptides, II. Selection of clones recognized by a protective monoclonal antibody against the *Bordetella pertussis* toxin from phage peptide libraries. *Gene* **128**, 21-27.

Feng J.N., M. Russel & P. Model (1997). A permeabilized cell system that assembles filamentous bacteriophage. *Proc.Natl.Acad.Sci.USA* **94**, 4068-4073.

Fernandez R.C. & A.A. Weiss (1994). Cloning and sequencing of a *Bordetella pertussis* serum resistance locus. *Infect.Immun.* **62**, 4727-4738.

Forood B., E.J. Feliciano & K.P. Nambiar (1993). Stabilization of α-helical structures in short peptides via end capping. *Proc.Natl.Acad.Sci.USA* **90**, 838-842.

Fulford W. & P. Model (1988a). Bacteriophage f1 DNA replication genes. II. The roles of gene V protein and gene II protein in complementary strand synthesis. *J.Mol.Biol.* **203**, 39-48.

Fulford W. & P. Model (1988b). Regulation of bacteriophage f1 DNA replication. I. new functions for genes II and X. *J.Mol.Biol.* **203**, 49-62.

Galinski M.R., D.E. Arnot, A.H. Cochrane, J.W. Barnwell, R.S. Nussenzweig & V. Enea (1987). The circumsporozoite gene of the *Plasmodium cynomolgi* complex. *Cell* **48**, 311-319.

Geysen H.M., S.J. Rodda, T.J. Mason, G. Tribbick & P.G. Schoofs (1987). Strategies for epitope analysis using peptide synthesis. *J.Immunol.Methods* **102**, 259-274.

Gibson K.D. & H.A. Scheraga (1986). Predicted conformations for the immunodominant region of the circumsporozoite protein of the human malaria parasite *Plasmodium falciparum*. *Proc.Natl.Acad.Sci.USA* **83**, 5649-5653.

Glasner W., R. Merkl, V. Schellenberger & H.-J. Fritz (1995). Substrate preferences of Vsr DNA mismatch endonuclease and their consequences for the evolution of the *Escherichia coli* K-12 genome. *J.Mol.Biol.* **245**, 1-7.

Goldman I.F., S.H. Qari, P.G. Millet, W.E. Collins & A.A. Lal (1993). Circumsporozite protein gene of *Plasmodium simium*, a *Plasmodium vivax*-like monkey malaria parasite. *MBP* **57**, 177-180.

Goodman H.M., J. Abelson, A. Landy, S. Brenner & J.D. Smith (1968). Amber suppression: a nucleotide change in the anticodon of a tyrosine transfer RNA. *Nature* **217**, 1019-1024.

Gordon D.M., T.M. Cosgriff, I. Schneider, G.F. Wasserman, W.R. Majarian, M.R. Hollingdale & J.D. Chulay (1990). Safety and immunogenicity of a *Plasmodium vivax* sporozoite vaccine. *Am.J.Trop.Med.Hyg.* **42**, 527-531.

Gram H., U. Stritmatter, M. Lorenz, D. Gluck & G. Zenke (1993). Phage display as a rapid gene expression system: production of bioactive cytokine-phage and generation of neutralizing monoclonal antibodies. *J.Immunol.Methods* **161**, 169-176.

Gramatikoff K., O. Georgiev & W. Schaffner (1994). Direct interaction rescue, a novel filamentous phage technique to study protein-protein interactions. *Nucleic Acids Res.* **22**, 5761-5762.

Gramatikoff K., W. Schaffner & O. Georgiev (1995). The leucine zipper of c-Jun binds to ribosomal protein L18a: a role in Jun protein regulation. *Biol.Chem.Hoppe Seyler* **376**, 321-325.

Greenwood J., A.E. Willis & R.N. Perham (1991). Multiple display of foregin peptides on a filamentous bacteriophage. Peptides from *Plasmodium falciparum* sporozoite protein as antigens. *J.Mol.Biol.* **220**, 821-827.

Griess G.A., E.T. Moreno, R. Herrmann & P. Serwer (1990). The sieving of rod-shaped viruses during agarose gel electrophoresis. I. Comparision with the sieving of spheres. *Biopolymers* **29**, 1277-1287.

Grihalde N.D., Y.-C.J. Chen, A. Golden, E. Gubbins & W. Mandecki (1995). Epitope mapping of anti-HIV and anti-HCV monoclonal antibodies and characterization of epitope mimics using a filamentous phage peptide library. *Gene* **166**, 187-195.

Gutierrez G., J. Casadesus, J.L. Oliver & A. Marin (1994). Compositional Heterogeneity of the *Escherichia coli* genome: a role for VSP repair? *J.Mol.Evol.* **39**, 340-346.

Hammer J., B. Takacs & F. Sinigaglia (1992). Identification of a motif for HLA-DR1 binding peptides using M13 display libraries. *J.Exp.Med.* **176**, 1007-1013.

Hannah J.H., F.D. Menozzi, G. Renauld, C. Locht & M.J. Brennan (1994). Sulfated glycoconjugate receptors for the *Bordetella pertussis* adhesin filamentous hemagglutinin (FHA) and mapping of the heparin-binding domain on FHA. *Infect.Immun.* **62**, 5010-5019.

Harlow E. & D. Lane (1988). *Antibodies: a laboratory manual.* Cold Spring Harbor, USA: Cold Spring Harbor Laboratory.

Harper E.T. & G.D. Rose (1993). Helix stop signals in proteins and peptides: the capping box. *Biochemistry.* **32**, 7605-7609.

Hartl F.U. (1996). Molecular chaperones in cellular protein folding. *Nature* **381**, 571-580.

Healy J.M., O. Murayama, T. Maeda, K. Yoshino, K. Sekiguchi & M. Kikuchi (1995). Peptide ligands for integrin $\alpha_v\beta_3$ selected from random phage display libraries. *Biochemistry* **34**, 3948-3955.

Hemmingsen J.M., K.M. Gernert, J.S. Richardson & D.C. Richardson (1994). The tyrosine corner: a feature of most Greek key β-barrel proteins. *Protein Sci.* **3**, 1927-1937.

Higashitani A., N. Higashitani & K. Horiuchi (1997). Minus-strand origin of filamentous phage versus transcriptional promoters in recognition of RNA polymerase. *Proc.Natl.Acad.Sci.USA* **94**, 2909-2914.

Higashitani N., A. Higashitani & K. Horiuchi (1995). SOS induction in *Escherichia coli* by single-stranded DNA of mutant filamentous phage: monitoring by cleavage of LexA repressor. *J.Bacteriol.* **177**, 3610-3612.

Higashitani N., A. Higashitani, M. Roth & K. Horiuchi (1992). SOS induction in *Escherichia coli* by infection with mutant filamentous phage that are defective in initiation of complementary-strand DNA synthesis. *J.Bacteriol.* **174**, 1612-1618.

Hutchinson F. (1996). Mutagenesis. In *Escherichia coli and Salmonella: cellular and molecular biology*, pp. 2218-2235. Edited by F. C. Neidhardt. Washington DC: ASM Press.

Iannolo G., O. Minenkova, R. Petruzzelli & G. Cesareni (1995). Modifying filamentous phage capsid: limits in the size of the major capsid protein. *J.Mol.Biol.* **248**, 835-844.

Inokuchi H., F. Yamao, H. Sakano & H. Ozeki (1979). Identification of transfer RNA suppressors in *Escherichia coli.* I. Amber suppressor *su*$^+$2, an anticodon mutant of $tRNA_2^{Gln}$. *J.Mol.Biol.* **132**, 649-662.

Ippolito J.A., R.S. Alexander & D.W. Christianson (1990). Hydrogen bond stereochemistry in protein structure and function. *J.Mol.Biol.* **215**, 457-471.

Ishibashi Y., S. Claus & D.A. Relman (1994). *Bordetella pertussis* filamentous hemagglutinin interacts with a leukocyte signal transduction complex and stimulates bacterial adherence to monocyte CR3 (CD11b/CD18). *J.Exp.Med.* **180**, 1225-1233.

Izard J.W. & D.A. Kendall (1994). Signal peptides: exquisitely designed transport promoters. *Mol.Microbiol.* **13**, 765-773.

Jacob-Dubuisson F., C. Buisine, N. Mielcarek, E. Clement, F.D. Menozzi & C. Locht (1996). Amino-terminal maturation of the *Bordetella pertussis* filamentous hemagglutinin. *Mol.Microbiol.* **19**, 65-78.

Jacobsson K. & L. Frykberg (1995). Cloning of ligand-binding domains of bacterial receptors by phage display. *Biotechniques* **18**, 878-885.

Jeffrey G.A. & W. Saenger (1991). *Hydrogen bonding in biological structures*. Berlin: Springer-Verlag.

Jellis C.L., T.J. Cradick, P. Rennert, P. Salinas, J. Boyd, T. Amirault & G.S. Gray (1993). Definining critical residues in the epitope for a HIV-neutralizing monoclonal antibody using phage display and peptide array technologies. *Gene* **137**, 63-68.

Kabsch W. & C. Sander (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637.

Kaiser C.A., D. Preuss, P. Grisafi & D. Botstein (1987). Many random sequences functionally replace the secretion signal sequence of yeast invertase. *Science* **235**, 312-317.

Kane J.F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr.Opin.Biotechnol.* **6**, 494-500.

Kang A.S., C.F. Barbas, K.D. Janda, S.J. Benkovic & R.A. Lerner (1991). Linkage of recognition and replication functions by assembling combinatorial antibody Fab libraries along phage surfaces. *Proc.Natl.Acad.Sci.USA* **88**, 4363-4366.

Karlin S., C. Burge & A.M. Campbell (1992). Statisical analysis of counts and distributions of restrictions sites in DNA sequences. *Nucleic Acids Res.* **20**, 1363-1370.

Karpen M.E., P.L. de Haseth & K.E. Neet (1992). Differences in the amino acid distributions of 3_{10}-helices and α-helices. *Protein Sci.* **1**, 1333-1342.

Kaslow D.C. (1993). Transmission-blocking immunity against malaria and other vector-borne diseases. *Curr.Opin.Immunol.* **5**, 557-565.

Kay B.K., N.B. Adey, Y.-S. He, J.P. Manfredi, A.H. Mataragnon & D.M. Fowlkes (1993). An M13 phage library displaying random 38-amino-acid peptides as a source of novel sequences with affinity to selected targets. *Gene* **128**, 59-65.

Kim M.H., J.C. Hines & D.S. Ray (1981). Viable deletions in the M13 complementary strand origin. *Proc.Natl.Acad.Sci.USA* **78**, 6784-6788.

Krebber C., S. Spada, D. Desplancq & A. Pluckthun (1995). Co-selection of cognate antibody-antigen pairs by selectively-infective phages. *FEBS Lett.* **377**, 227-231.

Kuhn A. (1995). Major coat proteins of bacteriophage Pf3 and M13 as model systems for Sec-independent protein transport. *FEMS Microbiol.Rev.* **17**, 185-190.

Kushwaha A., P.S. Chowdhury, K. Arora, S. Abrol & V.K. Chaudhary (1994). Construction and characterization of M13 bacteriophages displaying functional IgG-binding domains of Staphylococcal protein A. *Gene* **151**, 45-51.

Ladner R.C. (1995). Constrained peptides as binding entities. *Trends Biotechnol.* **13**, 426-430.

Lah M., A. Goldstraw, J.F. White, O. Dolezal, R. Malby & P.J. Hudson (1994). Phage surface presentation and secretion of antibody fragments. *Hum.Antibodies Hybridomas* **5**, 48-56.

Lal A.A., V.F. De La Cruz, G.H. Campbell, P.M. Procell, W.E. Collins & T.F. McCutchan (1988). Structure of the circumsporozoite gene of *Plasmodium malariae*. *MBP* **30**, 291-294.

Lal A.A. & I.F. Goldman (1991). Circumsporozoite protein gene from *Plasmodium reichenowi*, a chimpanzee malaria parasite evolutionarily related to the human malaria parasite *Plasmodium falciparum*. *J.Biol.Chem.* **266**, 6686-6689.

Lanar D.E. (1990). Sequence of the circumsporozoite gene of *Plasmodium berghei* ANKA clone and NK65 strain. *MBP* **39**, 151-154.

Lane D.P. & C.W. Stephen (1993). Epitope mapping using bacteriophage peptide libraries. *Curr.Opin.Immunol.* **5**, 268-271.

Le Questel J.-Y., D.G. Morris, P.H. Maccallum, R. Poet & E.J. Milner-White (1993). Common ring motifs in proteins involving asparagine or glutamine amide groups hydrogen-bonded to main-chain atoms. *J.Mol.Biol.* **231**, 888-896.

Leininger E., P.G. Probst, M.J. Brennan & J.G. Kenimer (1993). Inhibition of *Bordetella pertussis* filamentous hemagglutinin-mediated cell adherence with monoclonal antibodies. *FEMS Microbiol.Lett.* **80**, 31-38.

Li P., J. Beckwith & H. Inouye (1988). Alteration of the amino terminus of the mature sequence of a periplasmic protein can severely affect protein export in *Escherichia coli*. *Proc.Natl.Acad.Sci.USA* **85**, 7685-7689.

Lieb M. & A.S. Bhagwat (1996). Very short patch repair: reducing the cost of cytosine methylation. *Mol.Microbiol.* **20**, 467-473.

Lieb M. & S. Rehmat (1995). Very short patch repair of T:G mismatches in vivo: importance of context and accessory proteins. *J.Bacteriol.* **177**, 660-666.

Livnah O., E.A. Stura, D.L. Johnson, S.A. Middleton, L.S. Mulcahy, N.C. Wrighton, W.J. Dower, L.K. Jolliffe & I.A. Wilson (1996). Functional mimicry of a protein hormone by a peptide agonist: the EPO receptor complex at 2.8 Å. *Science* **273**, 464-471.

Lloyd R.G. & K.B. Low (1996). Homologous recombination. In *Escherichia coli and Salmonella: cellular and molecular biology*, pp. 2236-2255. Edited by F. C. Neidhardt. Washington DC: ASM Press.

Locht C., P. Bertin, F.D. Menozzi & G. Renauld (1993). The filamentous hemagglutinin, a multifaceted adhesin produced by virulent *Bordetella* spp. *Mol.Microbiol.* **9**, 653-660.

Lockyer M.J., C.S. Davies, A. Suhrbier & R.E. Sinden (1990). Nucleotide sequence of the

Plasmodium berghei circumsporozoite gene from the ANKA clone 2.34L. *Nucleic Acids Res.* **18**, 376.

Lockyer M.J. & R.T. Schwarz (1987). Strain variation in the circumsporozoite protein gene of *Plasmodium falciparum. MBP* **22**, 101-108.

Long C.A. (1993). Immunity to blood stages of malaria. *Curr.Opin.Immunol.* **5**, 548-556.

Luzzago A., F. Felici, A. Tramontano, A. Pessi & R. Cortese (1993). Mimicking of discontinuous epitopes by phage-displayed peptides. I. Epitope mapping of human H ferritin using a phage library of constrained peptides. *Gene* **128**, 51-57.

Lyu P.C., H.X. Zhou, N. Jelveh, D.E. Wemmer & N.R. Kallenbach (1992). Position-dependent stabilizing effects in α-helices: N-terminal capping in synthetic model peptides. *J.Am.Chem.Soc.* **114**, 6560-6562.

MacArthur M.W. & J.M. Thornton (1991). Influence of proline residues on protein conformation. *J.Mol.Biol.* **218**, 397-412.

MacIntyre S. & Henning.U. (1990). The role of the mature part of secretory proteins in translocation across the plasma membrane and in regulation of their synthesis in *Escherichia coli. Biochimie* **72**, 157-167.

Makhov A.M., J.H. Hannah, M.J. Brennan, B. Trus, E. Kocsis, J.F. Conway, P.T. Wingfield, M.N. Simon & A.C. Steven (1994). Filamentous hemagglutinin of *Bordetella pertussis.* A bacterial adhesin formed as a 50-nm monomeric rigid rod based on a 19-residue repeat motif rich in beta strands and turns. *J.Mol.Biol.* **241**, 110-124.

Malik P., T.D. Terry, L.R. Gowda, A. Langara, S.A. Petukhov, M.F. Symmons, L.C. Welsh, D.A. Marvin & R.N. Perham (1996). Role of capsid structure and membrane protein processing in determining the size and copy number of peptides displayed on the major coat protein of filamentous bacteriophage. *J.Mol.Biol.* **260**, 9-21.

Margalit H., N. Fischer & S.A. Ben-Sasson (1993). Comparative analysis of structurally defined heparin binding sequences reveals a distinct spatial distribution of basic residues. *J.Biol.Chem.* **268**, 19228-19231.

Markland W., B.L. Roberts, M.J. Saxena, S.K. Guterman & R.C. Ladner (1991). Design, construction and function of a multicopy display vector using fusions to the major coat protein of bacteriophage M13. *Gene* **109**, 13-19.

Marks J.D., H.R. Hoogenboom, A.D. Griffiths & G. Winter (1992). Molecular evolution of proteins on filamentous phage: mimicking the strategy of the immune system. *J.Biol.Chem.* **267**, 16007-16010.

Martin F., C. Toniatti, A.L. Salvati, G. Ciliberto, R. Cortese & M. Sollazzo (1996). Coupling protein design and *in vitro* selection strategies: improving specificity and affinity of

a designed ß-protein IL-6 antagonist. *J.Mol.Biol.* **255**, 86-97.

Martin F., C. Toniatti, A.L. Salvati, S. Venturini, G. Ciliberto, R. Cortese & M. Sollazzo (1994). The affinity-selection of a minibody polypeptide inhibitor of human interleukin-6. *EMBO J.* **13**, 5303-5309.

Maruyama I.N., H.I. Maruyama & S. Brenner (1994). Lambda foo: a lambda phage vector for the expression of foreign proteins. *Proc.Natl.Acad.Sci.USA* **91**, 8273-8277.

Matthews D.J., L.J. Goodman, C.M. Gorman & J.A. Wells (1994). A survey of furin substrate specificity using substrate phage display. *Protein Sci.* **3**, 1197-1205.

Matthews D.J. & J.A. Wells (1993). Substrate phage: selection of protease substrates by monovalent phage display. *Science* **260**, 1113-1117.

Mayhew M. & F.-U. Hartl (1996). Molecular chaperone proteins. In *Escherichia coli and Salmonella: cellular and molecular biology*, pp. 922-937. Edited by F. C. Neidhardt. Washington DC: ASM Press.

McCafferty J., A.D. Griffiths, G. Winter & D.J. Chiswell (1990). Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**, 552-554.

McCafferty J., R.H. Jackson & D.J. Chiswell (1991). Phage-enzymes: expression and affinity chromatography of functional alkaline phosphatase on the surface of bacteriophage. *Protein Eng.* **4**, 955-961.

McConnell S.J. & R.H. Hoess (1995). Tendamistat as a scaffold for conformationally constrained phage peptide libraries. *J.Mol.Biol.* **250**, 460-470.

McDonald I.K. & J.M. Thornton (1994). Satisfying hydrogen bonding potential in proteins. *J.Mol.Biol.* **238**, 777-792.

McDonald I.K. & J.M. Thornton (1995). The application of hydrogen bonding analysis in X-ray crystallography to help orientate asparagine, glutamine and histidine side chains. *Protein Eng.* **8**, 217-224.

McGregor M.J., A.M. Islam & J.E. Sternberg (1987). Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *J.Mol.Biol.* **198**, 295-310.

McLafferty M.A., R.B. Kent, R.C. Ladner & W. Markland (1993). M13 bacteriophage displaying disulfide-constrained microproteins. *Gene* **128**, 29-36.

Menozzi F.D., C. Gantiez & C. Locht (1991). Interaction of the *Bordetella pertussis* filamentous hemagglutinin with heparin. *FEMS Microbiol.Lett.* **62**, 59-64.

Menozzi F.D., R. Mutombo, G. Renauld, C. Gantiez, J.H. Hannah, E. Leininger, M.J. Brennan & C. Locht (1994). Heparin-inhibitable lectin activity of the filamentous

hemagglutinin adhesin of *Bordetalla pertussis*. *Infect.Immun.* **62**, 769-778.

Merkl R., M. Kroger, P. Rice & H.-J. Fritz (1992). Statistical evaluation and biological interpretation of non-random abundance in the *E. coli* K-12 genome of tetra- and pentanucleotide sequences related to VSP DNA mismatch repair. *Nucleic Acids Res.* **20**.

Miller C.G. (1996). Protein degradation and proteolytic modification. In *Escherichia coli and Salmonella: cellular and molecular biology*, pp. 938-954. Edited by F. C. Neidhardt. Washington DC: ASM Press.

Miller J.H. & A.M. Albertini (1983). Effects of surrounding sequence on the suppression of nonsense codons. *J.Mol.Biol.* **164**, 59-71.

Miller J.H. & K.B. Low (1984). Specificity of mutagenesis resulting from the induction of the SOS system in the absence of mutagenic treatment. *Cell* **37**, 675-682.

Miller J.M. (1994). Bacterial transformation by electroporation. *Methods Enzymol.* **235**, 375-385.

Model P. & M. Russel (1988). Filamentous bacteriophage. In *The bacteriophages*, pp. 375-456. Edited by R. Calendar. New York: Plenum Press.

Motti C., M. Nuzzo, A. Meola, G. Galfre, F. Felici, R. Cortese, A. Nicosia & P. Monaci (1994). Recognition by human sera and immunogenicity of HBsAg mimotopes selected from an M13 phage display library. *Gene* **146**, 191-198.

Mudgett J.S. (1991). Ultraviolet light-induced plasmid-chromosome recombination in *Escherichia coli*: the role of *recB* and *recF*. *Gene* **97**, 131-136.

Murphy C.K. & J. Beckwith (1996). Export of proteins to the cell envelope in *Escherichia coli*. In *Escherichia coli and Salmonella: cellular and molecular biology*, pp. 967-1004. Edited by F. C. Neidhardt. Washington DC: ASM Press.

Nardin E.H. & R.S. Nussenzweig (1993). T cell responses to pre-erythrocytic stages of malaria: role in protection and vaccine development against pre-erythrocytic stages. *Annu.Rev.Immunol.* **11**, 687-727.

Nelson F.N., S.M. Friedman & G.P. Smith (1981). Filamentous phage DNA cloning vectors: a noninfective mutant with a nonpolar deletion in gene III. *Virology* **108**, 338-350.

Nielsen H., J. Engelbrecht, S. Brunak & G. von Heijne (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1-6.

Nord K., J. Nilsson, B. Nilsson, M. Uhlen & P.-A. Nygren (1995). A combinatorial library of an α-helical bacterial receptor domain. *Protein Eng.* **8**, 601-608.

Nussenzweig V. & R.S. Nussenzweig (1989). Rationale for the development of an engineered

sporozoite malaria vaccine. *Adv.Immunol.* **45**, 283-334.

O'Neil K.T. & R.H. Hoess (1995). Phage display: protein engineering by directed evolution. *Curr.Opin.Struct.Biol.* **5**, 443-449.

Parmley S.F. & G.P. Smith (1988). Antibody-selectable filamentous fd phage vectors: affinity purification of target genes. *Gene* **73**, 305-318.

Parton R. (1996). New perspectives on *Bordetella* pathogenicity. *J.Med.Microbiol.* **44**, 233-235.

Pasqualini R. & E. Ruoslahti (1996). Organ targeting *in vivo* using phage display peptide libraries. *Nature* **380**, 364-366.

Patarroyo M.E., R. Amador, P. Clvijo, A. Moreno, F. Guzman, P. Romero, R. Tascon, A. Franco, L.A. Murillo, G. Ponton & G. Trujillo (1988). A synthetic vaccine protects humans against challenge with asexual blood stages of *Plasmodium falciparum* malaria. *Nature* **332**, 158-161.

Patarroyo M.E., P. Romero, M.L. Torres, P. Clavijo, A. Moreno, A. Martinez, R. Rodriguez, F. Guaman & E. Cabezas (1987). Induction of protective immunity against experimental infection with malaria using synthetic peptides. *Nature* **328**, 629-632.

Pattabiraman N., K. Namboodiri, A. Lowrey & B.P. Gaber (1990). NRL_3D: a sequence-structure database derived from the protein data bank (PDB) and searchable within the PIR environment. *Protein Seq. Data Anal.* **3**, 387-405.

Pedersen W.T. & J.F. Curran (1991). Effects of the nucleotide 3' to an amber codon on ribosomal selection rates of suppressor tRNA and Release Factor-1. *J.Mol.Biol.* **219**, 231-241.

Perham R.N., T.D. Terry, A.E. Willis, J. Greenwood, F. Di Marzo Veronese & E. Appella (1995). Engineering a peptide epitope display system on filamentous bacteriophage. *FEMS Microbiol. Rev.* **17**, 25-31.

Pessi A., E. Bianchi, A. Crameri, S. Venturini, A. Tramontano & M. Sollazzo (1993). A designed metal-binding protein with a novel fold. *Nature* **362**, 367-369.

Peters E.A., P.J. Schatz, S.S. Johnson & W.J. Dower (1994). Membrane insertion defects caused by positive charges in the early mature region of protein pIII of filamentous phage fd can be corrected by *prlA* suppressors. *J.Bacteriol.* **176**, 4296-4305.

Petersen G., D. Song, B. Hugle-Dorr, I. Oldenburg & E.K.F. Bautz (1995). Mapping of linear epitopes recognized by monoclonal antibodies with gene-fragment phage display libraries. *Mol.Gen.Genet.* **249**, 425-431.

Pittman M. (1984). Genus *Bordetella.* In *Bergey's manual of systematic bacteriology, Volume 1*, pp. 388-393. Edited by N. R. Krieg. Baltimore MD: Williams and Wilkins.

Poole E.S., C.M. Brown & W.T. Tate (1995). The identify of the base following the stop codon determines the efficiency of *in vivo* translational termination in *Escherichia coli. EMBO J.* **14**, 151-158.

Prasad S.M., Y. Yin, E. Rodzinski, E.I. Tuomanen & H.R. Masure (1993). Identification of a carbohydrate recognition domain in filamentous hemagglutinin from *Bordetella pertussis. Infect.Immun.* **61**, 2780-2785.

Presta L.G. & G.D. Rose (1988). Helix signals in proteins. *Science.* **240**, 1632-1641.

Pugsley A.P. (1993). The complete general secretory pathway in Gram-negative bacteria. *Microbiol.Rev.* **57**, 50-108.

Qiagen (1991). *The Qiagenologist: application protocols.* Chatsworth CA: Qiagen Inc.

Qiagen (1993). *Qiagen plasmid handbook.* Chatsworth CA: Qiagen Inc.

Ramachandran G.N., C. Ramakrishnan & V. Sasisekharan (1963). Stereochemistry of polypeptide chain configurations. *J.Mol.Biol.* **7**, 95-99.

Ramachandran G.N. & V. Sasisekharan (1968). Conformation of polypeptides and proteins. *Adv.Protein Chem.* **23**, 283-437.

Rampf B., P. Bross, T. Vocke & I. Rasched (1991). Release of periplasmic proteins induced in *E. coli* by expression of an N-terminal proximal segment of the phage fd gene 3 protein. *FEBS Lett.* **280**, 27-31.

Rappuoli R. (1994). Pathogenicity mechanisms of *Bordetella. Curr.Top.Microbiol.Immunol.* **192**, 319-336.

Rasched I. & E. Oberer (1986). Ff coliphages: structural and functional relationships. *Microbiol.Rev.* **50**, 401-427.

Relman D., E. Tuomanen, S. Falkow, D.T. Golenbock, K. Saukkonen & S.D. Wright (1990). Recognition of a bacterial adhesin by an integrin: macrohage CR3 ($\alpha_M\beta_2$, CD11b/CD18) binds filamentous hemagglutinin of *Bordetella pertussis. Cell* **61**, 1375-1382.

Relman D.A., M. Domenighini, E. Tuomanen, R. Rappuoli & S. Falkow (1989). Filamentous hemagglutinin of *Bordetella pertussis:* nucleotide sequence and crucial role in adherence. *Proc.Natl.Acad.Sci.USA* **86**, 2637-2641.

Renauld-Mongenie G., J. Cornette, N. Mielcarek, F.D. Menozzi & C. Locht (1996). Distinct roles of the N-terminal and C-terminal precursor domains in the biogenesis of the *Bordetella pertussis* filamentous hemagglutinin. *J.Bacteriol.* **178**, 1053-1060.

Richardson J.S. (1981). The anatomy and taxonomy of protein structure. *Adv.Protein Chem.* **34**, 167-339.

Richardson J.S. & D.C. Richardson (1988). Amino acid preferences for specific locations at the ends of α helices. *Science* **240**, 1648-1652.

Roberts B.L., W. Markland, A.C. Ley, R.B. Kent, D.W. White, S.K. Guterman & R.C. Ladner (1992a). Directed evolution of a protein: selection of potent neutrophil elastase inhibitors displayed on M13 fusion phage. *Proc.Natl.Acad.Sci.USA* **89**, 2429-2433.

Roberts B.L., W. Markland, K. Siranosian, M.J. Saxena, S.K. Guterman & R.C. Ladner (1992b). Protease inhibitor display M13 phage: selection of high-affinity neutrophil elastase inhibitors. *Gene* **121**, 9-15.

Roberts D., K. Guegler & J. Winter (1993). Antibody as a surrogate receptor in the screening of a phage display library. *Gene* **128**, 67-69.

Rozdzinksi E., J. Sandros, M. van der Flier, A. Young, B. Spellerberg, C. Bhattacharyya, J. Straub, G. Musso, S. Putney, R. Starzyk & E. Tuomanen (1995). Inhibition of leukocyte-endothelial cell interactions and inflammation by peptides from a bacterial adhesin which mim coagulation factor X. *J.Clin.Invest.* **95**, 1078-1085.

Russel M. (1991). Filamentous phage assembly. *Mol.Microbiol.* **5**, 1607-1613.

Russel M. (1994). Phage assembly: a paradigm for bacterial virulence factor export? *Science* **265**, 612-614.

Russel M. (1995). Moving through the membrane with filamentous phages. *Trends Microbiol.* **3**, 223-228.

Russel M. & P. Model (1981). A mutation downstream from the signal peptidase cleavage site affects cleavage but not membrane insertion of phage coat protein. *Proc.Natl.Acad.Sci.USA* **78**, 1717-1721.

Saggio I., I. Gloaguen & R. Laufer (1995). Functional phage display of ciliary neurotrophic factor. *Gene* **152**, 35-39.

Saggio I. & R. Laufer (1993). Biotin binders selected from a random peptide library expressed on phage. *Biochem.J.* **293**, 613-616.

Salivar W.O., H. Tzagoloff & D. Pratt (1964). Some physical-chemical and biological properties of the rod-shaped coliphage M13. *Virology* **24**, 359-371.

Sambrook J., E.F. Fritsch & T. Maniatis (1989). *Molecular cloning: a laboratory manual.* Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Sandros J. & E. Tuomanen (1993). Attachment factors of *Bordetella pertussis*: mimicry of eukaryotic cell recognition molecules. *Trends Microbiol.* **1**, 192-196.

Satterthwait A.C., L.-C. Chiang, T. Arrhenius, E. Cabezas, F. Zavala, H.J. Dyson, P.E.

Wright & R.A. Lerner (1990). The conformational restriction of synthetic vaccines for malaria. *Bull.World Health Organ.* **68(Suppl.)**, 17-25.

Saukkonen K., C. Cabellos, M. Burroughs, S. Prasad & E. Tuomanen (1991). Integrin-mediated localization of *Bordetella pertussis* within macrophages: role in pulmonary colonization. *J.Exp.Med.* **173**, 1143-1149.

Sawadogo M. & M.W. Van Dyke (1991). A rapid method for the purification of deprotected oligodeoxynucleotides. *Nucleic Acids Res.* **19**, 674.

Sayle R. & A. Bissell (1992). Rasmol: a program for fast realistic rendering of molecular structures with shadows. In *Proceedings of the 10th Eurographics UK '92 Conference.* Edinburgh, Scotland: University of Edinburgh.

Scarlato V., B. Arico, M. Domenighini & R. Rappuoli (1993). Environmental regulation of virulence factors in *Bordetella* species. *Bioessays* **15**, 99-104.

Scarselli E., G. Esposito & C. Traboni (1993). Display of functional domains of the human high affinity IgE receptor on the M13 phage surface. *FEBS Lett.* **329**, 223-226.

Schellekens G.A., E. Lasonder, M. Feijlbrief, D.G.A.M. Koedijk, J.W. Drijfhout, A.-J. Scheffer, S. Welling-Wester & G.W. Welling (1994). Identification of the core residues of the epitope of a monoclonal antibody raised against glycoprotein D of herpes simplex virus type 1 by screening of a random peptide library. *Eur.J.Immunol.* **24**, 3188-3193.

Schrauber H., F. Eisenhaber & P. Argos (1993). Rotamers: To be or not to be? An analysis of amino acid side-chain conformations in globular proteins. *J.Mol.Biol.* **230**, 592-612.

Schulz G.E. & R.H. Schirmer (1979). *Principles of protein structure.* New York: Springer-Verlag New York Inc.

Scott J.K. (1992). Discovering peptide ligands using epitope libraries. *Trends Biochem.Sci.* **17**, 241-245.

Scott J.K. & L. Craig (1994). Random peptide libraries. *Curr.Opin.Biotechnol.* **5**, 40-48.

Scott J.K. & G.P. Smith (1990). Searching for peptide ligands with an epitope library. *Science* **249**, 386-390.

Seale W.J., R. Srinivasan & G.D. Rose (1994). Sequence determinants of the capping box, a stabilizing motif at the N-termini of alpha helices. *Protein Sci.* **3**, 1741-1745.

Sharma S., P. Svec, G.H. Mitchell & G.N. Godson (1985). Diversity of circumsporozoite antigen genes from two strains of the malarial parasite *Plasmodium knowlesi. Science.* **229**, 779-782.

Sioud M., A. Dybwad, L. Jespersen, S. Suleyman, J.B. Natvig & O. Forre (1994).

Characterization of naturally occurring autoantibodies against tumor necrosis factor-alpha (TNF-α): *in vitro* function and precise epitope mapping by phage epitope library. *Clin.Exp.Immunol.* **98**, 520-525.

Smith G.P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315-1317.

Smith G.P. (1988). Filamentous phage assembly: morphogenetically defective mutants that do not kill the host. *Virology* **167**, 156-165.

Smith G.P. (1992). *Cloning in fUSE vectors*: handbook distributed by G.P. Smith, University of Missouri, U.S.A.

Smith G.P. (1993). Surface display and peptide libraries. *Gene* **128**, 1-2.

Smith G.P., D.A. Schultz & J.E. Ladbury (1993). A ribonuclease S-peptide antagonist discovered with a bacteriophage display library. *Gene* **128**, 37-42.

Smith G.P. & J.K. Scott (1993). Libraries of peptides and proteins displayed on filamentous phage. *Methods Enzymol.* **217**, 228-257.

Smith J.M. & N.H. Smith (1996). Site-specific codon bias in bacteria. *Genetics* **142**, 1037-1043.

Smith P.K., R.I. Krohn, G.T. Hermanson, A.K. Mallia, F.H. Gartner, M.D. Provenzano, E.K. Fujimoto, N.M. Goeke, B.J. Olson & D.C. Klenk (1985). Measurement of protein using bicinchoninic acid. *Anal.Biochem.* **150**, 76-85.

Soderlind E., M. Duenas & C.A.K. Borrebaeck (1995). Chaperonins in phage display of antibody fragments. *Methods Mol.Biol.* **51**, 343-353.

Soumillion P., L. Jespers, M. Bouchet, J. Marchand-Brynaert, P. Sartiaux & J. Fastrez (1994a). Phage display of enzymes and *in vitro* selection for catalytic activity. *Appl.Biochem.Biotechnol.* **47**, 175-190.

Soumillion P., L. Jespers, M. Bouchet, J. Marchand-Brynaert, G. Winter & J. Fastrez (1994b). Selection of ß-lactamase on filamentous bacteriophage by catalytic activity. *J.Mol.Biol.* **237**, 415-422.

Sparks A.B., L.A. Quilliam, J.M. Thorn, C.J. Der & B.K. Kay (1994). Identification and characterization of Src SH3 ligands from phage-displayed random peptide libraries. *J.Biol.Chem.* **269**, 23853-23856.

Steege D.A. (1983). A nucleotide change in the anticodon of an *Escherichia coli* serine transfer RNA results in *supD*$^-$ amber suppression. *Nucleic Acids Res.* **11**, 3823-3832.

Stephen C.W., P. Helminen & D.P. Lane (1995). Characterisation of epitopes on human p53

using phage-displayed peptide libraries: insights into antibody-peptide interactions. *J.Mol.Biol.* **248**, 58-78.

Sternberg N. & R.H. Hoess (1995). Display of peptides and proteins on the surface of bacteriophage lambda. *Proc.Natl.Acad.Sci.USA* **92**, 1609-1613.

Stickle D.F., L.G. Presta, K.A. Dill & G.D. Rose (1992). Hydrogen bonding in globular proteins. *J.Mol.Biol.* **226**, 1143-1159.

Stoute J.A., W.R. Ballou, K. Kolodny, C.D. Deal, R.A. Wirtz & L.E. Lindler (1995). Induction of humoral immune response against *Plasmodium falciparum* sporozoites by immunization with a synthetic peptide mimitope whose sequence was derived from screening a filamentous phage epitope library. *Infect.Immun.* **63**, 934-939.

Summers R.G. & J.R. Knowles (1989). Illicit secretion of a cytoplasmic protein into the periplasm of *Escherichia coli* requires a signal peptide plus a portion of the cognate secreted protein. *J.Biol.Chem.* **264**, 20074-20081.

Swimmer C., S.M. Lehar, J. McCafferty, D.J. Chiswell, W.A. Blattler & B.C. Guild (1992). Phage display of ricin B chain and its single binding domains: system for screening galactose-binding mutants. *Proc.Natl.Acad.Sci.USA* **89**, 3756-3760.

Tuomanen E. (1986). Piracy of adhesins: attachment of superinfecting pathogens to respiratory cilia by secreted adhesins of *Bordetella pertussis*. *Infect.Immun.* **54**, 905-908.

Tuomanen E.I., S.M. Prasad, J.S. George, A.I.M. Hoepelman, P. Ibsen, I. Heron & R.M. Starzyk (1993). Reversible opening of the blood-brain barrier by anti-bacterial antibodies. *Proc.Natl.Acad.Sci.USA* **90**, 7824-7828.

von Heijne G. (1986). Net N-C charge imbalance may be important for signal sequence function in bacteria. *J.Mol.Biol.* **192**, 287-290.

von Heijne G. (1994). Membrane proteins: from sequence to structure. *Annu.Rev.Biophys.Biomol.Struct.* **23**, 167-92.

Walker G.C. (1996). The SOS response of *Escherichia coli*. In *Escherichia coli and Salmonella: cellular and molecular biology*, pp. 1400-1416. Edited by F. C. Neidhardt. Washington DC: ASM Press.

Wall J.G. & A. Pluckthun (1995). Effects of overexpressing folding modulators on the *in vivo* folding of heterologous proteins in *Escherichia coli*. *Curr.Opin.Biotechnol.* **6**, 507-516.

Wang C.-I., Q. Yang & C.S. Craik (1995a). Isolation of a high affinity inhibitor of urokinase-type plasminogen activator by phage display of ecotin. *J.Biol.Chem.* **270**, 12250-12256.

Wang L.-F., D.H. Du Plessis, J.R. White, A.D. Hyatt & B.T. Eaton (1995b). Use of a gene-

targeted phage display random epitope library to map an antigenic determinant on the bluetongue virus outer capsid protein VP5. *J.Immunol.Methods* **178**, 1-12.

Weiss A.A. & E.L. Hewlett (1986). Virulence factors of *Bordetella pertussis*. *Annu.Rev.Microbiol.* **40**, 661-686.

Wells J.A. (1996). Hormone mimicry. *Science* **273**, 449-450.

Wells J.A. & H.B. Lowman (1992). Rapid evolution of peptide and protein binding properties *in vitro*. *Curr.Opin.Biotechnol.* **3**, 355-362.

Widersten M. & B. Mannervik (1995). Glutathione transferases with novel active sites isolated by phage display from a library of random mutants. *J.Mol.Biol.* **250**, 115-122.

Williams K.A., M. Glibowicka, Z. Li, H. Li, A.R. Khan, Y.M.Y. Chen, J. Wang, D.A. Marvin & C.M. Deber (1995). Packing of coat protein amphipathic and transmembrane helices in filamentous bacteriophage M13: role of small residues in protein oligomerization. *J.Mol.Biol.* **252**, 6-14.

Wilmot C.M. & J.M. Thornton (1988). Analysis and prediction of the different types of β-turns in proteins. *J.Mol.Biol.* **203**, 221-232.

Wilson D.R. & B.B. Finlay (1997). The "Asx-Pro turn" as a local structural motif stabilized by alternative patterns of hydrogen bonds, and a consensus-derived model of the sequence Asn-Pro-Asn. *Protein Eng.* **10**, 519-529.

Wilson D.R., R.A. Wirtz & B.B. Finlay (1997). Recognition of phage-expressed peptides containing Asx-Pro sequences by monoclonal antibodies that recognize *Plasmodium falciparum* circumsporozoite protein. *Protein Eng.* **10**, 531-540.

Winter G., A.D. Griffiths, R.E. Hawkins & H.R. Hoogenboom (1994). Making antibodies by phage display technology. *Annu.Rev.Immunol.* **12**, 433-455.

Wirtz R.A., Y. Charoenvit, T.R. Burkot, K.M. Esser, R.L. Beaudoin, W.E. Collins & R.G. Andre (1991). Evaluation of monoclonal antibodies against *Plasmodium vivax* sporozoites for ELISA development. *Med.Vet.Entomol.* **5**, 17-22.

Wirtz R.A., F. Zavala, Y. Charoenvit, G.H. Campbell, T.R. Burkot, I. Schneider, K.M. Esser, R.L. Beaudoin & R.G. Andre (1987). Comparative testing of monoclonal antibodies against *Plasmodium falciparum* sporozoites for ELISA development. *Bull.World Health Organ.* **65**, 39-45.

Wrighton N.C., F.X. Farrell, R. Chang, A.K. Kashyap, F.P. Barbone, L.S. Mulcahy, D.L. Johnson, R.W. Barrett, L.K. Jolliffe & W.J. Dower (1996). Small peptides as potent mimetics of the protein hormone erythropoietin. *Science* **273**, 458-464.

Yamamoto K.R., B.M. Alberts, R. Benzinger, L. Lawhorne & G. Treiber (1970). Rapid

bacteriophage sedimentation in the presence of polyethylene glycol and its application to large-scale virus purification. *Virology* **40**, 734-744.

Yamane K. & S. Mizushima (1988). Introduction of basic amino acids after the signal peptide inhibits protein translocation across the cytoplasmic membrane of *Escherichia coli*. *J.Biol.Chem.* **263**, 19690-19696.

Yao J., H.J. Dyson & P.E. Wright (1994a). Three-dimensional structure of a type VI turn in a linear peptide in water solution. *J.Mol.Biol.* **243**, 754-766.

Yao J., V.A. Feher, B.F. Espejo, M.T. Reymond, P.E. Wright & H.J. Dyson (1994b). Stabilization of a type VI turn in a family of linear peptides in water solution. *J.Mol.Biol.* **243**, 736-753.

Yatagai F., J.A. Halliday & B.W. Glickman (1991a). Specificity of *recA441*-mediated (*tif-1*) mutational events. *Mol.Gen.Genet.* **230**, 75-80.

Yatagai F., M.J. Horsfall & B.W. Glickman (1991b). Specificity of SOS mutagenesis in native M13*lacI* phage. *J.Bacteriol.* **173**, 7996-7999.

Young J.F., W.T. Hockmeyer, M. Gross, W.R. Ballou, R.A. Wirtz, J.H. Trosper, R.L. Beaudoin, M.R. Hollingdale, L.H. Miller, C.L. Diggs & M. Rosenberg (1985). Expression of *Plasmodium falciparum* circumsporozoite proteins in *Escherichia coli* for potential use in a human malaria vaccine. *Science* **228**, 958-962.

Zacher A.N.I., C.A. Stock, J.W.I. Golden & G.P. Smith (1980). A new filamentous phage cloning vector: fd-tet. *Gene* **9**, 127-140.

Zavala F., J.P. Tam, M.R. Hollingdale, A.H. Cochrane, I. Quakyi, R.S. Nussenzweig & V. Nussenzweig (1985). Rationale for development of a synthetic vaccine against *Plasmodium falciparum* malaria. *Science* **228**, 1436-1440.

Zavala F., J.P. Tam & A. Masuda (1986). Synthetic peptides as antigens for the detection of humoral immunity to *Plasmodium falciparum* sporozoites. *J.Immunol.Methods* **93**, 55-61.

Zhong G., G.P. Smith, J. Berry & R.C. Brunham (1994). Conformational mimicry of a chlamydial neutralization epitope on filamentous phage. *J.Biol.Chem.* **269**, 24183-24188.

Zinder N.D., R.C. Valentine, M. Roger & W. Stoeckenius (1963). f1, a rod shaped male-specific bacteriophage that contains DNA. *Virology* **20**, 638-640.

Appendix

A.1. PUBLICATIONS

The studies reported in Chapters 4 and 5 have been published as follows:

Wilson D.R. & B.B. Finlay (1997). The "Asx-Pro turn" as a local structural motif stabilized by alternative patterns of hydrogen bonds, and a consensus-derived model of the sequence Asn-Pro-Asn. *Protein Eng.* **10**, 519-529.

Wilson D.R., R.A. Wirtz & B.B. Finlay (1997). Recognition of phage-expressed peptides containing Asx-Pro sequences by monoclonal antibodies that recognize *Plasmodium falciparum* circumsporozoite protein. *Protein Eng.* **10**, 531-540.

A.2. ABBREVIATIONS

A_{260}, A_{269}, A_{280}, A_{320}, A_{490}, A_{570}	absorbance at 260, 269, 280, 320, 490, 570 nm
AP	alkaline phosphatase (conjugated to antibody)
BCA	bicinchoninic acid, colorimetric assay for protein quantification (Smith *et al.* 1985)
BCIP/NBT	5-bromo-4-chloro-3-indolyl phosphate p-toluidine salt (BCIP) + nitroblue tetrazolium chloride (NBT), substrate for alkaline phosphatase; §2.13.5
bp	base pairs
BSA	bovine serum albumin (Sigma fraction V) or bovuminar standard powder (Intergen)
cfu	colony-forming units
CS, CSP	*Plasmodium falciparum* circumsporozoite protein
dNTPs	equimolar solution of the four deoxynucleotide triphosphates
ds, d.s.	double stranded
DTT	dithiothreitol
ELISA	enzyme-linked immunosorbent assay

FHA	*Bordetella pertussis* filamentous hemagglutinin
FHA-H/H/H library	*B. pertussis fhaB Hha*I/*Hin*P1/*Hpa*II restriction fragment library
FHA-S library	*B. pertussis fhaB Sau*3AI restriction fragment library
FHA-70 library	*B. pertussis fhaB* DNase I fragment library constructed with vector fDRW70
FHA-80 library	*B. pertussis fhaB* DNase I fragment library constructed with fDRW8*nn* vectors.
gI, gII, gIII, ...	Ff gene I, II, III, ...
gIII, gVIII	encode coat proteins pIII and pVIII respectively
kbp	kilobase pairs
LB-Tet	LB culture medium (broth or plates) + 20 μg tetracycline mL^{-1}
lin.	linear fragment of DNA, digested with a restriction endonuclease
mAb	monoclonal antibody
MBN	Mung bean nuclease
MW	molecular weight (marker)
Kan^R	kanamycin resistant(ce)
NEB	New England Biolabs
oc, o.c.	open circular form of DNA, possessing a "nick" in one of the strands
OD_{595}, OD_{600}	optical density at 595 nm, 600 nm
OPD	*o*-phenylenediamine, peroxidase substrate; §2.13.5
pAb	polyclonal antibody
Pb	*Plasmodium berghei* (monoclonal antibody identification)
PBS	phosphate buffered saline

PDB	Brookhaven National Laboratory Protein Data Bank
PDL	phage display library
PEG	polyethylene glycol
Pf	*Plasmodium falciparum* (monoclonal antibody identification)
pfu	plaque-forming units
pI, pII, pIII, ...	Ff protein encoded by gene I, II, III, ...
Pv	*Plasmodium vivax* (monoclonal antibody identification)
RF	bacteriophage replicative form (double stranded) DNA
RPL	random peptide library
rpm	shaker-incubator, revolutions per minute
sc, s.c.	super-coiled form of DNA
ss, s.s.	single-stranded
TBS	Tris-buffered saline
TE	10 mM Tris·Cl, 1 mM EDTA, pH 7.4 or 8.0
Tet^R	tetracycline resistant(ce)
TU	transducing unit
U	units of DNA-modifying enzyme activity
UV	ultraviolet (spectroscopy)
v/v	volume/volume (%)
w/v	weight/volume (%)