Characterization of the *S. cerevisiae* Rap1 DNA binding domain and biochemical analysis of its interactions with telomeric and silencer DNA

by

Geoffrey Stone

B. Sc., Dalhousie University, 1989

M. Sc., University of British Columbia, 1992

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Doctor of Philosophy in

THE FACULTY OF GRADUATE STUDIES

Department of Biochemistry and Molecular Biology

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 2000

© Geoffrey Stone, 2000
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 Biocchemistry
The University of British Columbia
Vancouver, Canada

Date Aug 24 2000
ABSTRACT

The *Saccharomyces cerevisiae* gene product Rap1 is a multifunctional protein that binds DNA via a 235 amino acid domain. Rap1 binds *in vivo* to telomeres and silent mating type genes, where the protein anchors a heterochromatin complex. Two Helix-Turn-Helix (HTH) motifs within the DNA-binding domain (DBD) have been elucidated by X-ray crystallography (Konig et al., 1996).

In work performed before a crystal structure of DBD was available, weak identity between portions of the DBD and the c-Myb DNA binding motif was used to align the two sequences. Conservation of key hydrophobic residues comprising the c-Myb HTH “core” was used to confirm functional identity.

The second HTH motif (HTH2), located C-terminal to HTH1, was also studied. Analysis of the DBD revealed a highly conserved stretch of amino acids with functional identity to the HTH1 recognition helix. This HTH2 recognition helix was then modeled from various data.

After the DBD crystal structure determination (Konig et al., 1996), a second set of experiments was performed to study the affinity of Rap1 for telomere and silencer sites. Dissociation constant (Kd) data showed high affinity (1X10^{-11} M) of Rap1 for telomere and silencer HML-E binding sites *in vitro*. Competition assays to generate apparent IC_{50} values provided qualitative data on the relative affinity of Rap1 for telomere, HMR-E silencer and HML-E silencer sites. Rap1 showed slightly reduced affinity for HML-E sites. *In vivo* studies of Rap1 binding were performed using the CAT mammalian assay system as a reporter. These studies showed that Rap1 binds with higher affinity to telomeric than to silencer sites.

A variety of RAP1 site-directed mutants were also assayed for Kd, IC_{50} and CAT activity. Proteins bearing mutations affecting HTH1 were found to bind telomeres with high affinity, but a dramatic loss in binding was seen with HML-E sites. These data, combined with experiments using altered telomere and HMR-E sites, suggested that HTH1 and HTH2 bind DNA in a cooperative manner: the affinity of one subdomain is enhanced when the second subdomain is bound to DNA. Expression of HTH1 in *E. coli* resulted in a protein able to bind DNA only non-specifically.

Deletion of a “tail” region at the N-terminus of the DBD was also performed. This region has been implicated in DNA binding and makes base-specific contacts in the major
Abstract

groove. Deletion of this region resulted in a protein able to bind telomere and silencer DNA with high affinity, suggesting that this region is not essential for DNA binding or HTH1-HTH2 cooperative interaction. Deletion of a “loop” region between HTH1 and HTH2 was also performed, and results suggested a role for this region in optimal binding of Rap1 to tandemly repeated telomere sequences.

Rap1 is a multifunctional protein in *S. cerevisiae*. It plays a role in chromatin organization, gene silencing and transcriptional activation. Activation alone takes place using up to 100 Rap1 DNA binding sites on the genome. All of these functions are mediated by Rap1 binding to the DNA duplex, and the Rap1 DBD structure must be adaptive to the requirements of the various cellular processes. This thesis suggests that the HTH1, HTH2, “tail” and “loop” regions are specifically designed to optimize Rap1 function. Cooperative binding to DNA by HTH1-HTH2 allows a certain degree of degeneracy at high affinity binding sites. “Tail” interactions, while not essential for DNA binding, maintain binding site specificity at a “core” ACC sequence. Finally, the “loop” region allows proper alignment of HTH1 and HTH2, increasing the cooperative association of these two motifs.
# Table of Contents

ABSTRACT ........................................................................................................ ii

TABLE OF CONTENTS .................................................................................... iv

LIST OF FIGURES .......................................................................................... ix

LIST OF ABBREVIATIONS .......................................................................... xi

ACKNOWLEDGMENTS .................................................................................... xiii

I. INTRODUCTION ......................................................................................... 1

1. DNA packaging and regulation - overview ............................................. 1

2. *S. cerevisiae* as a model system for chromatin study ........................... 2

3. Telomeres, RAP1 and chromatin regulation at interphase .................. 2

3.1. The role of telomeres ........................................................................... 2

3.2. Telomere regulation ........................................................................... 3

3.3. Rap1 and the Sir gene products ......................................................... 3

3.4. Involvement of histones H3 and H4 .................................................. 4

3.5. The telomeric position effect .............................................................. 6

3.6. Telomeres and the formation of DNA quadruplexes ....................... 6

4. Mechanism of RAP1 action ................................................................... 7

4.1. Rap1 overview .................................................................................... 7

4.2. Sir protein mediation to the chromatin complex ............................ 9

4.3. The Rap1 DNA binding domain ...................................................... 9

4.4. DNA binding sequence ................................................................... 10

4.5. Gene activation region ..................................................................... 11

4.6. DNA bending mechanism and single stranded binding ............... 12

4.7. Phosphorylation of Rap1 and regulation by PKA ........................... 12

5. Silencing at the HM Loci of Chromosome 3 ....................................... 14

5.1. Mating type and mating type switching in yeast ............................ 14

5.2. Components of HM loci silencing .................................................. 15

5.3. Involvement of Sir gene products and histones H3/H4 .................. 17

5.4. Silencing and the telomeric position effect .................................... 17

6. Gene Activation by Rap1 ................................................................. 18

6.1. Gene activation - overview .............................................................. 18
# Table of Contents

6.2. Gene activation by Rap1 ........................................... 19

7. Protein-DNA interaction - Overview ........................................... 20
   7.1. Overview of DNA binding proteins ........................................... 20
   7.3. Various models of HTH DNA binding ........................................... 26
   7.4. Myb protein DNA binding characteristics ........................................... 27
   7.5. Prediction of DNA binding domain structure ........................................... 28
   7.6. Cooperative binding of proteins to DNA ........................................... 30

8. Crystal structure of the RAP1 DBD - implications for DNA binding .......... 32

9. The present study .......................................................... 36

II. MATERIALS AND METHODS .................................................. 38
   1. Strains, plasmids and oligonucleotides ........................................... 38
      1.1. Strains used .......................................................... 38
      1.2. Handling of strains and general techniques ........................................... 38
      1.3. Plasmids .......................................................... 39
         1.3.1. Rap1 and DBD expression plasmids ........................................... 39
         1.3.2. CAT assay plasmids ........................................... 40
         1.3.3. HTH1 expression plasmid ........................................... 40
      1.4. General techniques for plasmids ........................................... 40
         1.4.1. Plasmid preparation techniques ........................................... 40
         1.4.2. Agarose gel electrophoresis ........................................... 41
      1.5. Oligonucleotides .................................................. 42
   2. Plasmid manipulations .................................................. 42
      2.1. Cloning techniques .................................................. 42
         2.1.1. PCR .......................................................... 42
         2.1.2. DNA restriction digests ........................................... 44
         2.1.3. Ligation .......................................................... 44
         2.1.4. DNA concentration estimation ........................................... 44
      2.2. Site directed mutagenesis ........................................... 44
      2.3. Sequencing .................................................. 46
   3. Protein expression and purification ........................................... 46
Table of Contents

3.1. General techniques for proteins ................................................. 46
   3.1.1. SDS-PAGE analysis of protein ....................................... 46
   3.1.2. Western blot analysis .................................................. 46
   3.1.3. Raising polyclonal antibodies against Rap1 protein .............. 47

3.2. Protein expression .................................................................. 47
   3.2.1. BL21(λDE3) expression of Rap1 and derivatives ................. 47
   3.2.2. pRG50 expression .......................................................... 48
   3.2.3. pET22b(+) expression of HTH1 ...................................... 48

3.3. Protein purification ............................................................... 48
   3.3.1. Mono Q/S purification .................................................... 48
   3.3.2. Preparation of crude extracts from E. coli ....................... 49

4. Mobility shift techniques ........................................................ 50
   4.1. General techniques ........................................................... 50
   4.2. Kd determination ............................................................... 50
       4.2.1. Determination of active protein concentration ............... 51
       4.2.2. Kd determination ....................................................... 52
   4.3. Competition analysis ........................................................ 52
   4.4. Cross linking mobility shift assay ....................................... 53

5. CAT assay .............................................................................. 55
   5.1. Handling of Cos-7 Cells ..................................................... 55
   5.2. CAT assay protocol ........................................................... 55
   5.3. Measurement of chloramphenicol acetylation ....................... 55
   5.4. Statistical analysis ............................................................ 56

6. BLAST analysis of RAP1 .......................................................... 56

7. Visualizing the RAP1 three dimensional structure ....................... 57
   7.1. SETOR analysis of crystal structure .................................... 57
   7.2. Modeling of site-directed mutations ................................... 57
   7.3. Cartoon structure of DBD subdomains ................................ 57

8. Phosphorylation of RAP1 ......................................................... 58

III. RESULTS ............................................................................. 59
Section A. Prediction of Rap1 DNA binding domain recognition helices from its primary sequence

1. Analysis of the Rap1 DNA binding domain using BLAST (Basic Local Alignment Search Tool) ........................................ 59
2. Model prediction of HTH1 ........................................................................ 59
3. Testing of mutants ..................................................................................... 62
4. Model prediction of HTH2 ........................................................................ 64
5. Comparison of the Rap1 DBD model prediction and DBD crystal structure .......... 66
6. Determination of Rap1 and site-directed mutant dissociation constants .......... 66

Section B. Comparison of Rap1 affinity for telomeric and silencer DNA binding sites .. 68

1. Affinity of Wild-type Rap1 for telomere and silencer sites ......................... 73
   1.1. $K_d$ determination .............................................................................. 73
   1.2. Competition assay .............................................................................. 73
   1.3. CAT assay .......................................................................................... 76
2. Arginine 404 to lysine mutant has reduced affinity to silencer sites .......... 76
   2.1. $K_d$ determination .............................................................................. 77
   2.2. Competition assay .............................................................................. 77
   2.3. CAT assay .......................................................................................... 82
3. Affinity studies of other mutations at positions 404 and 408 ..................... 82
   3.1. $K_d$ determination .............................................................................. 86
   3.2. Competition assays ............................................................................. 86
4. Lys-446 and Lys-360 mutations affect binding affinity ......................... 86
   4.1. $K_d$ determination for mutants K446A and K360A ............................ 88
   4.2. K446A and K360A competition assays ............................................ 88
   4.3. K446A CAT assay .............................................................................. 96
5. Mutants deleting the residue 588 to 594 region and residue 418 to 438 region affect DNA binding .................................................. 96
   5.1. Determination of $K_d$ for 588 to 594 deletion .................................... 99
   5.2. Competition assay .............................................................................. 99
   5.3. CAT assay .......................................................................................... 99
6. The effect on binding of a G590V, R404K Rap1 double mutant .......... 99
6.1. G590V $K_d$ determination ................................................................. 102
6.2. G590V competition assay ................................................................. 102
6.3. Determination of $K_d$ for the R404K/G590V double mutant .............. 102

7. Affinity of Rap1 for TEL$^{A3}$ and HMR-E$^{A10C11}$ binding sites ........... 107

Section C. Studies of the HTH1 binding motif ........................................ 107
1. Expression of the HTH1 motif .............................................................. 107
2. Assaying HTH1 DNA binding affinity by cross-link mobility shift analysis .... 109
3. \textit{In vitro} PKC phosphorylation of DBD and HTH1 deletions .............. 109

IV. Discussion ............................................................................................. 113
1. Primary sequence of the RAP1 DBD and studies of RAP1 DNA binding can be
   used to model a bipartite Helix-Turn-Helix DNA binding domain .............. 113
2. Modeling other HTH DNA binding domains - various issues .................. 114
3. Wild-type RAP1 has similar affinity to telomeric and silencer binding sites \textit{in vitro}
   ................................................................................................................. 115
4. \textit{In vivo} studies suggest RAP1 has varied affinities for its binding sites ...... 117
5. Mutation of HTH1 affects binding to RAP1 sites with altered 3’ binding sequences
   ................................................................................................................. 118
6. Minor groove interactions help stabilize RAP1-DNA binding \textit{in vitro} and \textit{in vivo}...
   ................................................................................................................. 119
7. The DBD “tail” region is not essential for RAP1 DNA binding .............. 120
8. The DBD “loop” region affects DNA binding affinity ....................... 120
9. RAP1 DNA binding involves cooperative interaction between HTH1 and HTH2,
   independent of the “tail” region ............................................................... 121
10. Relation of Rap1 structure/function to chromatin organization ............ 123
11. Phosphorylation and DBD binding affinity ........................................ 125
12. Conclusions ............................................................................................ 125
13. Future Prospects ................................................................................... 126

V. REFERENCES .......................................................................................... 127
LIST OF FIGURES

Fig. 1. Model of telomere silencing by Rap1 and the Sir gene products..........................5
Fig. 2. Rap1 functional domains.................................................................................8
Fig. 3. Rap1 transcriptional activation......................................................................13
Fig. 4. Mating Type Switching in Yeast.................................................................16
Fig. 5. DNA Binding Motifs.....................................................................................21
Fig. 6. Comparison of Helix-Turn-Helix DNA Binding Domains..........................24
Fig. 7. Determination of HTH and homeodomain by primary structure homology......29
Fig. 8. Three dimensional crystal structure of DBD and ribbon diagram of DBD subdomains..............................................................................................................34
Fig. 9. Oligonucleotides used in present study..........................................................43
Fig. 10. Sample competition assay plot......................................................................54
Fig. 11. Protein sequence comparison......................................................................60
Fig. 12. Rap1 mobility shift of TEF2 binding site....................................................63
Fig. 13. Putative HTH2 sequence comparison...........................................................65
Fig. 14. Purification of Rap1 proteins........................................................................67
Fig. 15. $K_d$ determination of DBD site directed mutants........................................69
Fig. 16. Comparing Rap1 telomere and silencer DNA binding sites......................72
Fig. 17. Rap1 protein interactions with telomeric and silencer binding sites under various conditions........................................................................................................74
Fig. 18. Three-dimensional model of the Rap1 Arg-404 to lysine mutation............78
Fig. 19. R404K protein interactions with telomeric and silencer binding sites under various conditions........................................................................................................80
Fig. 20. Three dimensional model of the R408N mutation.......................................83
Fig. 21. Three dimensional model of the R404D and the R408D mutations............83
Fig. 22. R404D, R408N and R408D protein interactions with telomeric and silencer binding sites under various conditions.................................................................87
Fig. 23. Three dimensional model of the K446A mutation.......................................89
Fig. 24. Three dimensional model of the K360A mutation.......................................89
Fig. 25. Interactions of K446A protein with telomeric and silencer binding sites under various conditions........................................................................................................92
List of figures

Fig. 26. K360A protein interactions with telomeric and silencer binding sites under various conditions .................................................. 94
Fig. 27. Three-dimensional model of deletions 418 to 438 and 588 to 594 ............... 97
Fig. 28. Interactions between “tail” and “loop” deletions and Rap1 binding sites under various conditions ........................................................................................................... 100
Fig. 29. Three dimensional model of the G590V mutation .................................. 103
Fig. 30. G590V and G590V/R404K double mutant protein interactions with telomeric and silencer binding sites under various conditions .................................................. 105
Fig. 31. Competition analysis of altered TEL and HMR-E binding sites .................. 108
Fig. 32. Analysis of HTH1 DNA binding affinity .................................................. 110
Fig. 33. Rap1 DBD and HTH1 Heart Muscle Kinase Reactions ............................. 111
Fig. 34. Summary of Rap1 binding affinity data .................................................... 116
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Å</td>
<td>angstrom</td>
</tr>
<tr>
<td>Amp</td>
<td>ampicillin</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine-5’-triphosphate</td>
</tr>
<tr>
<td>BLAST</td>
<td>basic local alignment search tool</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>C-</td>
<td>Carboxy-</td>
</tr>
<tr>
<td>cAMP</td>
<td>cyclic adenosine-5’-monophosphate</td>
</tr>
<tr>
<td>CAT</td>
<td>chloramphenicol acetyl transferase</td>
</tr>
<tr>
<td>Ci</td>
<td>Curie</td>
</tr>
<tr>
<td>cpm</td>
<td>count per minute</td>
</tr>
<tr>
<td>CTP</td>
<td>cytidine-5’-triphosphate</td>
</tr>
<tr>
<td>dATP</td>
<td>deoxyadenosine-5’-triphosphate</td>
</tr>
<tr>
<td>dCTP</td>
<td>deoxycytidine-5’-triphosphate</td>
</tr>
<tr>
<td>DBD</td>
<td>DNA binding domain</td>
</tr>
<tr>
<td>dIdC</td>
<td>deoxyinosine - deoxycytidine</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNase</td>
<td>deoxyribonuclease</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>E. coli</td>
<td>Escherichia coli</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EtOH</td>
<td>ethanol</td>
</tr>
<tr>
<td>FBS</td>
<td>fetal bovine serum</td>
</tr>
<tr>
<td>FPLC</td>
<td>forced pressure liquid chromatography</td>
</tr>
<tr>
<td>HMK</td>
<td>heart muscle kinase</td>
</tr>
<tr>
<td>HTH</td>
<td>Helix Turn Helix</td>
</tr>
<tr>
<td>IC_{50}</td>
<td>Concentration Required for 50% Inhibition</td>
</tr>
<tr>
<td>ICD</td>
<td>interchromosome domain</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl-β-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>kb</td>
<td>kilo base</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>kDa</td>
<td>kilo dalton</td>
</tr>
<tr>
<td>$K_d$</td>
<td>dissociation constant</td>
</tr>
<tr>
<td>K. lactis</td>
<td>Kluyveromyces lactis</td>
</tr>
<tr>
<td>MES</td>
<td>methylethylsulfonate</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>N-</td>
<td>Amino-</td>
</tr>
<tr>
<td>NaDOC</td>
<td>sodium deoxycholate</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>Oligo</td>
<td>Oligodeoxyribonucleotide</td>
</tr>
<tr>
<td>ORC</td>
<td>origin recognition complex</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PIC</td>
<td>pre-initiation complex</td>
</tr>
<tr>
<td>PKA</td>
<td>protein kinase A</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethylsulfonyl fluoride</td>
</tr>
<tr>
<td>rDNA</td>
<td>ribosomal DNA</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>rpm</td>
<td>rotations per minute</td>
</tr>
<tr>
<td>SAAB</td>
<td>selected and amplified binding</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>Saccharomyces cerevisiae</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium docecyl sulphate</td>
</tr>
<tr>
<td>SV40</td>
<td>simian virus 40</td>
</tr>
<tr>
<td>TEL</td>
<td>telomere</td>
</tr>
<tr>
<td>TLC</td>
<td>thin layer chromatography</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer ribonucleic acid</td>
</tr>
<tr>
<td>ts</td>
<td>temperature sensitive</td>
</tr>
<tr>
<td>UAS</td>
<td>upstream activating sequence</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>VP16</td>
<td>viral protein 16</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

First and foremost I would like to thank my supervisor Dr. Michel Roberge for providing me with a dynamic research project and boundless support. His suggestions have made this thesis what it is. I am grateful to my committee members, Dr. George Mackie and Dr. Lawrence McIntosh, for their advise on various aspects of my research. Their diverse backgrounds have greatly enhanced my understanding of the scientific theories behind my work.

I would like to acknowledge the important contribution to my research from members of the McIntosh lab, especially Logan Donaldson and Manish Joshi. Logan’s knowledge of the Myb DNA binding domain, and his aid in RAP1 NMR analysis were instrumental in providing the initial direction to my research. Manish provided both spectroscopic analysis of Rap1, as well as helpful suggestions too numerous to describe.

To the past and present members of the Roberge lab: Hilary, Deb, LiLi, Pat and Wayne. I would like to sincerely thank you for the positive environment you have provided me.
I. INTRODUCTION

1. DNA PACKAGING AND REGULATION – OVERVIEW

While the importance of regulating chromatin structure during mitosis has been known for many years, the involvement of chromatin in DNA function during interphase is not clearly understood. Evidence indicates that the chromosomes are not left in a random state during interphase. In various studies, including *in situ* hybridization, a clear separation of functional regions of the chromosome can be observed (reviewed in Cremer et al., 1993). Incubation of mammalian cells with whole chromosome probes demonstrates that the chromosomes occupy specific territories and are not intertwined in the nucleoplasm. Such organization provides a framework to organize transcription, the processing and transport of mRNA, replication and recombination. In their model (Cremer et al., 1993), the chromosome territories are relatively impenetrable areas of chromatin, with transcription and other processes taking place on the exterior of their surfaces and extending into the "inter-chromosome domain" (ICD) which exists between chromosome territories. Aside from the nucleolus, the production of RNA and gene regulation take place at the interface of the chromosome territory and the ICD in mammalian cells.

A key aspect in DNA packaging is the role of the histone proteins. Histone proteins, including H1, H2A, H2B, H3 and H4, are involved in the creation of nucleosomes and higher order levels of chromatin folding. These proteins form a stable complex essential to chromosome packaging (Campbell, 1996).

Another aspect of chromatin regulation, related to "chromosome territories" and the histone complex, is the existence of areas of condensed chromatin termed heterochromatin. Heterochromatin was originally described as DNA which stains darkly, even in interphase cells (Heitz, 1928). In *Drosophila melanogaster*, up to a third of the chromatin is condensed in interphase. Heterochromatin is late-replicating and its histones are typically hypoacetylated (Turner et al., 1992). When genes from euchromatin are transposed near heterochromatin, they may be repressed (see Locke et al., 1988), suggesting that heterochromatin is involved in transcriptional silencing. This silencing, termed position-effect variegation, includes resistance to bacterial restriction endonucleases (Loo and Rine, 1996) and histone H4 hypoacetylation.
2. S. CEREVISIAE AS A MODEL SYSTEM FOR CHROMATIN STUDY

The budding yeast S. cerevisiae has become a popular system for the modeling for various eukaryotic cell processes. Genetic and biochemical studies of cell cycle regulation, signal transduction, chromatin regulation and transcriptional activation have been aided by the ease with which yeast can be cultured and manipulated.

In S. cerevisiae, the primary compartmentalization of the nucleus involves separation of the majority of the chromatin (roughly two thirds by volume) from the nucleolus, where rDNA is transcribed and ribosomes are assembled (Henriquez et al., 1990). Yeast chromatin has also been found to interact via specific regions with the nuclear envelope and nuclear pores (Gasser, 1991). In addition, the yeast telomeres are known to be isolated at perinuclear foci in the nucleus, resulting in a specialized heterochromatin complex (Klein et al., 1992). This heterochromatin complex may play a role in gene silencing, by inhibiting the interaction of transcription factors (Gilson and Gasser, 1995). The role of heterochromatin in telomere regulation (Grunstein, 1997) is described in the following section.

Beyond this compartmentalization, the genome of yeast contains specific sites which show increased sensitivity to DNase upon the induction of gene transcription (Lohr and Hopper, 1985). These studies support the theory that the histone complex does not protect regions of actively transcribed genes. Beyond these regions of hypersensitivity, DNase sensitivity of the yeast chromosome is comparable between transcribed and nontranscribed regions of the genome (Lohr and Hereford, 1979), suggesting that S. cerevisiae DNA is always in a transcriptionally poised state during interphase.

3. TELOMERES, RAP1 AND CHROMATIN REGULATION AT INTERPHASE

3.1. The role of telomeres

Telomeres, the ends of linear eukaryotic chromosomes, are specific protein-DNA complexes essential for the stability and replication of the chromosome (Zakian, 1989). Telomeres from various species, including vertebrates, fungi, plants and protozoa contain G-rich repeats. These repeats protect the ends of the DNA and allow the replication of the chromosomal termini (Biessman and Mason, 1992). One of the key roles of telomeres is the
maintenance of chromosome length. During DNA synthesis, the removal of RNA primers results in a 5'-terminal gap in one of the daughter molecules. Over numerous generations, repetitive priming would result in a loss of chromosome ends. In addition, exonucleases have been shown to degrade broken chromosomes (Blackburn and Szostak, 1984), suggesting that without telomeres, chromosomes would be subject to a gradual shortening.

3.2. Telomere regulation

Three mechanisms have been identified which help maintain the chromosome ends. First, a telomere-specific terminal transferase activity results in the elongation of the telomere ends. This telomerase enzyme is a ribonucleoprotein reverse transcriptase, which carries its own template RNA (reviewed in Shore, 1997). In yeast, telomerase activity results in maintenance of the G-rich repeat length at 150 to 350 base pairs. This repeat terminates in a 3' single-stranded overhang of variable length (Wellinger et al., 1993).

A second mechanism is recombination between telomeres, which has been observed in S. cerevisiae (Wang and Zakian, 1990). Recombination events occur between Y' elements, 5 to 6 kb in size, located directly adjacent to telomeric repeat sequences. Y' elements have no known function, but do include an open reading frame for RNA helicase (Louis et al., 1994). These elements are highly conserved, suggesting that they recombine preferentially (Louis et al., 1994).

A third mechanism of telomere maintenance involves creation of a heterochromatic state. As stated previously, telomeres are concentrated at foci in the nucleus. Telomeric foci can be dispersed through disruption of the amino termini of histones H3 and H4 (Hecht et al., 1995). As discussed in the next section, this heterochromatin is composed of a protein complex involving the yeast proteins Rap1, Sir2, Sir3, Sir4 and histones H3/H4 (Hecht et al., 1995). These results suggest that the foci contain heterochromatin.

3.3. Rap1 and the Sir gene products

In S. cerevisiae, the Rap1 gene product binds with high affinity to the G-rich DNA sequence of telomeres (Berman et al., 1986). In experiments involving the under-expression of Rap1 protein, telomere length is reduced, while overproduction increases both telomere length
and heterogeneity (Conrad et al., 1990; Lustig et al., 1990). This suggests that Rap1 and similar proteins in other species help to stabilize the telomere. In yeast, this stability is maintained through the creation of heterochromatin at the telomeres (for review see Grunstein, 1997).

The first step in this process involves the binding of Rap1 protein to the telomere ends. A DNA binding domain (DBD), contained in the middle of the Rap1 protein (Henry et al., 1990), binds strongly to the telomeric repeat sequence \([\text{(C)}_{1,3}\text{A}]_n\) \textit{in vitro} (Vignais et al., 1990). \textit{In vivo} studies have found that telomeres can be immunoprecipitated \textit{in vitro} by anti-Rap1 antiserum (Conrad et al., 1990). Rap1 also interacts through protein-protein contacts with the gene products of \textit{SIR3} and \textit{SIR4} (Cockell et al., 1995; Strahl-Bolsinger et al., 1997). Two-hybrid studies and immunoprecipitation (Cockell et al., 1995; Hecht et al., 1996) have indicated that Sir3 and Sir4 form homodimers and interact with each other \textit{in vivo}.

Interaction between Rap1 and Sir3 or Sir4 is through the C-terminal portion of the Rap1 protein (Liu et al., 1994). Another protein, Sir2, is a histone deacetylase involved in the hypoacetylation of this complex, and this hypoacetylation relative to active chromatin suggests a repressed chromatin complex (Braunstein et al., 1993; Landry et al., 2000).

Fluorescence microscopy using Rap1 and Sir protein antibodies shows that Rap1 localizes with the telomeres at distinct foci (Klein et al., 1992; Gotta et al., 1996). In addition, it was estimated that at least 20% of the Rap1 protein in the cell is localized to the telomeres. This high concentration is partly explained by the ability of Rap1 to bind up to once every 18 base pairs to a telomeric site \textit{in vitro} (Gilson et al., 1993). All of the current data are consistent with a model in which Rap1 binds tightly and with high affinity to telomeres, and that Sir3 and Sir4 gene products bind to the Rap1 matrix. A suggestion that this complex forms heterochromatin was supported by both genetic data and by association between Sir3/Sir4 and histone H3/H4 (Gilson and Gasser, 1995). A model of these interactions is shown in Figure 1.

### 3.4. Involvement of histones H3 and H4

Sir3 and Sir4 gene products interact directly with histone H3 and H4 \textit{in vitro} (Hecht et al., 1995). This association is through the histone N-termini, and these N-termini are required for the subnuclear localization of Sir3, and perinuclear localization of telomeres. These data confirm a model whereby Rap1 and the Sir gene products "label" telomeric DNA for
Introduction

Figure 1. Model of telomere silencing by Rap1 and the Sir gene products

Conclusions from the study of Rap1 telomere silencing (Cockell et al., 1995) are shown above. Rap1 binds with a frequency of approximately one molecule per 18 base pairs at the telomere ends. Rap1 interacts with Sir3 and Sir4 homodimers via the Rap1 C terminus (Liu et al., 1994). The Sir proteins then form protein-protein interactions with histones H3 and H4, forming a hetero-chromatin state.
condensation into heterochromatin.

3.5. The telomeric position effect

The heterochromatin state of telomeres can also influence the transcriptional activity of the region. Typically, electron dense, negatively stained heterochromatin remains highly condensed and localized at the nuclear periphery, as in the case of the inactive X chromosome of mammalian females (Walker et al., 1991). In addition, S. cerevisiae chromosomal regions adjacent to the telomeres are found near the nuclear envelope and repress gene activity (Pryde and Louis, 1999). This repression, termed the telomeric position effect, spreads 3-5 kb from the telomeres (Renauld et al., 1993). This repression is dependent on the Rap1/Sir/histone interaction.

Taken together, these studies suggest that in yeast a multiprotein complex containing Rap1, the Sir protein and histones H3 and H4 maintain repression at nearby genes. Rif1, a protein which antagonizes the action of the Sir proteins for silencing, may play a role to balance the activity of this multiprotein complex, perhaps by binding Rap1 and blocking interaction with the Sir proteins (Kyrion et al., 1993).

3.6. Telomeres and the formation of DNA quadruplexes

In addition to binding double-stranded telomeric repeats, Rap1 has been shown to bind single stranded telomeric DNA with high affinity (Giraldo and Rhodes, 1994). This interaction is significant given the 3’ single-stranded telomere extensions created by telomerase.

Association of telomere ends in perinuclear foci raises the possibility of telomere-to-telomere associations. Indeed, Rap1 has been found to stabilize a unique quadruplex formation of guanine rich strands in vitro (Giraldo and Rhodes, 1994). If Rap1 is the protein associated with the single stranded telomere ends, it may be involved in quadruplex formation to assemble the various telomere ends into a single complex. The mechanism of quadruplex formation is as yet unknown.
4. MECHANISM OF RAP1 ACTION

4.1. Rap1 overview

The *S. cerevisiae* Rap1 DNA binding protein has been found to play a key role in the regulation of various cell processes, including telomere length regulation, regulation of silencing at the yeast silent mating type loci, and transcriptional activation of glycolytic enzymes (described below, reviewed in Gilson and Gasser, 1995). Rap1 is an essential protein in yeast, and removal of the Rap1 gene results in cells unable to undergo mitotic growth (Shore and Nasmyth, 1987).

Rap1 is a 827 amino acid protein with a molecular weight of 116 kDa. The primary sequence of Rap1 contains only weak identity to other proteins, besides a Rap1 homologue from the fission yeast *Kluveromyces lactis* (Larson et al., 1994). Identity between *S. cerevisiae* and *K. lactis* Rap1 is mainly clustered in the middle of the protein, at the DBD (Henry et al., 1990). A moderate level of identity (50%) is seen in the C-terminal 159 amino acids of both proteins, suggesting that this region has a conserved role. The various subdomains of Rap1 are indicated in Figure 2.

Henry et al. (1990), who delineated the DBD through deletion analysis, found no similarity between this DBD and other DNA binding motifs. In the subsequent characterization of the *K. lactis* Rap1 homologue (Larson et al., 1994) the authors noted a weak identity between the Rap1 DBD and the c-Myb proto-oncogene DNA binding motif. This identity was confirmed by X-ray crystallography (Konig et al., 1996).

As noted above, Rap1 was characterized in various contexts as a transcriptional activator (Huet et al., 1985), a telomere binding protein (Berman et al., 1986) and a silencer of transcription (Buchman et al., 1988; Longtine et al., 1989). Roughly 8x10³ molecules of Rap1 are present in the cell, giving about one molecule of Rap1 bound per 2 kb of genomic DNA (reviewed in Gilson and Gasser, 1995). This is a high concentration for a transcription factor, and puts Rap1 in the class of a sub-chromatin component. Sub-chromatin components are defined as proteins present at a lower concentration than the histones, but still involved in the regulation of chromatin.
Figure 2. Rap1 functional domains

Rap1 domains elucidated by deletion analysis are shown above. The N terminal bending domain is non-essential. The activation and SIR binding domains overlap, with the activation domain extending to amino acid 695 (Reviewed in Gilson and Gasser, 1995).
4.2. Sir protein mediation to the chromatin complex

As noted in section 3.3, Rapl interacts specifically with various Sir proteins, including Sir3 and Sir4 both in vitro and in vivo (Moretti et al., 1994; Cockell et al., 1995). In the work of Cockell et al. (1995), immunofluorescence staining of yeast nuclei showed that Rapl, Sir3 and Sir4 proteins localize to foci at the nuclear periphery. Deletion of the 28 C-terminal residues of Rapl results in a loss of Sir3 immunolocalization. In a second set of experiments, Sir4 and Rapl were found to co-immunoprecipitate.

Interaction with both Sir proteins occurs via the C-terminal portion of Rapl (Kyrion et al., 1992), and various studies suggest that the interaction of Rapl with Sir3 and Sir4 causes the creation of a heterochromatic state at telomeres and silencer sites. This heterochromatic state is dependent on the location of telomeres and the silencer genes, discussed in section 5.

In a study by Maillet et al. (1996), transfer of a silenced reporter gene away from the chromosome ends resulted in a loss of silencing, which was restored by the overexpression of Sir3 and Sir4 genes. These data suggest that the ends of the chromosome, already known to be foci containing high concentrations of Rapl and Sir3/Sir4, provide a pool of proteins required to maintain silencing. The association of Sir3/Sir4 with histones H3/H4 provides a mechanism for the formation of heterochromatin, and the repression of gene expression at these sites.

4.3. The Rapl DNA binding domain

Henry et al. (1990) used in vitro transcription/translation to determine the minimal fragment of Rapl required for efficient binding to a phosphoglycerate kinase (PGK) upstream activating sequence (UAS) Z+ subfragment. The Z+ fragment contains a PGK promoter sequences from –473 to –409, and is a relatively weak Rapl binding site (Henry et al., 1990). This fragment was radiolabeled and then incubated with the various truncated Rapl proteins. Their study determined that amino acids 361 to 596 of Rapl contain a minimal DBD to the Z+ subfragment. An internal Rapl fragment from 378 to 691 was unable to bind the Z+ site, as was the fragment 302 to 583. This suggests that residues between 361 and 378 are required for optimal binding of Rapl to DNA targets as are residues from 583 to 596.

The minimal DBD fragment (361-596) bound strongly to HMR-E (a silencer binding site), TEF2 (translational elongation factor 2 promoter), RP51 (a ribosomal protein gene promoter) and ENO1 (a glycolytic gene promoter). In contrast, binding to the PGK Z+
fragment in the same assay was weak, suggesting that the minimal DBD bound quite poorly to this site.

4.4. DNA binding sequence

Rap1 binds to a variety of sites in the yeast genome and a 15 base pair consensus has been determined (Graham and Chambers, 1994). This consensus (5’- (A/G) T (A/G) C A C C C A N N C (C/A) C C -3’) is loosely conserved at each Rap1 binding site on the genome. Binding of Rap1 to a consensus telomeric site (A C A C C A C A C A C C C) exhibits a dissociation constant ($K_d$) of $1.3 \times 10^{-11}$ M (pH 8, 50 mM KCl, 5 mM MgCl$_2$, 0.5 mM CaCl$_2$, 20 mM Tris-HCl), indicating that Rap1 binds strongly to telomeric sites (Vignais et al., 1990). It has been shown that Rap1 can bind once every 18 base pairs at telomeres (Gilson et al., 1993). The $K_d$ for non-specific binding of Rap1 to DNA is $8.7 \times 10^{-6}$ M (pH 8, 50 mM KCl, 5 mM MgCl$_2$, 0.5 mM CaCl$_2$, 20 mM Tris-HCl).

In order to determine the specificity of the Rap1 DBD to its binding site, Graham and Chambers (1994) used the selected and amplified binding protocol (SAAB) to determine the optimal binding site for the Rap1 protein. Using in vitro translated Rap1 the researchers sequenced a series of SAAB products, and derived the consensus 15 base pair Rap1 binding site. Forty-seven SAAB products were sequenced to derive this consensus, and only one sequence was found not to bind Rap1 in vitro. Within this consensus, the central region, nucleotides 4 through 9 (sequence C A C C C A), were highly conserved, with nucleotides on either side being quite variable. Of the 47 sites tested, mismatches frequently occurred at the 3’ end of the binding site. Between 22 and 25 mismatches occurred at nucleotides 10 to 14 (the sequence C C/A C C). Within that context, significant change from the consensus was tolerated at the 5’ end of the site or at the 3’ end but changes at both the 5’ and 3’ ends were not tolerated. These results suggest that the “core” region (bases 4-9) could maintain DNA binding with additional contacts at either the 5’ or 3’ ends. This result would explain why previously determined in vivo binding sites HML-E and TEF2, which contain a 3’ TA and 3’ TT, respectively, are still considered strong binding sites. Previous to the SAAB study, the Rap1 consensus was 5’- (A/G) (A/C) A C C C A N N C A (T/C) (T/C) -3’ (Buchman et al., 1988). This consensus, derived from known Rap1 binding sites, would have considered TEF2 an optimal binding site.
Graham and Chambers (1990) searched the EMBL and GenBank databases to determine the number of \textit{in vitro} Rap1 binding sites present in the promoters of known yeast genes. Using rules derived from their studies, they selected 102 sequences capable of Rap1 binding. This suggests that Rap1 may control a large number of genes, acting as a general transcription factor. Rap1 may alternatively be considered a sub-chromatin component, rather than a gene-specific transcriptional regulator.

In a set of experiments related to those discussed above, purified Rap1 protein was bound to a telomeric or HMR-E binding site and the DNA in the complex tested for sensitivity to KMnO$_4$ or DNase I (Gilson et al., 1993). The results were compared to previous data on methylation interference by Rap1 and DMS protection. The results suggest that the Rap1 protein protects both the "core" region (bases 4 through 6, as noted above) and the 3' end of the consensus site (bases 12 to 15). These two sets of contacts reveal a set of bases (C10 and C12) which have increased sensitivity to KMnO$_4$. These data suggest that the Rap1 DBD, which is large for a DBD, may contain two DNA binding motifs, each contributing to half of the DNA binding specificity. These two contacts distort the DNA slightly, producing increased KMnO$_4$ reactivity.

4.5. Gene activation region

Gene activation by the Rap1 protein involves both the DBD and other elements involved in the transactivation of the RNA polymerase complex. Activation by Rap1 alone involves a small activating region of 65 residues near the C-terminus (Hardy et al., 1992). In their studies, the C terminus of Rap1 was fused to the GAL4 DBD, expressed using a Rap1 promoter from a single copy site integrated into the yeast genome. Activation was measured by a UASgal1-\textit{lacZ} reporter gene in a strain lacking GAL4. While the observed activation may be due to the fortuitous display of an activating domain which does not normally function for the intact Rap1 protein, it suggests that Rap1 contains its own activation domain (Hardy et al., 1992).

Kurtz and Shore (1991) showed that a temperature-sensitive mutation in Rap1 results in defective expression of the MATa gene, which contains a Rap1 binding site in its promoter. To control for the effect of Rap1 mutations on the expression of a-specific genes, these researchers found that they could reverse the MATa phenotype by replacing the Rap1 binding site at MATa with a binding site for the GAL4 transcriptional activator. These results suggest that
Rap1 can activate transcription by binding at the promoter of a gene. Whether Rap1 activates on its own or in concert with other proteins is as yet unclear. In the case of MATα, the GAL11 gene product appears to be required for expression in addition to Rap1 (Nishizawa et al., 1990).

An example of the interaction between Rap1 and another activator protein involves the in vivo protein-protein binding of Rap1 and the GCR1 gene product (see Figure 3). Rap1 and GCR1-dependent activation of the ADH1 and TEF1/TEF2 genes does not require a GCR1 binding site, and a Rap1/GCR1 complex can be co-immunoprecipitated from whole cell extracts (Tornow et al., 1993). This suggests that Rap1 and GCR1 combine to provide optimal activation of these genes. One can view Rap1 as a form of molecular anchor. Rap1 binds to various DNA sites on the genome, and depending on its location on the chromosome, either anchors the Sir gene products for the creation of heterochromatin, or anchors GCR1, GAL11 and others for gene activation.

### 4.6. DNA bending mechanism and single stranded binding

In addition to binding double stranded DNA, Rap1 can bend DNA and bind to various structures present at the telomere ends including single stranded DNA and G-quartets. DNA bending by Rap1 appears to involve the N-terminal portion of Rap1 (residues 1 to 235). Bending has been shown by DNA mobility shift, as well as by Scanning Tunneling microscopy. Using a DBD deletion protein results in a complex which is unable to bend DNA (Vignais and Sentenac, 1989). The function of bending is not known, though it may play a role in the fine tuning of promoters to exclude nucleosomes, which are not present at Rap1 binding sites (Wright et al., 1992).

Rap1 single stranded DNA binding, with a $K_d$ of approximately $3 \times 10^{-8}$ M (Giraldo and Rhodes, 1994), is implicated in the regulation of the telomere ends. These telomere ends contain a stretch of single stranded G-rich DNA. In addition, Rap1 promotes the formation of so-called G-quartets; a structure where four Guanine rich strands form an antiparallel complex.

### 4.7. Phosphorylation of Rap1 and regulation by PKA

A number of yeast transcription factors, including the GAL4 gene product, are regulated by phosphorylation of key residues (Karin and Hunter, 1995). In the case of Rap1, in vivo studies indicate that increased protein kinase A (PKA) activity results in increased transcription
Figure 3. Rap1 transcriptional activation

Representation of Rap1 transcriptional activation at glycolytic gene upstream activating sequences. The combination of Rap1 and GCR1 is required for activation. Protein-protein contacts between Rap1 and GCR1 allow activation, even in the absence of GCR1 binding sites (Reviewed in Gilson and Gasser, 1995).
of ribosomal protein genes, and that this transcription is dependent on the presence of Rap1 (Klein and Struhl, 1994). This effect is mediated through the Rap1 transcriptional activation domain rather than DBD. A yeast strain with constitutively high PKA levels gave a two-fold increase in ribosomal protein mRNA synthesis. This increase was dependent on Rap1 binding sites in the promoter, implicating Rap1, and not another factor, in the process. An extract from a strain with high PKA levels actually displayed less Rap1 DNA binding activity than a wild-type extract, implicating transcriptional activation in this process. Using a LexA DBD hybrid (LexA-Rap1) and a promoter containing LexA rather than Rap1 binding sites, it was found that a region of Rap1 is able to act as a transcriptional activator even without the use of its DNA binding activity.

In addition, in vitro studies indicate that Rap1 and the Rap1 DBD are involved in phosphorylation-dependent binding to the PGK promoter (Tsang et al., 1990). Studies of the in vitro phosphorylation of Rap1 by Heart Muscle Kinase (HMK), the catalytic subunit of PKA, indicate that full length Rap1 contains four phosphorylation sites (Yamazaki, 1995). Two of these sites are located within the DBD, and it was found by phoshoamino analysis that all four phosphorylation sites are serines. Yamazaki’s (1995) studies indicate that in vitro phosphorylation of Rap1 or its DBD increased binding to an 80 bp probe from the TEF2 promoter. Possible sites of phosphorylation were determined by comparing the DBD primary sequence to the target sequence for HMK. One potential site is located at Serine 450, the other at Serine 516, as determined by sequence analysis.

5. SILENCING AT THE HM LOCI OF CHROMOSOME 3

5.1. Mating type and mating type switching in yeast

Haploid cells of the yeast S. cerevisiae contain two alleles, MATa and MATα, involved in the regulation of mating-type specific genes (reviewed in Haber, 1998). Both MATa and MATα contain two divergently transcribed open reading frames. The gene products include MATα1, a positive regulator of α-specific genes, and MATα2, a protein which is both a negative regulator of α-specific genes, and which, in conjunction with the MATα1 gene product, represses haploid-specific genes in a diploid cell. MATα2 function has not yet been assigned. Initial mating type studies of S. cerevisiae determined that there are two strains. The first is a
heterothallic strain (ho) where the MAT genes are stable and mating type is maintained over multiple generations. The second is a homothallic strain (HO) where the haploid cell can switch mating type as often as every generation (Haber et al., 1998).

Sterile mutants of yeast (MacKay and Manney, 1974), and a final model (Hicks et al., 1977) provided the rationale for mating type gene switching (Figure 4). According to this model, haploid cells of either mating type contain complete copies of both a and α information at either end of Chromosome 3, but these genes are maintained in the unexpressed state. These two silent loci are termed HMLα and HMRα. Mating type switching involves the replacement of mating-type specific sequences near the middle of chromosome 3 by one of the two silenced copies.

Preference is toward recombining with the opposite mating type gene. This replacement involves site-specific mitotic gene conversion using the information at HML or HMR. This conversion apparently occurs before replication, since the two daughter cells switch in pairs (for review, see Haber, 1998). During mating type switching, HO endonuclease performs a site-specific mitotic gene conversion of, for example, 750 bp of the MATα region with 650 bp of the HMRα region. This gene-specific conversion is turned off in diploid cells.

The silent state of the HMRα and HMLα genes is maintained through a set of elements located at either end of the coding sequence. These regions, termed E (for essential) and I (for important), were found to contain binding sites for three “general factors” in yeast: the origin recognition complex (ORC), the Abf1 general transcription factor, and the Rap1 protein (Laurenson and Rine, 1992). This is investigated further in the following section.

5.2. Components of HM loci silencing

Despite the fact that HML and HMR contain complete copies of the α and a genes, including their promoters, they are silenced in the yeast cell. It has been shown that this silencing involves a change in chromatin structure in that UV induced damage repair and endonuclease activity are reduced in these regions compared to the mating type locus. A search for cis-acting elements determined that the establishment and maintenance of silencing depended on two sites, E and I, surrounding each silent locus (Feldman et al., 1984). These sites are not identical at the two loci and have different properties. At HML, the presence of either site is sufficient for silencing, while at HMR, deletion of the I site maintains silencing,
Figure 4. Mating Type Switching in Yeast

Schematic diagram of mating type switch from MATα to MATα on Chromosome 3. Chromosomal recombination is mediated by HO endonuclease, allowing the replacement of Ya with Ya. Silencing elements E and I are marked, as well as the divergently transcribed genes α1/α2 and a1/a2 (Reviewed in Haber, 1998).
but deletion of the E site abolishes silencing.

As noted above, HMR-E contains binding sites for three factors: the origin recognition complex (ORC), the Rap1 protein and the Abf1 protein. In vitro mutation of HMR-E (McNally and Rine, 1991) has shown that any one of the three sites can be eliminated without affecting silencing, as long as the other two sites remain.

Abf1 protein is a multifunctional protein, involved in the activation of numerous "housekeeping" genes, some of which are also regulated by Rap1 (Kimmerly et al., 1988). Abf1 protein can be made to function as an upstream activation sequence (UAS) element in constructed plasmids (Buchman et al., 1988), but otherwise may aid transcription by aiding transcription termination (Snyder et al., 1988).

The ORC appears to have a number of roles in the yeast genome, including its function in DNA replication. ORC binding sites are spread throughout the yeast genome, and there are 10 times as many ORC sites as are required for replication (Fangman and Brewer, 1992). The ORC HMRE binding site is required for silencing, and passage through S phase has been found to be a necessary step in silencing (Pillus and Rine, 1989).

5.3. Involvement of Sir gene products and histones H3/H4

A single Rap1 binding site is found at the various silencer sites including HMR-E and HML-E. In parallel with studies of heterochromatin organization of telomeres, it was found that Rap1 anchors a Sir/Histone protein complex that establishes a heterochromatic state at silencers. Various SIR genes are involved in the silencing process, including a HM-specific Sir protein, Sir1. Sir1 appears to promote the initiation of silencing by binding to the ORC complex and to the Sir4 gene product (Triolo and Sternglanz, 1996).

Histones H3 and H4 were shown to be essential for silencing (Laurenson and Rine, 1992). The involvement of these genes in silencing, in addition to evidence that proteins Sir2 and Sir3 influence histone deacetylation (Braunstein et al., 1993), suggest that silencing involves the creation of a heterochromatic state similar to that observed at the telomeres (see section 3.5).

5.4. Silencing and the telomeric position effect

Studies on Rap1 silencing also suggest that Rap1 may form a concentration gradient in
the nucleus, with the molarity of Rap1 decreasing with increasing distance from the chromosome ends (Buck and Shore, 1995). Support for this theory comes from the work of Renauld et al. (1993). By inserting a URA3 gene at various distances from the chromosome end, these authors found that silencing occurred up to around 1.7 kb from the telomere. But with the simultaneous overexpression of Sir3, silencing extended inwards nearly 25 kb.

This position effect was also seen with HML. When HML is carried on a plasmid or integrated at a centromere-proximal location, a single silencer is no longer able to repress transcription. Instead, both HML-E and HML-I are required. Proximity of the telomeres, in addition to one or more silencer sites, is required for silencing to occur. Silencers need to be located near the high concentration of Rap1 at chromosome ends for proper function (Boscheron et al., 1996).

Rap1 can also have the opposite effect on chromatin, opening the heterochromatin. Recent results have shown that Rap1 bound to activator sites can suppress silencing at HM silencers when inserted in front of silenced genes. Bi and Broach (1999) inserted a TEF2 promoter segment containing Rap1 binding sites between the E silencer and HMLα genes. This insertion suppressed silencing of the HMLα genes. At least two Rap1 binding sites were required for this effect. In a related set of experiments, Yu and Morse (1999) found that Rap1 interferes with nucleosome binding at GCN4 sites in vivo. These results contradict the role of Rap1 in initiating a nucleosome complex. It appears that in the presence of multiple Rap1 binding sites, and in the absence of silencer factors ABF1 and the ORC, Rap1 acts as a nucleosome disruptor. Silencing involves the interaction of Rap1, ABF1 and ORC. At telomeres, Rap1 action as a silencer is likely determined by the location of telomeres at distinct foci within the nucleus and the concentration gradient of Rap1 and Sir proteins. Bi and Broach (1999) present a model whereby Rap1 bound to activator sites blocks the spread of the Sir protein complex from the E site towards the MATα genes. They term this disruption a nucleosomal ‘hole’.

6. GENE ACTIVATION BY RAP1

6.1. Gene activation-overview

In eukaryotes, gene activation involves various proteins including RNA polymerase II,
which transcribes the nuclear genes that encode both mRNA and the small nuclear RNAs. Other gene activation involves RNA polymerases I and III. Whether or not RNA polymerase II starts transcription at a particular gene is controlled first by the core promoter, containing the TATA box, and secondly by a binding site for the transcription factor protein complex (Roeder, 1991). The TATA box and other initiator elements fix the transcription start site. However, neither RNA polymerase II nor the transcription factor complex (which together comprise the pre-initiation complex (PIC)) can adjust the rate of initiation of RNA synthesis from basal levels (Hampsey, 1998). Instead, transcription factors, which bind to either proximal or distal sequences, modulate the rate of RNA synthesis via specific interactions with the PIC (Zawel and Reinberg, 1992).

In yeast, binding of transcription factors occurs in the UAS upstream of the TATA box. One or more factors may be involved in gene activation. In some cases distinct gene-specific DNA binding sites are used, while in other instances, one protein binds DNA while the other interacts with the DNA bound protein to activate transcription.

6.2. Gene Activation by Rap1

Potential binding sites for Rap1 are present in the promoters of a large number of yeast genes. Determination of a consensus Rap1 binding site and analysis of the yeast genome sequence indicate over a hundred potential Rap1 binding sites are present in gene promoters (Graham and Chambers, 1994). A role for Rap1 is known at a small number of these potential sites. Originally, Rap1 was identified as a factor controlling the expression of ribosomal protein genes and genes encoding protein components of the translational machinery (Huet et al., 1985; Huet and Sentenac, 1987; Vignais et al., 1987). All ribosomal protein genes (except TCM1, RPS33, RPL2A and RPL2B) contain Rap1 binding sites.

Rap1 was also found to be a UAS binding factor at the MATα promoter (Kurtz and Shore, 1991) and is involved in its transcription in vivo. MATα transcription levels can be decreased in rap1α strains, with transcription levels reversed by replacing the Rap1 binding site with a GAL4 binding site.

In addition, experiments imply that Rap1 is involved in the activation of several glycolytic promoters (Butler et al., 1990; Chambers et al., 1990). Deletion of Rap1 binding sites results in reduced transcriptional activation. It has been shown that Rap1 aids in the activation
Introduction

of various glycolytic genes (ADH1, PYK1, TPI1, ENP1 and TDH3) in concert with the GCR1 protein. Evidence also suggests that Rap1 auto-regulates its own expression via four binding sites in the Rap1 promoter region (Graham and Chambers, 1996).

Rap1 gene activation often involves a complex combination of cis acting elements, rather than a single transcriptional activator (see section 4.5 and Figure 3). This role for Rap1, as a link between other proteins and the DNA, is a consistent theme in all three processes of telomere regulation, silencing and activation.

7. PROTEIN-DNA INTERACTIONS – OVERVIEW

Key processes in the eukaryotic cell, including DNA replication, transcriptional regulation and DNA repair require proteins to interact specifically with the DNA double helix. In the case of transcriptional regulation, protein-DNA interaction is a reversible process, with no modification to the DNA.

7.1. Overview of DNA binding proteins

Of the many mechanisms of DNA-protein binding, three basic motifs define the majority of DBDs from transcriptional regulatory proteins (reviewed in Pabo and Sauer, 1992). These are illustrated in Figure 5. The first family comprises the Helix-Loop-Helix and the Leucine zipper ("coiled coil") motifs (Figure 5A). The Helix-Loop-Helix comprises a basic region involved in DNA binding, two alpha helical motifs joined by a loop region, and occasionally an additional Leucine Zipper for protein dimerization. The HLH segment is involved in forming either homo- or hetero-dimers by a parallel four helix bundle (Ferre-D’Amare et al., 1993). These helices pack together around conserved hydrophobic amino acids.

The Leucine zipper motif also involves hydrophobic interactions between α-helices. In this case the helices, which can be parallel or antiparallel, contain a leucine every seven residues which interacts with a corresponding hydrophobic residue on the other helix (Alber, 1992). The structure forms a so-called "coiled coil", with the α-helices wrapping around each other in a shallow supercoil. The Leucine Zipper allows the dimerization of a variety of transcriptional activators.

Each half of the Helix-Loop-Helix interacts with the major groove via its recognition α-
Figure 5. DNA Binding Motifs

helix. As a monomer it is unable to bind DNA specifically, but the dimer binds with high affinity. Due to the highly conserved mechanism of dimerization, monomers can form a large variety of hetero-dimer species. More than 50 different complexes exist in vivo between members of the fos, jun and selected members of the ATF/CREB family of Helix-Loop-Helix proteins (Kovary and Bravo, 1991). This allows rapid and specific responses to changes in the cellular environment.

The second DNA binding motif includes one or more variations of the Helix-Turn-Helix (HTH) motif (Figure 5B). This motif was first described for the DNA binding proteins of the bacteriophage λ lysogenic to lytic switch mechanism (Ptashne, 1992). A HTH DNA binding motif is composed of two alpha helices and a conserved glycine residue, which allows a turn between the helices. Conserved residues on both helices interact within a hydrophobic core, stabilizing the three dimensional structure. The second helix, the recognition helix, sits in the major groove of the DNA, interacting with bases as well as the sugar-phosphate backbone. These interactions often involve hydrophilic residues, giving the recognition helix an amphipathic character: hydrophobic residues facing the second helix, and hydrophilic residues facing the major groove.

The HTH motif and variants are present in bacterial systems, phage, viruses and eukaryotic DNA binding proteins. The HTH and its variants contain three (as in the case of the Homeobox proteins) or more α-helices stabilized by a hydrophobic core. The two helices (comprising the HTH) are arranged in a more or less perpendicular orientation. Additional α-helices or β-sheets, present in other HTH domains, help maintain the complex.

The third family of DNA binding motifs includes various DNA binding proteins containing a zinc molecule chelated by four cysteine and/or histidine residues (Figure 5C). The zinc finger motif contains a mixture of α-helices and β-sheets arranged such that the ends of the α-helix extend into the major groove. Zinc finger motifs bind DNA via a number of fingers repeated in tandem (Pabo and Sauer, 1992). The zinc finger motif of TFIIIA, where this type of DNA binding protein was first discovered, contains tandem repeats of a 30 residue zinc finger. Each motif contains the sequence pattern Cys-X_{2-4}-Cys-X_{12}-His-X_{3.5}-His. Proteins such as zif268 contain three zinc fingers, while the GLI-DNA complex was shown to contain five zinc fingers (Pavletich and Pabo, 1993). The two cysteines and two histidines coordinate a central Zinc molecule to stabilize the structure.
There are other DNA binding motifs that are structurally homologous to the TFIIIA fingers, including yeast transcription factor GAL4. GAL4 has the sequence pattern Cys-X2-Cys-X6-Cys-X6-Cys-X2-Cys-X6-Cys, which forms a binuclear metal cluster.

Additional motifs to these three families include proteins which use an antiparallel β-sheet for DNA binding (Figure 5D). The crystal structure of the TATA-Box binding protein isoform 2 (TBP2) reveals a saddle-like structure whose β-sheets interact with the TATA-Box minor groove (Nikolov et al., 1992). Within each of the families noted above are numerous variations, indicating that evolution provides many ways for proteins to interact in a specific manner with DNA sequences. Variations within the HTH motif are discussed below.

### 7.2. Principles of Helix-Turn-Helix DNA recognition in eukaryotes

A comparison of λ repressor and HTH-variant DNA binding domains is shown in Figure 6. In eukaryotic systems, a number of HTH motifs have been defined. The Homeodomain family of DNA binding motifs is one grouping that contains an assortment of eukaryotic proteins (Pabo and Sauer, 1992). The POU

Homeodomain

motif (Figure 6C) illustrates this motif. An N-terminal arm fits in the minor groove of the DNA, with helix 3 fitting directly in the major groove. Helix 1 and 2 form an antiparallel arrangement, packing against each other and lying perpendicular to helix 3.

The MATα2 DNA binding motif, while the most divergent member of the Homeodomain family, is structurally conserved. The three dimensional structure (Figure 6B) is comparable to that of POU

Homeodomain

.

Other HTH-containing DNA binding proteins in eukaryotes contain two conserved structures that are essential for sequence-specific binding (for review see Brennan and Matthews, 1989). These include a DNA recognition helix, which makes base-specific contacts in the major groove, and one or more helices involved in stabilizing the recognition helix and providing additional contacts with the DNA backbone. In addition to these structures, other mechanisms for DNA binding may be present. These include N-terminal arms which interact with DNA via the minor groove (Kissinger et al., 1990), regions involved in protein dimerization (Li et al., 1995), and novel regions used to stabilize the structure or provide additional DNA contacts (Brennan, 1993).

While the recognition helix sits in the major groove in all HTH proteins, the orientation...
Figure 6. Comparison of Helix-Turn-Helix DNA Binding Domains

Three-dimensional models of six HTH and HTH variant DNA binding domains are shown. (A) λ repressor DNA binding domain (Jordan and Pabo, 1988). (B) MATα2 homeodomain DNA binding domain (Wolberger et al., 1991). (C) POU-homeodomain DNA binding domain (Klemm et al., 1994). (D) POU specific DNA binding domain (Klemm et al., 1994). (E) C-Myb DNA binding domain repeat 3 (Ogata et al., 1994). (F) ETS DNA binding motif (Donaldson et al., 1996). DNA is not indicated in this model.
of the helix relative to the DNA can vary significantly (Pabo and Sauer, 1992). Some helices lie parallel to the local direction of the major groove, while others can vary up to 15° in either direction. This variation appears to be due to the surrounding regions of the protein, and not simply to the sequence-specificity of the recognition helix. For example, in the Myb DNA binding motif (Figure 6E) the recognition helix is tilted more toward the helical axis when compared to both the homeodomain as well as λ repressor motifs (Ogata et al., 1994).

The stability of this recognition helix is maintained by the other α-helices through a hydrophobic core. Originally, HTH motifs were determined by the conserved spacing of these hydrophobic residues. From crystal structure data we know that these hydrophobic residues play a key role in maintaining the stability of the HTH structure. Hydrophobic residues extend from the face of the recognition helix opposite the DNA helix. These residues are able to form interactions with hydrophobic residues extending from the complementary face of the other helices (Li et al., 1995). Bacterial HTH proteins λ repressor (Figure 6A) and cro contain hydrophobic valine, isoleucine and leucine residues on helix 2 and helix 3, allowing these helices to form a stable structure in the aqueous environment of the cell.

7.3. Various models of HTH DNA binding

A large number of HTH variants have been found to date, with more expected with future three dimensional structural data. HTH variants can either be arranged into sub-groups, or classified as isolated structures. The POUı domain (Figure 6D) varies from λ repressor, with helix 2 and its turn (T') both extended by 3 residues (Klemm et al., 1994). However, the POUı motif is similar to λ in its DNA binding with a similar orientation to the bases of the major groove. Even the position of glutamine residues and the extended hydrogen bonding network of POUı are analogous to those present in λ repressor and 434 repressor. Despite its similarities to λ, the motif contains four helices, with helices 2 and 3 arranged in a variant HTH unit and helices 1 and 4 stabilizing the structure (Klemm et al., 1994).

Another variation of HTH is the hepatocyte transcription factor LFB1/HNF1. Its DNA binding motif contains an extra 21 residue insertion compared to λ repressor, which is divided between an 8 residue extension of helix 2 and a 13 residue addition to the turn.

The most extensive variation within the HTH grouping is with the HNF-3γ and ETS
domain ‘winged helix’ (Figure 6F). This motif contains α-helix (for the HTH motif) in addition to β-sheet. The HTH motif binds DNA in a manner similar to other eukaryotic proteins. H3F-3γ binds DNA as a monomer, with the recognition helix lying in the major groove (Brennan, 1993). The β-sheet motif contains a set of loops (W1 and W2) which interact with more distal parts of the DNA recognition site. These loop interactions permit the total number of DNA contacts for H3F-3γ to approach those of dimeric HTH protein-DNA contacts, allowing high affinity binding with only one recognition helix (Brennan, 1993).

### 7.4. Myb protein DNA binding characteristics

The c-Myb protein is a mammalian transcriptional regulator. Its DBD consists of three imperfect direct repeats, with two repeats (R2 and R3) required for sequence-specific binding. NMR analysis indicated that c-Myb repeat 3 contains a HTH motif with three α-helices (Ogata et al., 1992). Differences with the λ repressor HTH motif include an extra residue in the turn and, more importantly, a novel organization of the hydrophobic core. This core orientation alters the juxtaposition of the three α-helices. In c-Myb helix 1 and 2 are in an antiparallel arrangement similar to that of the homeodomain motif (Figure 6E).

Ogata et al. (1994) determined the three-dimensional structure of Myb motif repeat 3 (R3) and repeat 4 (R4) bound to the Myb DNA binding site. This work demonstrated a role for various conserved hydrophobic residues on the stability of the HTH structure. This hydrophobic core consists of three tryptophan residues, characteristic of various Myb-related proteins. In addition to these three residues, one per helix in the three-helix motif, a number of hydrophobic side chains are consistently present. In the known three-dimensional structures of Myb, these side chains stabilize the core, maintaining the amphipathic nature of the helices.

Myb also has an altered orientation to the DNA when compared to bacterial HTH proteins. As noted in section 7.2, c-Myb helix 3 lies more toward the DNA axis than both λ repressor and homeodomain motifs. The recognition helix lies at the floor of the major groove, with, in the case of R3, a total of five residues making base specific contacts (Ogata et al., 1994). Despite this number of contacts, each repeat alone is unable to bind DNA specifically (Tanakawa et al., 1993) and the two motifs work cooperatively to bind the Myb DNA binding site. This cooperative interaction is believed to occur via direct contact between the two repeats.
7.5. Prediction of DNA binding domain structure

One of the more daunting tasks in the field of molecular biology has been the prediction of the three dimensional structure of a protein from its primary sequence. One method for predicting tertiary structure is through comparison of identity between a known structural motif and a recently discovered protein sequence. In the case of a DNA binding domain, various techniques are used to compare a new sequence with known DNA binding families. Zinc finger-like proteins will often contain conserved cysteine and histidine residues at a particular spacing. Evenly spaced leucine residues characterize leucine zipper-like proteins. In the case of HTH proteins, the most widely used method to predict structure is by comparison of conserved residues, often hydrophobic and/or aromatic side chains involved in bringing together α-helices (Gehring et al., 1994). Figure 7 denotes various conserved amino acids from the HTH and homeodomain DNA binding motifs. As another example, the c-Myb proto-oncogene is characterized by the unique spacing of three tryptophan residues and various other conserved hydrophobic residues. These amino acids provide a hydrophobic core for interaction of the three α-helices of Myb.

Determining that a primary sequence encodes a DNA binding protein, predicting which residues make base contacts in the major groove, and predicting to which specific base a particular residue will form hydrogen bonds, are difficult tasks at best. Work by Suzuki (1994) provides a general framework for a DNA-protein recognition code. By comparing DNA-protein interactions from known data, it was found that there are preferences for recognition of specific bases by particular residues. Major groove interactions by DNA-transcription factor complexes including HNF3, Max, Gal4, E2, GCN4, Mata2 were analyzed from their crystal structure. By studying the amino acid to base preferences of all known DNA binding transcription factors, a set of general rules was determined. For example, recognition helix positions 1, 4, 5 and 8 were used for recognition by HTH proteins, but not all sites were used in all cases. Mata2 uses only sites 4, 5, and 8 for recognition, despite the fact that the residue at site 1 (Asn), is used for binding site recognition at site 1 of GCN4 (Suzuki, 1994). While there are many exceptions to his predictions, the rules permit us to make an educated guess about which amino acids are involved in hydrogen bonding to particular bases. Combined with experimental evidence about the protein's consensus binding site, it is possible to assign a probable amino acid to a particular base in its binding site sequence. For example, the arginine and lysine residues typically
A

<table>
<thead>
<tr>
<th>CAP</th>
<th>RQEQGQIVGSRESIVGRILK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lac</td>
<td>LYDVAEYAGVSYQTGTVVSN</td>
</tr>
<tr>
<td>434 Cro</td>
<td>QTELATKAGVQIQSSLIEA</td>
</tr>
<tr>
<td>λ rep</td>
<td>QESVADMDMGQSGWALFN</td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th>MATα2</th>
<th>SRIRKNVSNRRKKEKTI</th>
</tr>
</thead>
<tbody>
<tr>
<td>MATα1</td>
<td>TPLQRRVIWFINKRQSK--</td>
</tr>
<tr>
<td>oct-2</td>
<td>EKEVIRVFCNNRWEKKRIN</td>
</tr>
<tr>
<td>ubx</td>
<td>TERQIKWFNRMLKKEI</td>
</tr>
</tbody>
</table>

Figure 7. Determination of HTH and homeodomain by primary structure homology

HTH and HTH variant motifs can be distinguished by comparison of the amino acid primary sequence. (A) Alignment of HTH motif sequences. Residues generally conserved, including several hydrophobic residues and a glycine in the “Turn”, are shaded. (B) The recognition helix of four homeodomain proteins are shown. The shaded residues are typically conserved between proteins. These residues are involved in either DNA recognition or form a hydrophobic core to stabilize the HTH motif.
interact with a guanine in the major groove via a hydrogen bond, with interactions to other bases possible but not often found.

7.6. Cooperative binding of proteins to DNA

Cooperativity in protein DNA interactions can be defined as an increased affinity by a DBD for its DNA site when another DBD is bound to an adjacent site (Senear et al., 1998). A classic example is that of the bacteriophage λ cl repressor. A cl repressor homodimer binds a set of three O₅ sites, and affinity by the dimer for one site is enhanced if a dimer is already bound to an adjacent site (Ptashne, 1992).

From a biochemical viewpoint, there are known to be various instances where the binding of a dimer, A-B, to DNA is greater than the sum of its individual binding energies. In the work of Jencks (1981), this additional energetic factor was described according to the Gibbs free energy change. The Gibbs free energy changes for the binding to a protein of a molecule A-B must include both the "intrinsic binding energies" of A and B ($\Delta G^{i}_A$ and $\Delta G^{i}_B$), as well as a "connection Gibbs energy", $\Delta G^{c}$, that is derived from changes in translational and rotational entropy. In terms of DNA binding, the dimerization of a DNA binding protein from two monomers, which bind DNA at low affinity, would result in more than simply the additive effect of the two recognition helices involved. Essentially the creation of a dimer increases the local concentration of the second molecule when the first molecule is bound to its site. This is termed the chelate effect.

Changes in Gibbs free energy can result from hydrogen bonding between two DNA recognition helices. Connection Gibbs energy can be related to the loss of entropy, to the fact that the local molarity of each molecule is increased when its dimer partner is bound to DNA, and by the fact that each monomer maintains the other monomer in a specific spatial orientation. This cooperative binding of protein to DNA has been shown for a number of transcriptional activators including the MATα1/MATα2 complex (Goutte and Johnson, 1993). Binding of each protein as a monomer is in the order of $10^{-6}$ M for MATα2 and is undetectable for MATα1. When incubated with the hsg operator binding site as a dimer, the complex binds at concentrations as low as $10^{-9}$ M, a thousand-fold increase in binding affinity. Such increases in affinity are labeled as a cooperative binding event, though no direct allosteric conformation change of one subunit by the other subunit can be detected.
Another form of cooperativity through an increased “connection Gibbs energy” is the cooperative binding of two HTH monomers contained within the same primary sequence. The Oct-1 POU domain contains two DNA recognition helices, a POU-specific DBD (POUs) connected by a 24 residue linker to the POU homeodomain. Each domain contains a variant of the HTH motif as shown by crystal structure (Klemm et al., 1994). Previous work on the protein indicated that optimal binding required both subdomains, with the POU homeodomain binding with 1000 fold lower affinity when compared to full length protein (Ingraham et al., 1990).

Is this interaction comparable to the cooperative binding observed between the two subunits of a dimerized protein? In the case of the POU domain, the 24 residue linker between the two domains is flexible, with a portion of the sequence being untraceable in electron density maps. This flexibility would appear to indicate that the two subdomains interact as if separate molecules, with the exception that binding of one subdomain to the DNA will increase the local concentration of the other subdomain to its binding site. Despite the absence of stable protein-protein contacts, there are contacts by each subunit to overlapping phosphates in the DNA backbone and one overlapping interaction through a base pair. Klemm et al. (1994) speculate that cooperation in binding between the two POU subdomains results from either a chelate effect in line with the theories of Jencks (1981), or that thermodynamic coupling may result from the overlapping phosphate contacts.

That the two domains behave as monomers was confirmed in the work of Klemm and Pabo, 1996. They expressed the POU domain as two monomers, POUs and POU homeodomain, by removing the linker region. While POUs alone was unable to bind its DNA site, it was able to form a complex in the presence of the POU homeodomain even without a linker between the two motifs. The resulting $\Delta G_{\text{cooperativity}}$ was calculated as $-1.6 \text{ kcal/mol}$, indicating a favourable cooperative interaction. This cooperativity points to thermodynamic coupling from overlapping phosphates, as well as overlapping contacts at base pair 5 (POUs via the major groove, POU homeodomain via the minor groove).

However, this cooperative interaction does not negate the much larger effect from increased local concentration. Cooperative binding of the two subdomains when “linker-less” results in a $K_d$ of $1.7 \times 10^{-6}$ M for POUs and $1.5 \times 10^{-7}$ M for POU homeodomain. In contrast the wildtype domain containing the flexible linker between the two subdomains results in a $K_d$ of
Another example of cooperative DNA binding occurs between DNA binding motifs repeat two (R2) and repeat three (R3) of the Myb proto-oncogene. Neither R2 nor R3 alone is able to bind DNA specifically (Tanikawa et al., 1993), suggesting a cooperative interaction. Ogata et al. (1994) speculate that this cooperativity is induced by salt bridges between residues on R2 and on R3, hydrogen bonding between Glu-132 of R2 and Asn-179 of R3, as well as the interaction of amino acids from R2 and R3 with overlapping phosphates. In the free DBD, unbound to DNA, R2 is much less stable than R3. When bound to DNA, R2 and R3 are both stable structures, suggesting that conformation changes occur in R2 and R3 during binding.

In the various cooperative protein-DNA associations described, the exact mechanism of cooperative interactions has not been deduced. Cooperativity, reflected as $\Delta G_{\text{cooperativity}}$, can be provided by a number of factors, including protein-protein interactions, interactions between proteins mediated by the DNA, DNA conformation changes and protein mediated allostery.

In studies of DNA binding where two subdomains or two proteins are required for high affinity binding, it can be assumed that the two components work cooperatively if each component alone is unable to bind with affinities that are additive to the binding affinity of the dimer. This increased affinity reflects both $\Delta G_{\text{cooperativity}}$ as well as the connection free energy ($\Delta G^c$) from increased local concentration.

However, the contribution from the connection free energy does not always increase protein-DNA affinity as expected. In the work of Senear et al. (1998), the connection energy required for a $\lambda$ repressor dimer to bind DNA is actually positive ($\Delta G^c$ is +4 to +5 Kcal/mol reflecting reduced affinity). Binding to one site on the $\Omega_R$ interferes with the potential interaction with a neighboring site. This reduced affinity may be due to improper orientation of the two DNA binding domains to the DNA when they are dimerized. Senear et al. (1998) speculate that this interference is used to moderate the effect of cooperativity, allowing the protein to function as a sensitive genetic switch rather than always binding to the $\Omega_R$ site as a dimer.

8. CRYSTAL STRUCTURE OF THE RAP1 DBD - IMPLICATIONS FOR DNA BINDING

During the course of this work, the crystal structure of the Rap1 DBD bound to a
Introduction

consensus telomeric site was determined to 2.25 angstrom (Å) resolution (Konig et al., 1996). This model (shown in Figure 8) presents evidence that Rap1 binds to DNA via a pair of HTH motifs. Figure 8B displays the various subdomains of the DBD in the same orientation as the DBD shown in 8A. HTH1, with its recognition helix lying across the major groove is shown at left. Also interacting in this major groove is the "tail" region, positioned in front of HTH1. HTH2 is twisted roughly 90° backward from HTH1, lying in the major groove at the right of the model in Figure 8A. The loop region, comprising the residues extending upwards to the right of HTH1 in Figure 8A, forms a stable structure visible in electron density maps.

The DBD was shown to have structural similarity to the Myb HTH and the Engrailed homeodomain. Despite the ability to superimpose Myb and Engrailed over the two Rap1 HTH motifs, there is only weak identity between their primary sequences. The two Rap1 HTH domains are positioned in a tandem orientation, each binding to a similar 5'-ACACC-3' sequence. This tandem array is similar to that of the Myb motif, though the distance between binding sites is increased. Another similarity to Myb is an N-terminal minor groove contact on both HTH1 and HTH2. These contacts, through residue Lys-360 for HTH1 and Lys-446 for HTH2, are known to stabilize DNA binding of other HTH variants, and are likely to play a role in Rap1 DNA binding.

In addition to the HTH motifs, the crystal structure suggests that the Rap1 DBD contains a novel "tail" region, composed of proline and glycine residues involved in base-specific interactions via the major groove. These unique contacts, which allow the DBD to envelop the Rap1 binding site, could potentially be artifacts of the crystallization process. Because the "tail" region is preceded by a random loop that is not visible in the electron density map, it is possible that the tail region extends from a different Rap1 molecule in the crystal (Konig et al., 1996). Despite this uncertainty, the "tail" region is also known to be essential for binding to a PGK promoter Z+ fragment (Henry et al., 1990), suggesting it does interact with the DNA in some manner.

Other unusual aspects of the DBD I have inferred from the crystal structure include the existence of various regions of structural disorder between the helices of HTH2; between HTH2 and the tail region; and a highly structured "loop" region between HTH1 and HTH2. These areas explain why the DBD of Rap1 is significantly larger than other HTH binding domains including the Myb DBD (235 amino acids for Rap1, compared to 103 amino acids for Myb.
Figure 8. Three dimensional crystal structure of DBD and ribbon diagram of DBD sub-domains

The crystal structure of the Rap1 DNA binding domain bound to a telomeric site was provided by Daniela Rhodes (Konig et al., 1996), and was viewed using the SETOR program. (A) side view of DBD bound to a telomeric site. The N-terminus is shown at the far left, the C-terminus is located just to the right of the N-terminus. (B) DBD side view. HTH1 and HTH2 are shown in yellow. The DBD “loop” region and the “tail” region are shown in red.
R2R3). It also suggests that the “loop” region, which is conserved between *S. cerevisiae* and *K. lactis* homologs, may have a role in the binding of Rap1 to DNA. This region contains two structurally similar omega turns (residues 422-427 and 430-435) which may play a role in its stability. The “loop”, as a structured linker between the two DNA binding motifs, aids in the tandem orientation of the two HTH domains. These two domains do not form a protein-protein interface, so their distance apart must be determined by constraints imposed by the “loop”. This would aid in the binding of Rap1 to telomeres, which, due to the nature of telomerase activity, are tandemly repeated sequences.

Other results suggest a role for the two HTH domains. In the work of Indrissi et al. (1998) binding of Rap1 to telomere and ribosomal protein sites were distinguishable by chemical probes. Further studies (Indrissi and Pina, 1999) related these differences to the two half sites (each being ACAYYY where Y=T or C) bound by each Rap1 DBD HTH domain. In the consensus telomere site Y=C, whereas in the UAS$^g$ site Y=T. Indrissi and Pina (1999) note that a UAS$^g$ site led to higher levels of gene activation under identical circumstances to that for a telomere site. Their data suggest that the differences in binding sites induce structural changes in Rap1, and that this has an allosteric effect on activation by Rap1. When Rap1 binds to a telomere site, its activation domain is hidden, while binding to UAS$^g$ reveals the activation domain. In their model, Rap1 is able to play different roles in the genome by changing its conformation depending on the DNA site it is bound to. Consistent with this, *in vivo* telomere sites and silencer sites contain a higher proportion of C at the second half site, while ribosomal protein genes and glycolytic-enzyme genes contain a higher proportion of T (Indrissi and Pina, 1999).

9. THE PRESENT STUDY

Genetic and biochemical studies of telomeres, silencers, and gene promoters have provided a model whereby Rap1 DNA binding helps to anchor a wide range of gene products involved in heterochromatin formation or gene activation. Rap1 binds telomere ends, forming a heterochromatin state. This same molecule binds at UAS$^g$ sites to activate transcription, and binds at HM loci to silence gene transcription. The crystal structure of the Rap1 DBD bound to a telomeric site has provided an opportunity to compare what we know about Rap1 interactions
in vivo and its DNA binding function. Rap1 contains a novel bipartite DNA binding domain. The DBD is composed of two Myb-like HTH motifs and an extended region between the two motifs. The structure also contains a c-terminal region involved in further protein-DNA contacts.

These results provide key questions regarding Rap1 function. With the molecule binding such functionally disparate sites, is a region of the protein involved in determining the function for each site? Could the variable nature of the DNA binding site play a role in Rap1 action at telomeric versus silencer or activation sequences? One key question is whether differences in Rap1 action at telomeres and silencers is a result of the DBD as opposed to other regions of Rap1, and whether this effect is related to novel aspects of the DBD, as determined by crystal structure analysis.

The work presented in this thesis will outline the determination of key residues of the Rap1 DBD by primary sequence analysis. This is followed by studying the effect of key mutations in the DBD HTH1 recognition helix on binding affinity for telomeric and silencer binding sites, both in vivo and in vitro. The results indicate that wildtype Rap1 binds various binding sites at the same affinity in vitro, but mutations in the Rap1 DBD, and in vivo studies suggest Rap1 actually has higher affinity for telomeric sites. Finally, a brief analysis of Rap1 phosphorylation and DNA binding affinity will be presented.
II. MATERIALS AND METHODS

1. STRAINS, PLASMIDS AND OLIGONUCLEOTIDES

1.1. Strains used

Plasmid manipulations used E. coli strain DH5α (recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 ΔlacU169 (φ80 lacZΔM15)), grown in TB medium with 100 µg/ml ampicillin. Protein expression of Rap1, DBD and their respective R404K, R404D, R408N and R408D mutants was performed using the BL21(λDE3) strain (FompT[on] hsdS (rB·mB·; an E. coli B strain) with DE3, a lambda prophage carrying the T7 RNA polymerase gene), grown in M9ZM medium with 100 µg/ml ampicillin. This strain allows T7 transcription to take place from plasmids containing a T7 promoter in front of the RAP1 gene.

All other protein expression utilized the E. coli strain TGI (F' traD36 lacIQΔ(lacZ)M15 proA+B+supE Δ(hsdM-mcrB)5 (rB·mB·mcrB· thiΔ(lac-proAB) ) grown in 2xYT medium with 100 µg/ml ampicillin. This strain is used for the recA+ system of protein expression.

For site-directed mutagenesis, the strain NM522 mutS was utilized. This strain (thi, supE, Δ(hsdMS-mcrB)5, Δ(lac-proAB) [mutS::Tn10] F' [proAB+ lacI+ lacZ ΔM15]) is unable to repair base mismatches in plasmids containing mutant and wild-type sequence on opposite strands. This allows the plasmid to be replicated in NM522 mutS strains, forming one wild-type and one mutant replicate. If the mismatched plasmid were transformed into another strain, such as DH5α, the mismatch would be repaired.

1.2. Handling of strains and general techniques

All strains were initially isolated as single colonies on LB plates and grown overnight at 37°C with moderate shaking in LB medium (M9ZM medium in the case of BL21(λDE3)). The cultures were then adjusted to 15% glycerol with sterile 50% glycerol, and frozen at -70°C as stock.

Subsequent insertion of plasmids was performed by making the strains electrocompetent. A culture grown overnight in SOB medium was inoculated into 500 ml of SOB medium and grown at 37°C with shaking to an A_{600} of 1.0. The culture was then centrifuged for 10 minutes at
4,000x g in a GSA rotor, supernatant removed, and the pellet gently resuspended in 500 ml of ice cold deionized distilled sterile water containing 10% glycerol by volume. After centrifugation at 4,000x g, the supernatant was gently discarded, leaving a small amount of liquid (around 2 ml) which was then used to resuspend the cell pellet. This resuspended cell solution was then aliquoted into microfuge tubes and quick-frozen in a dry ice ethanol bath. The tubes were then stored at -70°C for later use.

In the case of the NM522 mutS strain, cells were made competent by the calcium chloride method, as outlined in the USE mutagenesis kit protocol (Pharmacia).

1.3. Plasmids

1.3.1. Rapl and DBD expression plasmids

Two systems were used for the expression of full length Rapl, DBD and their various site-directed mutants. Initial expression of Rapl, DBD and their R404K, R404D, R408N and R408D mutants involved the T7 gene 10 expression plasmids pL3S5 (Gilson et al., 1993) for full length Rapl and pRBD (Ho et al., 1994) for the Rapl DBD. Both of these plasmids express the first 11 amino acids of the T7 phage gene 10 which is fused to all but the 19 amino-terminal residues for full length Rapl. For pRBD, the first 11 amino acids of gene 10 are linked by leucine and glutamine to residues 361 to 598 of Rapl, with residues serine, glycine, and cysteine comprising the C-terminus of the expressed protein.

A second set of expression plasmids, containing wild-type Rapl or the Rapl DBD (residues 353 to 598) was utilized for most Rapl studies discussed in this paper. The plasmid pRG50 (His 6 tag), a gift of Dr. Daniela Rhodes, expresses full length Rapl with a 6 His-tag using the recA promoter fused to the T7 gene 10 ribosome binding site. This system, unlike pL3S5, does not express any residues from gene 10, but instead expresses the full RAP1 gene combined with a 10 residue C-terminus containing 6 histidine residues linked to Rapl by the residues aspartate, leucine, leucine and glycine. For the DBD, the plasmid pRG50 (DBD) expresses residues 353 to 598 of Rapl, with no additional residues (Konig et al., 1996; personal communication).
Materials and Methods

1.3.2. *CAT* assay plasmids

Expression of Rap1 in Cos-7 cells involved the introduction of the RAP1 gene into the pMV1 vector, a gift of Dr. Ivan Sadowski (Sadowski, submitted). This plasmid contains an SV40 early promoter regulating the expression of the VP16 activation domain linked via a multiple cloning site to the SV40 poly A region. PCR fragments containing various wild-type and mutant RAP1 sequences were ligated into the multiple cloning site, producing VP16/RAP1 hybrid genes.

The CAT reporter gene was derived from pBCAT, a gift of Dr. Ivan Sadowski. pBCAT contains an E1B TATA box in front of a chloramphenicol acetyl transferase gene. Rap1 reporter constructs were made by inserting five Rap1 binding sites in the upstream activating region. Briefly, oligonucleotides containing either telomeric or silencer binding sites (see Figure 9) were phosphorylated and self-ligated under optimal conditions to produce linear fragments containing 5 copies of each binding site. These five copy fragments were isolated from other ligation products by agarose gel electrophoresis and were inserted into a pBluescript II KS (-) vector cut with EcoRI and HindIII. The resulting clones were then digested with XhoI, PstI and the binding site fragment inserted into a XhoI, PstI-digested pBCAT vector.

1.3.3. *HTH1* expression plasmid

For expressing HTH1 (residues 359 to 415) of Rap1, PCR was performed on the full length RAP1 gene (plasmid pL3S5) using oligonucleotides designed for the expression of amino acids 359 to 415 of Rap1, with a C-terminal insertion of residues glycine, isoleucine and histidine. This fragment was inserted between the NdeI and HindIII sites of the pET-22b (+) plasmid, allowing expression of HTH1 from a T7 promoter.

1.4. General techniques for plasmids

1.4.1. Plasmid preparation techniques

Preparation of plasmid DNA from *E. coli* cells involved use of one of the following alkaline lysis techniques. For most cloning techniques, the Qiaprep system (Qiagen) was used for the preparation of small amounts of plasmid from overnight cultures grown in TB medium with 100μg/ml ampicillin. Occasionally, the Wizard Prep system (Promega) was used.
Materials and Methods

For the large scale preparation of plasmid DNA, especially to produce large quantities of plasmid DNA for the CAT assay procedure, a variation of the LiCl DNA purification protocol was used. Briefly, 50 ml of cells were grown overnight at 37°C with moderate shaking in TB ampicillin medium. The cells were pelleted and resuspended in 2 ml solution I (Maniatis et al., 1982). 4 ml of Solution II was added, the tubes gently mixed and left on ice for 5 minutes. 3 ml of Solution III was added, the tubes shaken and left on ice for a further 5 minutes. Tubes were then centrifuged at 14,500x g in an SS-34 Rotor for 10 minutes at 4°C. The supernatant was removed to a fresh tube, 20 ml of EtOH added, the tube vortexed for 45 seconds and left at room temperature for 5 minutes. After centrifuging for 5 minutes at 14,500x g in an SS-34 rotor, the supernatant was removed and the pellet allowed to air dry briefly. The pellet was then resuspended in 750 µl of H₂O after which 750 µl of 5M LiCl was added and the tube left at -20°C for 10 minutes. The tube was then centrifuged at 9,000x g in a microfuge for 10 minutes at 4°C to pellet much of the RNA, and the supernatant transferred to a new microfuge tube. 45 µl of 0.5M NaCl and 3 ml of EtOH were added, and the tubes left at -70°C for 10 minutes. The tubes were then centrifuged at 9,000x g in a microfuge for 10 minutes at 4°C, the supernatant aspirated, and the pellet washed once with 70% EtOH. The pellet was dried down and resuspended in a small volume of 10 mM Tris, pH 7.5. The concentration of the plasmid DNA was determined by separating 1-4 µl of the plasmid on an agarose gel beside a known quantity of the original plasmid and a known quantity of 1 kb ladder marker DNA (Gibco-BRL). These preparations contain both plasmid DNA as well as tRNA which is not precipitated by LiCl.

1.4.2. Agarose gel electrophoresis

During the cloning, restriction digestion and quantitation of plasmids, agarose gel electrophoresis was frequently employed. In most circumstances, the plasmid of interest was digested by one or more restriction enzymes, and separated in an agarose gel at a concentration appropriate for the molecular weight of the plasmid molecule. A size marker lane, containing either a 1 kb DNA ladder or a 100 bp DNA ladder (Gibco-BRL) was also run.

Each 50 ml gel contained 8 µl of a 1 mg/ml ethidium bromide solution (Maniatis et al., 1982), and had a total concentration of 1x TBE. 2% agarose was used for DNA of 100 to 600 base pairs, 1% agarose for DNA from 500 base pair (bp) to 2 kilo base pair (kb), and 0.6% agarose for DNA over 2 kb in length. For cloning protocols requiring ligation, DNA was
Materials and Methods

digested and separated on an agarose gel. The DNA was visualized under UV light. The appropriate band was excised with a razor, and purified from the agarose using the QIAquick DNA isolation kit (QIAGEN).

1.5. Oligonucleotides

The various oligodeoxyribonucleotides (oligonucleotides) used in this thesis are shown in Figure 9. Rap1 DNA binding oligonucleotides were designed to contain the appropriate site, flanked by overhanging EcoRI, HindIII sites. These overhangs allow subcloning of the oligonucleotide for the CAT assay protocol.

Oligonucleotides for site-directed mutagenesis contain one or more base mismatches from the wild-type gene sequence, flanked by at least 12 bases of Rap1 sequence on either side. Mismatches for R404K, R404D, R408N and R408D destroy a BanI restriction site, and loss of this site was used to identify mutants. The oligonucleotide for K360A destroys a HindIII site, the K446A oligonucleotide creates an NruI site, the G590V oligonucleotide adds an HpaI site, the Δ418-438 mutant oligonucleotide removes a HincII site, and the Δ588-594 oligonucleotide removes a BgII site. These various alterations were used to identify the mutants during final screening.

PCR primers for Rap1 cloning contain the start and end sites of the desired Rap1 sequence, in addition to restriction sites for ligation into the appropriate plasmid. Additional bases are included after restriction sites to ensure proper restriction digestion of the PCR products.

2. PLASMID MANIPULATIONS

2.1. Cloning techniques

2.1.1. PCR

All PCR used the Vent DNA Polymerase enzyme (New England Biolabs, NEB), and followed standard procedures. Each 100 μl reaction contained 1x Vent Buffer (NEB), 100 pmol of each primer, 1-10 ng of template, 1 μl of 100x BSA (NEB), 100 nM MgSO4, and 1 μl of Vent Polymerase. For cloning of CAT assay expression plasmids, PCR was performed with 1 ng of
### Materials and Methods

**Description** | **Sequence**
--- | ---
**Rap1 DNA Binding** |  
Telomeric site | 5'-AATTCACCCCCACACCCGGTGA-3'  
| | 3'-GTGTGGGTGCTGTCGAGATTCGA-5'  
HMR-E site | 5'-AATTTCAAAAAACACATCAACCTA  
| | 3'-GTTTTGGGTGCTGTCGAGATTCGA-5'  
HML-E site | 5'-AATTTCAAAACCCATACCA-3'  
| | 3'-GTTTTGGGTGCTGTCGAGATTCGA-5'  
**TelomericA3 site** | 5'-AATTTCAAAAGCCCACCCCTGA-3'  
| | 3'-GTTTTGGGTGCTGTCGAGATTCGA-5'  
**HMR-El0/C11 site** | 5'-AATTTCAAAAACCCATACCA-3'  
| | 3'-GTTTTGGGTGCTGTCGAGATTCGA-5'  
**Site Directed Mutagenesis** |  
R404K | 5'-GACCTCTAATTCGGTCTTATAGAATTACCGG-3'  
R404D | 5'-GACCTCTAATTCGGTCTTATAGAATTACCGG-3'  
R408N | 5'-GAAAGATAGACATTTTCTGCTTATAG-3'  
R408D | 5'-GAAAGATAGACATTTTCTGCTTATAG-3'  
K360A | 5'-CTGTAAAAGAAGCTGCTTGTGGGAGGG-3'  
K446A | 5'-CTGTAAGAAAATTCCCTGCGGATTGATGGTGCG-3'  
G590V | 5'-GGTCAAGAACATTTTCTGCGGATTGATGGTGCG-3'  
Δ418-438 | 5'-GATGGTGGCAAAACCTTTGCGTACTCTAGTCTTATAG-3'  
Δ588-594 | 5'-CTTGGCCCTTCTTGGCGGGAAGCCAGCCAGCC-3'  
**PCR Primers for Cloning** |  
Rap1 HTH Motif (5'-GGGGGGGGGGGGGGGGTAATATGAATAAAGCTTTTTCTTTTACAG-3'  
| 5'-GGGGGGGGGGGGGGTAATATGAATAAAGCTTTTTCTTTTACAG-3')  
VP16-Rap1 Hybrid | 5'-CTAGAAGCTTTCCATACAGGCTTCTCCTTTCTTTGAGATAG-3'  
| 5'-CTAGAAGCTTTCCATACAGGCTTCTCCTTTCTTTGAGATAG-3')  
VP16-DBD Hybrid | 5'-CTAGAAGCTTTCCATACAGGCTTCTCCTTTCTTTGAGATAG-3'  
| 5'-CTAGAAGCTTTCCATACAGGCTTCTCCTTTCTTTGAGATAG-3')

**Figure 9. Oligonucleotides used in present study**

A listing of oligonucleotides used for the various experiments performed. Rap1 DNA binding sites shown were ligated and cloned into the pBCAT plasmid for CAT assays of Rap1 activity. They were also used for mobility shift assays of Rap1 protein. PCR primers contain an overhang region and a unique restriction site to aid in cloning.
Materials and Methods

DNA from plasmids containing various full-length Rap1 mutants derived from plasmid pRG50 (His tag). For HTH1 PCR, 10 ng of pL3S5 plasmid was used.

2.1.2. DNA restriction digests

All restriction digests followed standard procedures. For digestion of DNA with two or more restriction enzymes, One-Phor-All buffer (Pharmacia) was used at the highest concentration recommended for the various enzymes. Incubations typically were at 37°C for one hour.

2.1.3. Ligation

Subcloning of inserts into the vector of interest was accomplished using T4 DNA Ligase (New England Biolabs). 100 ng of vector was incubated at 16°C overnight with a 4-fold molar excess of insert in a 1x ligase buffer (NEB). Because all insert/vector ligations involved two incompatible ends, dephosphorylation of the vector was not required.

For the ligation of Rap1 DNA binding site oligonucleotides to produce CAT expression plasmids, the oligonucleotides were first phosphorylated under standard conditions (Maniatis et al., 1982). 150 pmol (2 µg), was incubated with 1 mM ATP and 1 unit of T4 Polynucleotide kinase (Pharmacia) in 1x One-Phor-All buffer (Pharmacia) in a total volume of 30 µl. The phosphorylated oligonucleotides were then ligated in a total volume of 20 µl for one hour at room temperature. EDTA at a final concentration of 10mM was added to stop the reaction, and the mixture was separated on a 2% agarose gel to resolve the fragment containing five Rap1 binding sites.

2.1.4. DNA concentration estimation

For most cloning techniques, DNA concentration was determined by comparing the intensity of the band of interest with a known quantity of DNA size marker (1 Kb ladder, GIBCO-BRL)

2.2. Site-directed mutagenesis

All site-directed mutants were produced using the unique site elimination (USE) method (Deng and Nickoloff, 1992). The mutagenesis procedure and reagents were obtained in the
Materials and Methods

Pharmacia USE mutagenesis kit. In addition to the reagents provided with the kit, two selection primers were also obtained from Pharmacia which allow the mutation of a unique Scal site within the Amp resistance gene of the various plasmids used for this procedure. These primers toggle between a Scal and MluI restriction site at this location, allowing multiple mutations to be performed. Mutagenesis involved the annealing of 0.025 pmol of plasmid DNA, 1.25 pmol of selection primer, and 1.25 pmol of the target mutagenic primer (see Table 1) in 1x One-Phor-All buffer (Pharmacia) at a total volume of 20μl. This mixture was incubated at 100°C in water under a rolling boil for 5 minutes, and immediately chilled on ice for 5 minutes. After a brief centrifugation, a USE nucleotide mix and reaction mix (Pharmacia) were added, and the tube incubated at 37°C for one hour. This procedure allows extension of the mutagenic and selection primer and ligation to form a complete circularized double-stranded plasmid with one strand containing the wild-type sequence, the other containing both the mutant selection primer and the mutant Rap1 sequence. After this reaction, the mix was completely digested with Scal enzyme to linearize all plasmid containing a double-stranded wild-type Scal site. This reaction mix was then transformed into competent NM522 cells. After growth overnight in TB ampicillin medium, plasmid DNA was isolated and 200 ng was completely digested by Scal. The resulting digest was then electroporated into DH5α cells and grown overnight on LB ampicillin plates. Plasmid DNA from single colonies was isolated. Typically 50-90% of colonies selected contained both a mutated Scal site and the desired mutation, as determined by restriction site analysis.

Mutagenesis was performed directly on the plasmids pL3S5, pRBD, pRG50 (His tag) and pRG50 (DBD). To aid in the selection of mutants, site-directed mutations were designed such that an easily monitored restriction site was added to or removed from the sequence. The presence of the desired mutation could then be confirmed by a restriction digest. In the case of Arg-404 and Arg-408 mutants, mutations were also confirmed by DNA sequencing. Mutants include K360A, R404K, R404D, R408N, R408D, K446A, A418-438, G590V, and Δ588-594. All mutants except Δ588-594 were created for both full length Rap1 and the DNA binding domain.
2.3. Sequencing

Mutants R404K, R404D, R408N and R408D were initially selected by the loss of a BanI restriction site. To confirm the mutations, the double-stranded plasmid was sequenced by the dideoxy chain termination method using a Sequenase kit (Pharmacia). Sequencing was performed using the manufacturer's protocol with a T7 promoter primer for DNA priming on the pRBD plasmid and $^{35}$S-CTP as label. Sequence reactions were separated on a denaturing polyacrylamide gel, dried and exposed to film. The sequences of wild-type DBD and the mutants were as expected in all instances.

3. PROTEIN EXPRESSION AND PURIFICATION

3.1. General techniques for proteins

3.1.1. SDS-PAGE analysis of protein

SDS-PAGE analysis of Rap1 followed standard procedures (Maniatis et al., 1982). Full length Rap1 was separated on a 10% gel, and migrated between 100 and 120 kDa. Rap1 DBD protein was run on a 10% or 15% gel, and migrated at 30 kDa. HTH1 was separated on a 15% gel, and migrated at 8 kDa, near the bottom of the gel.

3.1.2. Western blot analysis

Proteins separated by SDS-PAGE were electro-blotted to nitrocellulose membrane. The membrane was then blocked with milk protein in PBS buffer with 0.05% Tween 20, washed, and incubated with a 1:1,000 dilution of Rap1 antiserum in 5% milk protein 0.05% Tween 20 with rocking at room temperature for 1 hour. Membranes were washed under standard conditions, and then incubated in secondary antibody (1:5,000 dilution of peroxidase conjugated anti-rabbit antibody, Amersham) for 30 minutes at room temperature, followed by washing. Protein-antibody complexes were detected using an Enhanced chemiluminescence western blotting kit (Amersham).
3.1.3. Raising polyclonal antibodies against Rap1 protein

A New Zealand White rabbit was immunized with 130 μg of purified Rap1 protein expressed from plasmid pL3S5. The non-denatured protein, in 500 μl buffer, was mixed with an equal volume of Freund’s complete adjuvant, and injected into the rabbit. The animal was boosted with 60 μg of protein once after three weeks. Blood samples were taken and allowed to clot at 37°C for four hours. Serum was then collected and stored in aliquots at -70°C. Western blot analysis of total yeast cell extract, using a 1:5000 dilution of Rap1 antibody, revealed a single band at 100 kDa, corresponding to the Rap1 protein.

3.2. Protein expression

3.2.1. BL21(λDE3) expression of Rap1 and derivatives

Expression of pL3S5 and pRBD derivatives, including wild-type Rap1 and mutants R404K, R404D, R408N and R408D was performed using the T7 gene 10 promoter with fusion of Rap1 to the first 11 amino acids of the gene 10 product (Ho et al., 1994). Before transformation of the pL3S5 or other plasmid into the BL21(λDE3) strain, the cell was transformed with the pLysS plasmid (Ho et al., 1994), and transformants selected with 25 μg/ml chloramphenicol. This plasmid allows the production of T7 lysozyme, an inhibitor of gene 10 transcription. This reduces any toxic effect of low level Rap1 production. The plasmid pLysS also aids in lysis of cells after IPTG induction. Freeze/thawing of cells after IPTG induction breaks the cell membrane, allowing lysozyme to digest the surrounding cell wall. Induction of Rap1 transcription is initiated by the addition of IPTG to a final concentration of 1 mM.

The procedure for Rap1 expression was as follows. 100 ml of M9ZM-Amp-chloramphenicol medium was inoculated with 1 ml overnight culture of BL21(λDE3) cells containing the pLysS and pL3S5 or pRBD plasmids. After growth in a 37°C rotary shaker to an OD600 of 0.6, IPTG was added to a final concentration of 1 mM. The cells were grown for a further one and three quarter hours, allowing maximum expression of Rap1 protein as measured by western blotting.
3.2.2. pRG50 expression

Expression of all Rap1 mutants was also performed utilizing the recA expression system. Plasmids pRG50 (his tag), pRG50 (DBD) and their mutant derivatives express protein using a recA promoter and a phage T7 gene 10 leader ribosome binding site. Addition of nalidixic acid to the growing cells inhibits DNA Gyrase, thereby inducing the recA promoter and leading to high levels of protein expression.

In order to express proteins from the pRG50 plasmid, TGI cells containing the relevant plasmid were grown overnight at 37°C in 2x YT with ampicillin. 100 μl of the overnight culture was added to 3 ml 2x YT ampicillin medium and grown at 37°C to an OD₆₀₀ of 1.0. 3 μl of a 500 mg/ml solution of nalidixic acid was added and cells grown overnight at 37°C, to obtain maximal levels of protein expression.

3.2.3. pET22b(+) expression of HTH1

Expression of HTH1 followed the procedure for expression using BL21(λDE3) cells. After induction by IPTG, cells were incubated with shaking for one hour forty five minutes to allow maximum expression.

3.3. Protein purification

3.3.1. Mono Q/S purification

Derivatives of Rap1, expressed in BL21(λDE3) cells, were purified from crude extracts following a two step procedure. After IPTG induction, 100 ml of culture were pelleted at 7,700x g in a SS-34 rotor. The pellet was then resuspended in 15 ml lysis buffer (50 mM sodium phosphate, pH 6.0, 100 mM NaCl), pelleted again and suspended in 5 ml lysis buffer containing 0.1% (v/v) NP-40, 0.3 μg/ml leupeptin, 300 μg/ml benzamidine, 2 μg/ml pepstatin A and 0.5 mM PMSF. The cells were quick frozen in a dry ice/ethanol bath and thawed quickly in a 37°C water bath, then placed on ice. 5 mg of NaDOC and 50 μl of 20% Triton X-100 were added, and the lysate sonicated 3x 10 seconds on ice at a medium setting (dial 3-4). After centrifugation at maximum speed in a microcentrifuge to clear the lysate, the supernatant was adjusted to 10% glycerol, 0.5 mM DTT.
Materials and Methods

This cell lysate was incubated as a batch with 10ml Heparin Sepharose by gentle rocking for 45 minutes at 4°C. The Heparin Sepharose mix was then used to pack a column and washed with 10 column volumes of lysis buffer containing 50 mM sodium phosphate (pH 6.0), 100 mM NaCl, 10% glycerol, 0.5 mM DTT and 0.1% NP-40. After washing, proteins bound to the heparin sepharose matrix were eluted with lysis buffer containing from 0.2 to 0.9 M NaCl, applied in 2 ml aliquots. Full length Rap1 eluted between 0.4 and 0.8 M NaCl, while DBD fractions eluted between 0.5 and 0.7 M NaCl. Aliquots from each fraction were then separated on an SDS-PAGE gel and detected by western blotting with Rap1 antiserum to select peak fractions. Peak fractions were then combined and dialyzed overnight in lysis buffer.

The second step in Rap1 purification involved passing the protein through a Mono Q sepharose column in a gradient from 0.1 to 1 M NaCl by FPLC. Again lysis buffer containing 50 mM sodium phosphate (pH 6.0), 100 mM NaCl and 10% glycerol was used. Dialyzed extracts were applied to the column, and 2 ml fractions collected. Rap1 protein eluted at between 0.6 and 0.7 M NaCl. Fractions were assayed by SDS-PAGE and western blotting. Peak fractions were combined. Their purity was estimated by be from 30-90% homogeneity. In the case of pRBD derivatives, second step purification was performed on a Mono S, rather than Mono Q, sepharose column. Wild-type and mutant DBD eluted between 0.45 and 0.50 M NaCl. Total protein concentration was determined by OD$_{280}$. The DBD protein concentration was determined by multiplying the total protein concentration by the proteins percent purity. Concentration of active DBD protein was determined (see section 4.2) to quantify the actual concentration of protein in each sample.

3.3.2. Preparation of Crude Extracts From E. coli

In the case of pRG50 derived Rap1, experiments were performed using crude extracts of the protein. The procedure was as follows. The day after induction by nalidixic acid, cells were centrifuged at low speed and washed twice with cold M9 medium. The pellet was resuspended in 1 ml cold lysis buffer containing 40 mM MES-KOH (pH 6.0), 100 mM KCl, 2 mM DTT, 1 mM EDTA, 1 mM Benzamidine, 0.5 mM PMSF and 10% glycerol. Cells were kept on ice and sonicated at moderate levels (3-4) a total of three times for 20 seconds using a fine probe. The lysate was cleared by centrifugation for 20 minutes at 22,000x g in a Beckman TL-100
ultracentrifuge using a fixed angle rotor. The supernatant was collected and used immediately or stored as aliquots at -70°C.

4. MOBILITY SHIFT TECHNIQUES

4.1. General techniques

All mobility shifts were performed with the following final buffer conditions: 20 mM Tris (pH 8.0), 20 mM NaCl, 10% glycerol, 1 mM MgCl₂, 0.04 mg/ml BSA, and 1 mM DTT. Radio-labeled DNA, poly dIdC and protein concentrations are as indicated in figure legends, in a total volume of 10 µl. Three probes were typically used in the various experiments: a consensus telomeric Rap1 binding site, the Rap1 binding site from the HMR-E silencer, and the Rap1 binding site from the HML-E silencer. The sequence of each oligonucleotide is listed in Figure 9.

To radiolabel DNA probes, 10 pmol duplex oligonucleotide was incubated with excess γ³²P-ATP and 5 units of T4 polynucleotide kinase (Pharmacia) in kinase buffer and incubated at 37°C for one hour. The probe was then separated from free nucleotide by C-18 reverse phase chromatography in a SEP-PAK column, eluted in 1 ml 30% acetonitrile, dried under vacuum and resuspended in water to a concentration of 0.2 pmol/µl.

Mobility shift reactions were prepared on ice, then incubated for 10 minutes at room temperature before loading on a 5% non-denaturing polyacrylamide gel using 1x TBE buffer. Polyacrylamide gels were pre-run for 30 minutes before loading samples. Samples were gently loaded into wells, and a marker lane containing a small sample of DNA loading buffer was run in a spare lane to measure migration. Gels were run at room temperature at 200 volts for 1-2 hours, then dried and exposed to a Phosphor Imager screen for 3 hours to 3 days depending on the signal strength. The intensity of specific bands was analyzed using a PhosphorImager and the IP Lab gel analysis software program.

4.2. Kₐ determination

Determination of dissociation constant (Kₐ) values were as described in Jonsen et al. (1996). Briefly, full length protein was expressed and either purified to over 50% purity (in the case of wild-type Rap1, R404K, R404D, R408N and R408D) or prepared as a crude lysate.
4.2.1. Determination of active protein concentration

The first step in $K_d$ determination was to calculate the concentration of active protein in each sample. This removes error associated with denatured Rap1 protein and other protein in the samples. The procedure uses excess DNA to drive binding, where the saturation point of binding equals the active protein concentration. Protein activity was determined by combining 0.1 or 1 μl of protein extract with radio-labeled telomeric DNA ranging in concentration from $10^{-11}$ to $2 \times 10^{-8}$ M. These mixtures were then subjected to mobility shift electrophoresis. All mobility shifts were analyzed using the PhosphorImager system.

The concentration of active protein was determined as the maximal concentration of protein able to bind duplex oligonucleotide [PD] over the range of total duplex oligonucleotide concentrations [Dt]. Using an estimated concentration of active protein (around $10^{-8}$ to $10^{-10}$M), two sets of reaction mixtures were assembled: one with increasing labeled duplex and the same [protein] and one as a control with increasing labeled duplex and no protein. After separation by mobility shift, free duplex was measured for both sets. The ratio of free duplex [Df] in the set with protein to free duplex without protein is equal to the ratio [Df]/[Dt]. The concentration of protein DNA complex [PD] (which is equal to the concentration of bound duplex [D_{Bound}]) becomes:

\[
\frac{[PD]}{[Dt]} = 1 - \frac{[Df]}{[Dt]}
\]

Plotting bound duplex ([PD]) as determined by the above equation against [Dt] gives a plot whereby the saturation level is equal to the concentration of active protein [Pt] in the assay. Activity of stocks was determined as $3.3 \times 10^{-10}$ M for R404K, $9.4 \times 10^{-8}$ M for R404D, $1.3 \times 10^{-9}$ M for R408N and $1.3 \times 10^{-8}$ M for R408D. For crude extracts, activity was found to be $1 \times 10^{-9}$ M when 1 μl of Rap1 extract is used in an assay, giving a Rap1 concentration of $1 \times 10^{-8}$ M for each extract.
**Materials and Methods**

### 4.2.2. Kd determination

K$_d$ was determined by mobility shift of increasing concentrations of protein activity with $<10^{-12}$ M radio-labeled telomeric binding site [Dt]. Since [Dt] is less than the published K$_d$ for Rap1 ($1\times10^{-11}$), [Pt] will be equivalent to free protein concentration [P$_{free}$]. The following equation describes this relationship:

$$\frac{[PD]}{[Dt]} = \frac{1}{1 + Kd / [P_{free}]}$$

At the point where K$_d$ equals [P$_{free}$], the equation can be simplified to:

$$\frac{[PD]}{[Dt]} = \frac{1}{1 + 1} = \frac{1}{2} = 0.5$$

Mobility shifts were analyzed using the Phosphor Imager system, and values were obtained for total duplex oligonucleotide concentration [Dt] and concentration of protein bound to duplex oligonucleotide [PD]. By plotting the fraction of protein bound to duplex oligonucleotide over total duplex oligonucleotide concentration, [PD]/[Dt], K$_d$ was determined as the point where [PD]/[Dt] = 0.5. All K$_d$ values were calculated at least twice in separate experiments. In the case of R404 and R408 mutants, K$_d$ values were determined using both T7 and recA expression systems.

### 4.3. Competition analysis

For each assay, a mobility shift was performed whereby a 100x dilution of 1 µl of crude extract ($1\times10^{-11}$M final concentration) was combined with labeled telomeric Rap1 binding site at a final concentration of $10^{-10}$ M. For binding competition, a 100, 300 or 500-fold molar excess of either unlabelled telomeric Rap1 binding site, unlabelled Rap1 binding site from the HMR-E silencer or unlabelled Rap1 binding site from the HML-E silencer was added. Concentration of unlabelled binding site was determined by diluting a spectroscopically determined concentration of stock solution binding site duplex oligonucleotide. A control lane containing labeled telomeric binding site and protein was also separated in parallel. The ratio of bound to total signal was calculated for each lane. Each assay was performed in triplicate to determine mean values and
standard deviation. Once mean values were obtained, a graph of moles per liter unlabelled binding site against % bound signal was plotted (see Figure 10 for example). The concentration of unlabeled duplex able to reduce the signal of labeled bound duplex by 50% was determined as IC\textsubscript{50}. In order to provide consistency, all competition assays were performed using a pRG50-derived plasmid.

Competition assays were performed as a standardized qualitative tool to compare the relative affinity of Rap1 species to various Rap1 binding sites. It should be noted that IC\textsubscript{50} values are unrelated in themselves to Kd values. The competition results are only significant in comparison to each other. As all competitions were performed under standard conditions, differences in IC\textsubscript{50} suggest a relative difference in affinity between two Rap1 species, or between two Rap1 binding sites.

4.4. Cross-linking mobility shift assay

To assay the ability of the HTH1 motif to bind DNA, a cross-linking enhanced mobility shift experiment was performed. Briefly, a 40 µl mobility shift mix was prepared in six separate tubes, with 1 µl HTH1 protein extract and 10^{-10} M \textsuperscript{32}P-labeled telomeric binding site in each. Tubes 1-3 also contained 500 ng of poly dIdC. Tubes 2 and 5 contained 1 µl of a 1:100 dilution of Rap1 antiserum, and tubes 3 and 6 contained 10^{-9} M unlabeled telomeric binding site. All tubes were incubated at room temperature for 10 minutes, then subjected to a single-pulse ultraviolet laser irradiation (Ho et al., 1994) in an open 1.5 ml microfuge tube. The laser, a Quanta-Ray pulsed Nd:YAG laser (Spectra-Physics) provides a 6.4 mm diameter beam of monochromatic 266-nm light. Each 5 ns pulse provides energy of up to 60 mJ.

After irradiation, samples were stored on ice before being loaded on a 15% SDS-PAGE. After separation, the gel was dried and exposed to a Phosphor Imager screen. Any protein bound specifically to the labeled binding site during irradiation should crosslink to the DNA, providing a stable complex that would run as a single radioactive band on SDS-PAGE. The apparent molecular mass of the band should be roughly equal to the molecular mass of both the binding site and HTH1.
**Figure 10. Sample Competition Assay Plot**

Plot of Rap1 competition assay with unlabelled telomeric binding site. At the point of 50% competition of signal (intensity of approximately 400), IC$_{50}$ was calculated as 5x10$^{-9}$ M. The IC$_{50}$ data was plotted to a logarithmic curve to determine a standardized IC$_{50}$ value.

The competition assays were performed as a qualitative tool to compare the relative binding of Rap1 species to various Rap1 binding sites. IC$_{50}$ values are able to compare the relative binding affinity of two Rap1 species, or two binding sites, but the values obtained are unrelated to the more rigorous Kd determinations.
5. CAT ASSAY

5.1. Handling of Cos-7 Cells

CAT assays were performed on Cos-7 cells grown in 100 mm plates. Cells were grown in DMEM medium (pH 7.4) containing 1x antibiotic mix (Pen-Strep) and 10% FBS. Cells were grown at 37°C with 5% CO₂. Confluent plates were subcultured by standard procedures. Growth medium was aspirated from the plate and 2 ml of Trypsin medium (Gibco-BRL) added, incubating the plate for 10 minutes at 37°C. After checking under a low power microscope for floating cells, 10 ml of DMEM (10% FBS) was added, and the solution pipetted repeatedly to break up clumps of Cos-7 cells. The cells were then divided among into 10 plates, and media brought up to 10 ml DMEM (10% FBS) per plate. Typically, cells became confluent within 2-3 days. For long term storage, 1 ml aliquots of cells were brought up to 10% in DMSO after trypsinization, mixed gently, slowly frozen to -70°C, and finally stored under liquid nitrogen.

5.2. CAT assay protocol

Standard procedures were utilized for the CAT assay as outlined in Gill et al. (1990). Cells subcultured 1:5 the day before were washed twice in DMEM (no FBS) and layered with 3 ml DMEM transfection mix containing a total of 10 µg CAT reporter plasmid, 10 µg expression plasmid and 0.4 mg of DEAE dextran (Pharmacia). After incubation for eight hours at 37°C, 5% CO₂, the transfection mix was removed by aspiration, 3ml of 10% DMSO in PBS (pH 7.4) was layered on cells and plates were left 5 minutes at room temperature. The DMSO mix was aspirated and cells were washed twice gently with PBS. Cells were allowed to recover in DMEM with 10% FBS and incubated 37°C, 5% CO₂ for 48 hours.

5.3. Measurement of chloramphenicol acetylation

Measurement of CAT activity was performed as described (Gill et al., 1990). After transfection and a 48 hour incubation, cells were gently washed twice with 3 ml of ice cold PBS. The final wash was aspirated, and 1 ml PBS added. The cells were then scraped from the plate into a 1.5 ml microfuge tube, spun for 20 seconds at low speed in a microfuge and the supernatant aspirated. The cells were resuspended in 60 µl of 0.25 M Tris-HCl (pH 8.0), and quick frozen/quick thawed 3x using a dry ice ethanol bath and a 37°C water bath to produce a
crude protein lysate. The lysate was then spun for 5 minutes at high speed in a microfuge at 4°C, and the supernatant removed to a new tube. 5 μl of the lysate was set aside to determine protein concentration.

The other 55 μl of lysate was used for a CAT assay which included lysate, 70 μl 1 M Tris-HCl (pH 8.0), 2 μl \([^{14}C]\) chloramphenicol (0.05 μCi/μl) and 20 μl Acetyl Coenzyme-A (3.2 mg/ml). This reaction was vortexed, incubated at 37°C for 3 hours and stopped by the addition of 1 ml of ethyl acetate. The tubes were vortexed twice for 20 seconds, centrifuged briefly in a microfuge, and the top 950 μl ethyl acetate layer removed to a clean tube. The samples were then dried down under vacuum.

The resulting residue was resuspended in 20 μl of ethyl acetate, spotted on Whatman LK6D silica TLC plates and separated in chloroform/ethyl acetate (70:30) until the solvent front migrated 10-15 cm. After air drying, the plates were exposed overnight on Phosphor Imager screens. The ratio of acetylated to total \([^{14}C]\) chloramphenicol was calculated using the IP Labs gel program. Values obtained were standardized for variation in total protein concentration as calculated by BIORAD Protein Assay.

5.4. Statistical analysis

All samples were tested in triplicate and the data used to give mean and standard deviation values. All three samples were assayed together during the same experiment, ensuring that each plate of cells was exposed to the same conditions throughout. Control acetylation for each experiment, using cells transfected with only CAT reporter plasmid, was subtracted from the data to provide final % acetylation and standard deviation values. A set of control plates, transfected with 10 μg of expression plasmid and 10 μg of pMV1 were also assayed to ensure the Rap1 proteins were unable to activate transcription from any cryptic sites.

6. BLAST ANALYSIS OF RAP1

Using default parameters (a non redundant database with pairwise alignment), sequence alignment to the Rap1 DNA binding domain was searched using the BLAST program (Basic Local Alignment Search Tool, Altschul et al., 1990). The BLAST search used the SWISS-PROT Release 31.0 database of 43,470 protein sequences. Weak scores were produced to a
variety of Myb and Myb-like proteins including the *A. thaliana* Myb gene product, *P. hybrida* Myb-like protein 1, *B. Taurus* proto-oncogene c-Myb and *M. musculus* Myb protein.

7. VISUALIZING THE RAP1 THREE DIMENSIONAL STRUCTURE

7.1. SETOR analysis of crystal structure

Computer data for the three dimensional atomic structure of the Rap1 DBD bound to a telomeric site was provided by Dr. Daniela Rhodes. This data was visualized using the SETOR computer program. The original data contained two Rap1 molecules bound to a repeated telomeric site. To aid in visualization, one set of Rap1 DBD and telomere was removed, and the remaining protein DNA complex was displayed as a solid shape, circular ribbon. Distinguishing colors were provided to each molecule, Rap1 backbone in yellow, one DNA strand in red and the other in blue (see Figure 8).

7.2. Modeling of site-directed mutations

In order to highlight residues selected for site-directed mutagenesis, the single residue was displayed in blue extending from the protein backbone. All other residues of DBD were not displayed to aid visualization. Using the amino acid substitution feature of SETOR, a wild-type residue, such as Arg-404, was changed to the desired mutant amino acid. The program is unable to alter the three dimensional structure after mutagenesis, but does indicate a possible orientation of the mutant residue after substitution. In order to highlight regions 418-438 and 588-594, these regions were selected as loops, then displayed in red.

7.3. Cartoon structure of DBD subdomains

To create a cartoon of the various DBD subdomains, the various HTH1 and HTH2 alpha helical residues were selected as α-helices and displayed as yellow barrels. The “tail” and “loop” regions were selected as loop and displayed as circular ribbon in red. The rest of the DBD structure was displayed as “spaghetti” solid.
8. PHOSPHORYLATION OF RAP1

Purified DBD and HTH1 protein was phosphorylated by treatment with Bovine heart muscle cAMP-dependent kinase (Sigma). 100 ng of protein was incubated with 1 unit of kinase in a 10 µl reaction containing 20 mM Tris (pH 7.4), 1 mM DTT, 100 mM NaCl, 50 µM ATP and 12 mM MgCl₂. The reaction was incubated for 1 hour at room temperature. The reaction was stopped by the addition of EDTA to 10 mM final concentration, and the protein samples were mixed with SDS sample buffer, and separated on a 15% SDS-PAGE gel, along with a protein molecular weight marker. The gel was stained with Coomassie blue to confirm the relative amounts of DBD and HTH1 protein loaded. The stained gel was dried onto Whatman paper and exposed to a Phosphor Imager screen overnight.
III. RESULTS

SECTION A. PREDICTION OF RAP1 DNA BINDING DOMAIN RECOGNITION HELICES FROM ITS PRIMARY SEQUENCE

1. Analysis of the Rap1 DNA binding domain using BLAST (Basic Local Alignment Search Tool)

The major goal of my research was to elucidate the mechanism by which Rap1 binds DNA in a sequence specific manner. This project was initiated before the crystal structure of the Rap1 DNA binding domain bound to telomeric DNA (outlined in Figure 8) was published. In order to determine whether the Rap1 DBD contains known DNA binding motifs, I first performed a BLAST analysis between full Rap1, Rap1 DBD, and other known gene products. This search found a weak identity (27%) between a number of Myb-related DNA binding motifs and the N terminus of the Rap1 DBD. The Myb and Myb-related motifs included protein sequences from diverse organisms, such as A. thaliana, P. hybrida, B. taurus, M. musculus, G. gallus and H. sapiens. In addition, previous work had suggested that Rap1 contains a site of weak similarity to the *Xenopus* A-Myb oncoprotein DNA binding motif (Larson et al., 1994).

2. Model prediction of HTH1

While the number of Myb domains identified by BLAST search was encouraging, the extent of similarity between the Rap1 DBD and these Myb domains was very low. Therefore, this region of similarity was analyzed in greater detail. Figure 11 shows an alignment of *S. cerevisiae* Rap1 (residues 358-415) and the Myb DNA binding motif. This alignment is based on the regions of identity selected by the BLAST program. *S. cerevisiae* and *K. lactis* Rap1 show a high level of identity (95%) in this region, showing conservation in yeast, though *K. lactis* Rap1 is unable to complement *S. cerevisiae* Rap1 activity (Larson et al., 1994). Alignment of the Myb sequences and Rap1 reveal a number of key features. The shaded bars in Figure 11 show residues revealed by NMR analysis to form hydrophobic clusters within the R2R3 DNA binding domain of mouse c-Myb (Ogata et al., 1994). These clusters are presumed to form a stabilizing core which serves to maintain the three dimensional structure of the motif. If Rap1 were to contain a Myb-like motif, it would be expected to maintain the majority of these
A comparison between Rap1 residues 358 to 415 and the Myb HTH domain. Shown are the primary sequences of *S. cerevisiae* Rap1, *K. lactis* Rap1, and repeat three (R3) of the c-Myb and a-Myb DNA binding domains from mouse. The sequence identity between the two Rap1 molecules is also shown. "+" indicates similar amino acid function according to the BLAST sequence comparison program. The conserved tryptophan residues of the Myb motif are indicated by "*" and residues involved in assembly of the hydrophobic core of C-Myb R3 HTH are shaded. The three helices of C-Myb are underlined.

Figure 11. Protein sequence comparison
sites. As is shown in the Figure 11, Rap1 also contains hydrophobic residues that align well with those of the Myb domain. Hydrophobic residues are present at 10 of the 11 positions, with only arginine at position 406 differing. In addition, three highly conserved tryptophan residues are found in most Myb-like proteins. Rap1 does not contain any tryptophan residues in this region, but aromatic hydrophobic residues (phenylalanine and tyrosine) are located at the corresponding sites in the Rap1 sequence. These include Rap1 residues Phe-363, Tyr-388 and Phe-407. Our computer search revealed various Myb-like DNA binding motifs (proteins Hv1, AtG11, Am308, Ph3) which substitute phenylalanine for a conserved tryptophan, but maintain a Myb-like binding motif. The presence of residues with potential to form a hydrophobic cluster and of aromatic residues at the positions occupied by tryptophan in Myb motifs suggested the presence of a Myb-related motif spanning amino acids 358 to 415, in the N-terminal portion of the Rap1 DBD (Fig. 2).

If Rap1 contains a Myb-related motif involved in DNA binding, then it should also contain a DNA recognition helix between residues 401 and 411. In Myb motifs, residues 3, 6, 7 and 11 of the recognition helix (helix 3) are hydrophobic, with residue 7 typically being tryptophan. In this region, Rap1 contains a similarly spaced hydrophobic residue at 3 of the 4 positions: isoleucine 403, phenylalanine 407 and leucine 411.

For mouse c-Myb, base-specific DNA recognition takes place via the fourth and eighth residues of the recognition helix (Ogata et al., 1994), with contributions to major groove binding from the first, fifth, and ninth residues. In Rap1, the fourth and eighth residues correspond to Arg-404 and Arg-408. In order to substantiate a possible role for these residues in DNA binding, a comparison of their preferred base interactions with the Rap1 consensus binding site was determined. A review of interactions between HTH recognition helix residues and the major groove (Suzuki, 1994) indicates that an arginine residue would most likely interact with a guanine. Also, Rap1 protects the guanine residues located in the Rap1 core region from attack by KMnO₄ (Gilson et al., 1993), consistent with interaction between R404 and R408 in the putative HTH and guanines. In addition N401 and H405 would be predicted to form interactions as the first and fifth residues of a recognition helix. On the basis of these observations, it was decided to mutate R404 and R408 to determine their involvement in sequence-specific DNA recognition by Rap1.
Results

During the course of the mutagenesis experiments, Konig et al (1996) published the crystal structure of a Rap1 DBD-DNA complex. Their results verified my prediction of a Myb-like HTH. The crystal structure shows the existence of a hydrophobic cluster containing residues F363, F370, I371, V375, Y388, I391, V395, I403, F407 and L411, which were predicted from the BLAST analysis. Within the recognition helix, residues N401, S402, R404, H405, R408 and Y410 make base-specific contacts with telomeric DNA, with R404 interacting directly with a guanine and R408 interacting with a guanine via a water molecule.

3. Testing of mutants

After site-directed mutagenesis, crude extracts of wild-type DBD (residues 361 to 598) and two isolates of each mutation were assayed for DNA binding by electrophoretic mobility shift (see Materials and Methods) using a fragment from the TEF2 promoter containing a Rap1 binding site in the presence of poly dIdC (Figure 12A). The TEF2 site is known to bind strongly to the DNA binding domain in vitro (Henry et al., 1990) and a mobility shift corresponding to Rap1 DBD was observed (Figure 12A, lane 1). Mutants R404K (lanes 2 and 3), R404D (lanes 4 and 5) and R408D (lanes 8 and 9) showed no mobility shift under these conditions. A shift was observed for mutant R408N (lanes 6 and 7), at an intensity significantly lower than that of wild-type protein. Western blot analysis of the protein extracts indicated comparable levels of protein expression (data not shown).

Further analysis of these mutants was performed with a consensus telomeric Rap1 binding site and shows that, in contrast with the Tef2 site, all mutants were able to bind a telomeric site with high affinity. These experiments will be described in later sections. These data show the mutations did not disrupt DBD folding but that there are differences in affinity between Rap1 and TEF2 or telomeric sites. Comparison between a TEF2, telomeric (TEL) and consensus-binding site (Graham and Chambers, 1994) is shown in Figure 12B. Bases 1, 2, 13 and 14 differ between the telomere and TEF2 sites. Comparison of these three sites reveals deviation from the consensus at bases 13 and 14 of TEF2 (Figure 12B).
Results

Figure 12. Rap1 mobility shift of TEF2 binding site

(A) Crude extracts of E.coli expressing Rap1 DBD or various DNA binding mutants were incubated with a labeled TEF2 binding site at a concentration of $10^{-9}$ M (as described in Materials and Methods, section 4.1). Each mutation was tested in duplicate and a mutant from each screen was assayed. Extract (0.083 μg) was incubated with the probe for 10 minutes at room temperature and then loaded on a 5% non-denaturing polyacrylamide gel. Gels were separated for 2h at 4°C and exposed to X-ray film. (B) Comparison of the TEF2, telomeric and consensus Rap1 binding sites. Bases involved in hydrogen bonding to the DBD according to the crystal structure are underlined.
4. Model prediction of HTH2

The Rap1 DBD spans residues 361 to 596 and is therefore much larger than the Myb-like motif identified above, indicating the existence of additional DNA recognition elements. The C-terminal portion of the Rap1 DBD shows much lower identity between *S. cerevisiae* and *K. lactis* than the N-terminal portion of DBD, with extended areas of weak identity located between areas of similar protein sequence. Analysis of the spacing of hydrophobic residues in this region did not point to any HTH motifs similar to that of Myb. However, a conserved recognition helix should still be present. This recognition helix would likely be conserved between *S. cerevisiae* and *K. lactis* Rap1 and would contain an amphipathic array of hydrophobic residues. Selected and amplified binding (SAAB) studies of Rap1 DNA binding (Graham and Chambers, 1994) and reactivity assays (Gilson et al., 1993) of Rap1 suggested that a second recognition motif exists. This motif would interact with bases TGG (complement to ACC) at the 3' end of the consensus binding site. This putative HTH motif should contain an arrangement of Arg or Lys residues typically involved in base-specific contacts to guanine (Suzuki, 1994).

A search of the DBD primary sequence revealed a potential site at the C-terminal portion of the binding domain. The sequence contains a Myb-like spacing of hydrophobic residues, an aromatic hydrophobic residue at the corresponding site for Myb, and arginine residues at residues 4 and 8. This motif, from residues 539 to 549 (see Figure 13), also contains a high level of identity between *S. cerevisiae* and *K. lactis*. It includes hydrophobic residues at sites 3, 7, and 11, in a manner similar to that of c-Myb. It is missing a hydrophobic residue at site 6, but has an additional hydrophobic residue at site 10, which could substitute. This gives the region an identical spacing of hydrophobic residues to that of the Rap1 Myb-like HTH.

Residues 4 and 8 of this putative recognition helix (amino acids 542 and 546) are both arginine and would be expected to interact with a guanine residue according to Suzuki (1994). *In vitro* studies (Gilson et al., 1993) indicate that guanine residues are indeed protected by Rap1 at the site. Base-specific interactions with guanine residues are in fact seen in the DBD crystal structure (Konig et al., 1996). Mutations of Arg-542 and Arg-546 would be expected to disrupt Rap1 DNA binding.
Figure 13. Putative HTH2 sequence comparison

A comparison between Rapl residues 539 to 550 and the Myb recognition helix. Shown are the primary sequences of *S. cerevisiae* Rapl, *K. lactis* Rapl, and the recognition sequence of R3 of C-Myb and A-Myb. Identity between the two yeast Rapl sequences is also shown. "+" indicates similar amino acid function according to the BLAST sequence comparison program. Residues comprising the hydrophobic cluster of C-Myb are shaded (Ogata et al., 1994). In the case of Rapl, residue 544 is not hydrophobic, but other residues are consistent with forming a hydrophobic cluster.
5. **Comparison of the Rap1 DBD model prediction and DBD crystal structure**

With the publication of crystal structure for the DNA binding domain during the course of this study, direct comparisons could be made between predictions from Myb primary sequence and the three dimensional structure of the Rap1 DBD bound to a telomeric site. According to Konig et al. (1996), Rap1 contains two HTH motifs, with recognition helices between residues 400 to 409 and 538 to 552. The first motif, HTH1, contains a recognition helix involved in base-specific contacts at what would be sites 1, 2, 4, 5, 8 and 10 (Figure 11) on a Myb recognition helix (residues Asn 401, Ser 402, Arg 404, His 405, Arg 408 and Tyr 410 of Rap1). Of these sites, Arg-404 is involved in the most base-specific hydrogen bonding, making interactions with the N7 atoms of two guanine residues, and well as an O6 contact to one of these guanines via a water molecule. Arginine 408, while involved in hydrogen bonding, does so via a water molecule, and contacts a thymidine rather than a guanine. These results are in agreement with my findings that mutations of Arg-404 and Arg-408 will disrupt interactions with DNA, and that mutant R408N could still maintain partial DNA binding to a Rap1 site.

For HTH2 (See Figure 13), crystal structure data indicates that Rap1 residues 539, 542, 543, 546 and 547 are involved in base-specific contacts (corresponding to recognition residues 1, 4, 5, 8 and 9 of Myb). Of these residues, Arg-546 is involved in two base-specific hydrogen bonds to a guanine (one to an O6, the other to a N7 atom). Arg-542 is involved in a single hydrogen bond to the N7 atom of a guanine.

Taken together, these results confirm a number of my predictions concerning the DBD. As suggested by protein sequence comparison, Rap1 contains Myb-like HTH domains. In addition, prediction of the DNA recognition helix for HTH1 was correct, with Arg-404 a key component of the protein-DNA interaction. My prediction of the HTH2 recognition helix was also confirmed from the crystal structure data.

6. **Determination of Rap1 and site-directed mutant dissociation constants**

In order to examine the Arg-404 and Arg-408 site-directed mutants, the wild-type and mutant Rap1 DNA binding domains were expressed and purified using an *E. coli* T7 polymerase expression system (see Materials and Methods). Protein was purified by FPLC to between 30% to 90% homogeneity (see Figure 14). A single band at 29 kDa was present in all
Figure 14. Purification of Rap1 proteins

Rap1 DBD and Rap1 DBD site directed mutants were purified to 30% to 90% homogeneity as described in Materials and Methods, section 3.3. Purification involved expression in BL21(λDE3) cells, separation on a heparin sepharose column, and final purification by FPLC on a Mono S ion exchange column. Rap1 protein eluted at between 0.45 and 0.5 M NaCl. DBD runs as a 29 kd band, which was confirmed by western blot analysis. (A) Lane 1: wildtype DBD; 0.1 mg/ml. Lane 2: R404K DBD; 0.033 mg/ml. (B) Lane 1: R404D DBD; 2.16 mg/ml. Lane 2: R408N DBD; 0.2 mg/ml. Lane 3: R408D DBD; 0.64 mg/ml.
results samples. The DNA binding activity of the proteins was then quantified (see Materials and Methods, section 4.2).

Dissociation constants for the binding of each protein to a telomeric binding site were determined. Wild-type Rap1 DBD gave a dissociation constant of $1 \times 10^{-11}$ M, comparable to the previously published value for full length Rap1 of $1.3 \times 10^{-11}$ M (Vignais et al., 1990). All other assays are summarized in Figure 15. Mutant R404K DBD (Figure 15A) gave a gradual shift in mobility, with half dissociation (lane 5) occurring at $3 \times 10^{-11}$ M. This mutant results in a 3-fold loss of DNA binding affinity as measured by mobility shift. Mutant R404D DBD, which should disrupt two hydrogen bonds to bases via the major groove, results in a distinct mobility shift at a concentration of $3 \times 10^{-10}$ M (Figure 15B). This corresponds to a 30-fold loss in DNA binding affinity as compared to wild-type protein.

Dissociation constants were also determined for Arg-408 DBD mutant proteins. R408N DBD gave a distinct shift at a concentration of $1 \times 10^{-11}$ M (Figure 15C). This $K_d$, which corresponds to that of wild-type DBD, is consistent with results of the TEF2 mobility shift assay (Figure 12). The R408D DBD mutation gave a gradual shift in mobility (Figure 15D), resulting in half shifting (between Lanes 6 and 7) of $1 \times 10^{-10}$ M. The reason for the gradual shift is unclear, but was consistent for all R408D mobility shifts. This corresponds to a 10-fold decrease in DNA binding affinity when compared to wild-type DBD. This result appears to contradict the loss of signal seen in the TEF2 mobility shift. This may simply be due to the concentration of protein used in the TEF2 assay, allowing loss of signal with a 10-fold decrease in DNA binding affinity. Therefore, it was decided to perform a direct comparison between Rap1 binding to telomere and non-telomere sites.

SECTION B. COMPARISON OF RAP1 AFFINITY FOR TELOMERIC AND SILENCER DNA BINDING SITES

After carrying out TEF2 mobility shift assays and determining dissociation constants at telomeric sites, a number of questions were raised. First, if Rap1 has a $K_d$ of ca. $1 \times 10^{-11}$ M for a telomeric site, is this affinity maintained at other known in vivo Rap1 binding sites? Secondly, results indicate that Rap1 mutants bind TEF2 with low affinity, even at high protein and DNA concentrations. These same mutants bind a telomeric site at high affinity ($K_d$ of at least $3 \times 10^{-10}$
Results

Figure 15. Kd determination of DBD site directed mutants

Increasing concentrations of DNA binding domain mutants were incubated with $<10^{-12}$ M labeled telomeric binding site in a mobility shift assay (see Materials and Methods, section 4.2). Complexes were separated on a non-denaturing polyacrylamide gel and exposed to a phosphor-imager screen. $K_d$ was determined as the concentration resulting in half shifting of the radioactive label. Error estimates are within 10% for all $K_d$ determinations. Samples of mobility shifts are shown at left. All values are for DBD mutants. (A) R404K, (R404K concentrations (M): 1) 6x10^{-13}, 2) 1x10^{-12}, 3) 3x10^{-12}, 4) 6x10^{-12}, 5) 1x10^{-11}, 6) 3x10^{-11}, 7) 6x10^{-11}, 8) 1x10^{-10}) (B) R404D, (R404D concentrations (molar): 1) 1x10^{-11}, 2) 3x10^{-11}, 3) 6x10^{-11} 4) 1x10^{-10}, 5) 3x10^{-10}, 6) 6x10^{-10}, 7) 1x10^{-9}, 8) 3x10^{-9}). (C) R408N, (R408N concentrations (molar): 1) 6x10^{-13}, 2) 1x10^{-12}, 3) 3x10^{-12}, 4) 6x10^{-12}, 5) 1x10^{-11}, 6) 3x10^{-11}, 7) 6x10^{-11}, 8) 1x10^{-10}). (D) R408D, (R408D concentrations (molar): 1) 3x10^{-12}, 2) 6x10^{-12}, 3) 1x10^{-11}, 4) 3x10^{-11}, 5) 6x10^{-11}, 6) 1x10^{-10}, 7) 3x10^{-10}, 8) 6x10^{-10}). Concentration of protein was determined by protein activity assay (see Materials and Methods, section 4.2).
Results

A. R404K DBD
Kd = $3 \times 10^{-11}$ M

B. R404D DBD
Kd = $3 \times 10^{-10}$ M

C. R408N DBD
Kd = $1 \times 10^{-11}$ M

D. R408D DBD
Kd = $1 \times 10^{-10}$ M
M), well above non-specific binding levels (8.7x10^{-6} M, Vignais et al., 1990). Do non-telomeric Rap1 binding sites show increased sensitivity to Rap1 mutations, and if so, why? Third, Rap1 protein is known to affect silencing at the HMR and HML regions of chromosome 3. This silencing appears to be dependent on a concentration gradient of Rap1 (Buck and Shore, 1995). Does Rap1 bind with equal affinity to both silencer and telomeric sites, and what effect might this have on the sensitivity of silencers to the concentration gradient?

To examine these questions, a series of protein-DNA affinity experiments was designed. These compare the affinity of Rap1 to three known in vivo and in vitro binding sites: a consensus telomeric site, the Rap1 DNA binding site from the HMR-E region, and the Rap1 DNA binding site from HML-E. As shown in Figure 16, all three sites maintain key aspects of the Rap1 consensus binding site. All sites maintain a “core” ACC at bases 4 to 6, required for in vitro binding of Rap1. The telomeric site has a non-consensus cytidine at base 1, HMR-E has non-consensus sequence at bases 1, 3, and 11, and HML-E has non-consensus sequence at bases 1, 3, 13 and 14. Of these alterations, base 3 would affect interaction with HTH1 and the residue 588 to 594 region (for description, see Introduction section 8), while bases 11, 13 and 14 would affect interaction with HTH2.

Three methods were used to test the affinity of Rap1 to both telomeric and silencer binding sites. First, dissociation constants ($K_d$) were determined for the protein at telomeric and silencer sites. The in vitro affinity of a Rap1 derivative for its binding site can then be compared between binding sites, and between different mutations in the protein.

Secondly, competition assays were performed to determine the concentration required to effect 50% competition ($IC_{50}$) of a labeled telomeric binding site by unlabelled telomeric or silencer site in a mobility shift assay. The advantage of these assays is that they can be performed in parallel on the same polyacrylamide gel, and use the same protein preparation and radiolabelled DNA probe. The only variable is the unlabelled competitor DNA binding site. This allows for a direct comparison between competition by telomeric or silencer sites. This method also confirms dissociation constant data. This procedure allows the screening of more mutants and binding sites compared to the multi-step process of $K_d$ determination. Experiments can easily be performed in triplicate, allowing statistical analysis. It should be noted, however, that $IC_{50}$ data provide only a comparison in affinity between proteins and binding sites. The $IC_{50}$ data are not directly comparable to $K_d$ data.
### Results

**Figure 16. Comparing Rap1 telomere and silencer DNA binding sites**

Consensus telomeric, HMR-E and HML-E binding sites are shown, as well as the Rap1 consensus site derived from selected and amplified binding experiments (Graham and Chambers 1994). Bases involved in hydrogen bonding to the Rap1 DBD are underlined.

<table>
<thead>
<tr>
<th></th>
<th>TEL</th>
<th>HMR-E</th>
<th>HML-E</th>
<th>Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base Number</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14</td>
<td>C A C A C C C A C A C A C C</td>
<td>A A A A C C C A T C A A C C</td>
<td>T A/G C A C C C A N N C A/C C C</td>
</tr>
</tbody>
</table>

HTH1/"TAIL"  
HTH2
Finally, *in vivo* binding activity of Rapl was determined by CAT assay in Cos-7 cells. Again, assays were performed using both telomeric and silencer binding sites. Because Rapl is an essential protein in yeast, *in vivo* studies in the yeast cell would be complicated by effects of Rapl toxicity. CAT assays in a mammalian system quantify the effect of various mutations on Rapl binding, but are unlikely to affect the cell phenotype as no Rapl homologue exists in mammalian cells.

1. Affinity of Wild-type Rapl for telomere and silencer sites

1.1. \( K_d \) determination

Figure 17A summarizes the various Rapl \( K_d \) determination experiments. As shown in the autoradiogram, wild-type protein produces a distinct mobility shift at \( K_d \). Using a telomeric site, Rapl binds with a dissociation constant of \( 1 \times 10^{11} \) M, consistent with published values (Vignais et al., 1990). At an HML-E binding site, the dissociation constant is also \( 1 \times 10^{11} \) M, indicating equal *in vitro* affinity of Rapl to both sites.

These experiments were also performed in parallel with a Rapl DNA binding domain deletion mutant containing residues 353 to 598 of the protein. This DBD is the full binding domain of Rapl. The DNA binding domain exhibited a \( K_d \) identical to that of the full length protein.

1.2. Competition assay

Both full length Rapl and the DNA binding domain deletion were analyzed for differences in \( IC_{50} \) values between telomeric and silencer sites (Figure 17B). A sample competition experiment is shown at left. The control lane (lane 1) indicates binding of DBD to a labeled telomeric site. Lanes 2 to 4 contain increasing concentrations of unlabeled telomeric binding site. Lanes 5 to 7 and 8 to 10 contain increasing concentrations of unlabeled HMR-E and HML-E binding sites respectively. A duplex oligonucleotide containing a telomeric site inhibited Rapl binding to a labelled telomeric site with an \( IC_{50} \) of \( 5 \times 10^9 \) M (Figure 17B). A value of \( 1 \times 10^{-8} \) M was obtained using unlabeled HMR-E binding site, but using unlabeled HML-E binding site resulted in an \( IC_{50} \) of \( 3 \times 10^{-8} \) M. This \( IC_{50} \) value for HML-E suggests that
Results

Figure 17. Rap1 protein interactions with telomeric and silencer binding sites under various conditions

Wildtype Rap1 and Rap1 DBD protein was subjected to various studies. (A) Kd determination of the wt protein. A sample mobility shift used to determine the DBD Kd is shown. Concentrations of protein in sample mobility shift lanes [M]: 1. 6x10^{-13} 2. 1x10^{-12} 3. 3x10^{-12} 4. 6x10^{-12} 5. 1x10^{-11} 6. 3x10^{-11} 7. 6x10^{-11} 8. 1x10^{-10}. Protein was incubated with <10^{-12} M labeled oligo duplex. Values were obtained for binding to both telomere and HML-E sites. (B) Competition analysis of Rap1 affinity. Rap1 labeled telomeric binding site (1x10^{-10} M) was subjected to mobility shift with increasing concentrations of unlabeled telomeric or silencer site. A sample DBD competition assay is shown at left. Concentrations of duplex oligonucleotide: Lane 1; None, Lanes 2, 5 and 8; 1x10^{-8} M, Lanes 3, 6 and 9; 3x10^{-8} M, Lanes 4, 7 and 10; 5x10^{-8} M. IC50’s were calculated from the mean of three assays by determining the molarity of unlabeled duplex oligonucleotide that reduces the control band intensity by 50%. Plotting the intensity of the shifted band against the concentration of unlabelled duplex oligonucleotide gave an estimate for the concentration of unlabelled duplex required. See Materials and Methods section 4.3 for a description and a sample plot of competition data. (C) CAT assay analysis of Rap1. Rap1 expression plasmids were co-transfected with telomeric or silencer reporter plasmids into COS-7 cells, and levels of acetylated product were determined from three assays (see Materials and Methods section 5.3). Mean percent acetylation was calculated, along with standard deviation values (shown as error bars).
### Results

**Kd (M)**

<table>
<thead>
<tr>
<th></th>
<th>Rap1</th>
<th>DBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEL</td>
<td>$1 \times 10^{-11}$</td>
<td>$1 \times 10^{-11}$</td>
</tr>
<tr>
<td>HML-E</td>
<td>$1 \times 10^{-11}$</td>
<td>$1 \times 10^{-11}$</td>
</tr>
</tbody>
</table>

**IC$_{50}$ (M)**

<table>
<thead>
<tr>
<th></th>
<th>Rap1</th>
<th>DBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEL</td>
<td>$5 \times 10^{-9}$</td>
<td>$1 \times 10^{-8}$</td>
</tr>
<tr>
<td>HMR-E</td>
<td>$1 \times 10^{-8}$</td>
<td>$1 \times 10^{-8}$</td>
</tr>
<tr>
<td>HML-E</td>
<td>$3 \times 10^{-8}$</td>
<td>$2 \times 10^{-8}$</td>
</tr>
</tbody>
</table>

---

**Diagram A**

- Labelled lanes 1 to 8
- Image of gel electrophoresis

**Diagram B**

- Lanes 1 to 10
- Triangles for TEL, HMR-E, HML-E
- Specific bands indicated

**Diagram C**

- Bar chart showing % Acetyl
- Comparison between TEL and HML-E
- Rap1 and DBD sections
over 3-fold more unlabeled HML-E binding site than unlabeled telomeric binding site is required to titrate labeled telomeric binding site.

Similar trends were obtained with the DNA binding domain. IC$_{50}$ values of $1 \times 10^{-8}$ M were obtained for telomeric and HMR-E sites. The HML-E binding site produced an IC$_{50}$ of $2 \times 10^{-8}$ M, a 2-fold decrease in DBD affinity to HML-E when compared to the telomeric site. While this 2-fold difference is statistically significant, the results suggest DBD binds all sites with similar affinity.

1.3. CAT assay

In vivo binding of Rap1 was determined by a CAT assay in mammalian cells. Expression of full length Rap1 in cells containing a telomeric reporter plasmid gave acetylation values of 74%±24%. Acetylation values in cells containing a HMR-E reporter plasmid were 20%±5%. These standard deviations did not overlap indicating that Rap1 binds with higher affinity to telomeric as opposed to HMR-E sites in vivo. Acetylation using an HML-E reporter plasmid was 14%±3%, indicating similar affinity of HMR-E and HML-E sites in vivo. Values from control assays (without expression plasmid) were used to standardize results. Because transfection efficiency varies from experiment to experiment, all assays were performed simultaneously so cells on all transfection plates were treated identically.

CAT assays with the DNA binding domain gave similar results to full-length protein. Acetylation using a telomeric reporter was 80%±8%, well above that of an HMR-E reporter (15%±2%). Using an HML-E reporter construct, acetylation values were near background (2%), indicating that HML-E has lower affinity when compared to both HMR-E and telomeric binding sites.

2. Arginine 404 to lysine mutant has reduced affinity for silencer sites

The crystal structure of the Rap1 DBD bound to a telomeric site (Konig et al., 1996) indicates that the Rap1 residue Arg-404 plays a critical role in the interaction of Rap1 to the DNA sequence ACC found in all known Rap1 binding sites. In order to test this interaction, and its effect on binding to both telomeric and silencer sites, Arg-404 was mutated to a lysine residue, shortening the side chain by one carbon, but maintaining the charge between residue
404 and the DNA. This mutation is modeled in Figure 18. Figure 18A shows the DBD backbone, with residue Arg-404 indicated by an arrow. This residue extends from the backbone into the major groove, allowing three hydrogen bond contacts to occur. Figure 18B shows the hypothetical position of the Arg-404 to lysine mutant (R404K). The Lys-404 residue, shown in blue and indicated with an arrow, could be capable of maintaining at least one hydrogen bond in the major groove. Even if the mutant disrupts both Arg-404 DNA interactions, it should not otherwise disrupt protein-DNA binding or three-dimensional structure.

Various studies on R404K DNA binding affinity are shown in Figure 19.

2.1. $K_d$ determination

Figure 19A shows a sample mobility shift assay for mutant R404K at left. A gradual shift in label is observed with this mutant. $K_d$ values for full length protein are $5 \times 10^{-11}$ M using a telomeric binding site. Therefore a full length R404K mutant would bind telomeres with 5-fold lower affinity than the wild-type protein. A dissociation constant for binding to a HML-E site could not be determined, since $1 \times 10^{-9}$ M concentration, the highest obtainable for the experiment, was unable to shift the label fully. R404K therefore has a $K_d$ of greater than $10^{-9}$ M at HML-E sites, indicating at least 100-fold weaker affinity than wild-type protein ($1 \times 10^{-11}$ M).

Similar results were obtained for an R404K DNA binding domain. $K_d$ values of $3 \times 10^{-11}$ M using a telomeric binding site, and greater than $10^{-9}$ M using an HML-E binding site were measured. Again, this confirms weak binding to a HML-E site by R404K.

2.2. Competition assay

In order to compare telomeric affinity to both HMR-E and HML-E sites, competition assays were performed (Figure 19B). A sample competition experiment is shown at left. The experiment was repeated several times, and consistently showed strong competition by unlabelled telomeric binding site (lanes 2 to 4) and weak competition by HMR-E (lanes 5 to 7) and HML-E (lanes 8 to 10). These differences are also shown in $IC_{50}$ values.

Full length R404K had an $IC_{50}$ of $1 \times 10^{-8}$ M using unlabelled telomeric binding site, while unlabelled HMR-E gave an $IC_{50}$ of $8 \times 10^{-8}$ M. HML-E gave minimal competition, resulting in an $IC_{50}$ value of at least $5 \times 10^{-7}$ M. These silencer $IC_{50}$ concentrations are 8 and 50-fold higher than telomeric concentrations, respectively.
Figure 18. Three-dimensional model of the Rap1 Arg-404 to lysine mutation
(A) SETOR model of Rap1 DBD bound to a telomeric site. Residue arginine 404 is shown in blue. This residue makes two direct hydrogen bond contacts to bases via the major groove. (B) Putative position of lysine 404 in the site directed mutant. Residue 404 is shown in blue and indicated with an arrow. This mutant could still make one direct hydrogen bond to N7 of guanine via the major groove, and is unlikely to alter the three-dimensional structure of the protein.
Figure 19. R404K protein interactions with telomeric and silencer binding sites under various conditions

Full length RAP1, containing the R404K mutant, and a DBD R404K mutant protein were subjected to various studies (see Materials and Methods, section 4.2 and 4.3). (A) Kd determination of the protein. A sample mobility shift used to determine the R404K DBD Kd is shown. Values were obtained for binding to both telomere and HML-E sites. Concentrations of sample mobility shift lanes: R404K TEL (M): 1. 6x10^{-12} 2. 1x10^{-11} 3. 3x10^{-11} 4. 6x10^{-11} 5. 1x10^{-10} 6.3x10^{-10} 7.6x10^{-10} 8.1x10^{-9} (B) competition analysis of R404K affinity. Rap1 labeled telomeric binding site was subjected to mobility shift with increasing concentrations (10^{-8}M, 3x10^{-7}M, and 5x10^{-7}M) of unlabeled telomeric or silencer sites. A sample competition assay with full length R404K is shown at left. IC_{50}s were calculated from the mean of three assays by determining the molarity of unlabeled duplex required to halve the control amount of shifted label. (C) CAT assay analysis of R404K. R404K expression plasmid was co-transfected with telomeric or silencer reporter plasmids into COS-7 cells, and the level of acetylated product was determined from three assays. Mean percent acetylation was calculated, along with standard deviation values (shown as error bars). See Materials and Methods, section 5.
Results

<table>
<thead>
<tr>
<th>Kd (M)</th>
<th>R404K DBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEL</td>
<td>5X10^{-11}</td>
</tr>
<tr>
<td>HML-E</td>
<td>&gt;1X10^{-9}</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IC_{50} (M)</th>
<th>R404K DBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEL</td>
<td>1X10^{-8}</td>
</tr>
<tr>
<td>HMR-E</td>
<td>8X10^{-8}</td>
</tr>
<tr>
<td>HML-E</td>
<td>5X10^{-7}</td>
</tr>
</tbody>
</table>
Results

Values with the DNA binding domain deletion of R404K were comparable to full length protein. Unlabelled telomere gave an IC$_{50}$ of 2x10$^{-8}$ M, unlabelled HMR-E gave a 15-fold increase in IC$_{50}$ and HML-E gave a 20-fold increase (Figure 19B).

2.3. CAT assay

In vivo binding of Rapl mutant R404K was determined by CAT assay in mammalian cells. Only minimal acetylation of chloramphenicol was observed in CAT assays involving full-length R404K expression plasmids. Expression plasmid data for telomeric, HMR-E and HML-E sites gave values only slightly above background levels. Similar results were obtained using a DNA binding domain deletion. In the case of DBD, percent acetylation values of 10% (telomere and HMR-E) and 5% (HML-E) were obtained. Rapl DNA binding at a telomeric site appears to be unaffected by removal of both N-terminal and C-terminal portions of the protein.

3. Affinity studies of other mutations at positions 404 and 408

A number of affinity studies were performed on Rapl mutants R404D, R408N and R408D. Models of the various mutants are shown in Figure 20 and Figure 21. Wild-type DBD bound to a telomeric site is shown in Figure 20A, with residue Arg-408 extending from the backbone. This residue makes a hydrogen bond contact through a water molecule to the major groove. Mutant R408N is modeled in Figure 20B. This residue, shown here extending away from the major groove, could alternatively extend into the major groove and form a hydrogen bond, maintaining the affinity of Rapl for DNA. However, the shorter asparagine residue may be unable to bind due to distance constraints.

Figure 21 illustrates models of mutants R404D and R408D. Asp-404, an acidic residue, may disrupt interaction of HTH1 with the negatively charged DNA through electrostatic repulsion (Figure 21A). A similar situation could occur with R408D (Figure 21B). In addition, the shorter aspartate residue would not be able to hydrogen bond in the major groove.
Figure 20. Three dimensional model of the R408N mutation

Model of Rap1 DBD bound to a telomeric site. (A) Arginine 408 is shown in blue. This residue hydrogen bonds to a thymidine at base 4 via a water molecule. (B) R408N mutant. Putative location of the asparagine is shown in blue. This mutant may still be capable of hydrogen bonding via the major groove, and is unlikely to disrupt the stability of the protein.

Figure 21. Three dimensional model of the R404D and the R408D mutations

(A) Model of Rap1 DBD bound to a telomeric site. The putative position of mutant residue Asp-404 is shown in blue. This residue may disrupt the stability of the protein due to electrostatic repulsion of the aspartate carboxylate oxygen to guanine. However, distance from the residue to the DNA is increased from the 2.6 Å for the wildtype protein. This makes disruption unlikely. This mutant was able to bind DNA, suggesting that the protein is stable. (B) Putative position of Asp-408 in the site directed mutant. Residue 408 is shown in blue. This mutant would possibly cause electrostatic repulsion, but may also be involved in hydrogen bonding via the major groove.
Results

A

B
3.1. $K_d$ determination

$K_d$ values were obtained for all three mutants, as both full length and DBD deletion protein (Figure 22A). R404D gave $K_d$ values similar to those observed for R404K. Full length protein exhibited a dissociation constant of $5 \times 10^{-11}$ M using a telomeric site, but only $1 \times 10^{-8}$ M for a HML-E site, a 1000-fold loss in affinity compared to wild-type protein. The DBD version of R404D gave $K_d$ values of $3 \times 10^{-10}$ M for telomere and $1 \times 10^{-9}$ M for a HML-E site. The DBD values are distinct from the Rap1 values, with 6-fold higher values for telomere and 10-fold lower values for HML-E sites. It is unclear why these differences occur.

R408N protein gave wild-type levels of DNA binding affinity to telomeric sites. Binding to HML-E was reduced 5-fold for full length protein, and a 100-fold for the DNA binding domain mutant. R408D gave low binding affinity to both telomere and HML-E sites, binding with 10-fold less affinity to telomeres, and a hundred to a 1000-fold less affinity to HML-E.

3.2. Competition assays

Competition values for the mutants are shown in Figure 22B. Values were obtained for full length R404D and R408D. R408N values were obtained for both full length and DBD versions of the protein. R404D gave competition values comparable to those of R404K, though competition of HML-E was reduced. Telomeric competition resulted in an $IC_{50}$ of $2 \times 10^{-8}$ M. HMR-E and HML-E showed decreased affinity by 4-fold and 5-fold respectively.

$IC_{50}$ values for R408N show reduced affinity for a silencer site, with telomeric competition giving concentrations of $1 \times 10^{-8}$ M for full length protein. HMR-E and HML-E values are 2-fold and 10-fold higher, respectively. Similar reductions in affinity for HMR-E and HML-E are seen with the mutant DBD alone. A 5-fold increase in $IC_{50}$ is seen for HMR-E, a 10-fold increase for HML-E, when compared to telomeric competition.

$IC_{50}$ values for R408D are similar to those of R408N. Both HMR-E and HML-E exhibited a 5-fold reduction in affinity when compared to a telomeric site.

4. Lys-446 and Lys-360 mutations affect binding affinity

In addition to protein/DNA interactions via the major groove, the Rap1 protein also contains two residues, Lys-360 and Lys-446, which make base-specific interactions with
## Results

### A

<table>
<thead>
<tr>
<th>Kd (M)</th>
<th>R404D DBD</th>
<th>R408N DBD</th>
<th>R408D DBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEL</td>
<td>5X10⁻¹¹</td>
<td>1X10⁻¹¹</td>
<td>1X10⁻¹¹</td>
</tr>
<tr>
<td>HML-E</td>
<td>1X10⁻⁸</td>
<td>1X10⁻⁹</td>
<td>5X10⁻¹¹</td>
</tr>
</tbody>
</table>

### B

<table>
<thead>
<tr>
<th>IC₅₀ (M)</th>
<th>R404D</th>
<th>R408N</th>
<th>R408D</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEL</td>
<td>2X10⁻⁸</td>
<td>1X10⁻⁸</td>
<td>4X10⁻⁸</td>
</tr>
<tr>
<td>HMR-E</td>
<td>8X10⁻⁸</td>
<td>2X10⁻⁸</td>
<td>2X10⁻⁷</td>
</tr>
<tr>
<td>HML-E</td>
<td>1X10⁻⁷</td>
<td>1X10⁻⁷</td>
<td>4X10⁻⁷</td>
</tr>
</tbody>
</table>

**Figure 22.** R404D, R408N and R408D protein interactions with telomeric and silencer binding sites under various conditions

Full length and DBD deletions of R404D, R408N and R408D were subjected to Kₐ and competition studies. (A) Kₐ determination. Values were obtained for binding to both a telomere and HML-E site. Kₐ values were determined for protein expressed from both the pL3S5 and the pRG50 expression plasmids for the DBD mutants. See Materials and Methods, section 4.2. (B) Competition analysis of mutants. Protein and labeled telomeric binding site were subjected to mobility shift reactions with increasing concentration of cold telomeric or silencer site. Full length R404D and R408D Rap1 were assayed, as well as both the full length and the DBD versions of R408N. IC₅₀ values were calculated from the mean of three assays by determining the molarity of unlabeled oligonucleotide required to halve the control amount of shifted label. See Materials and Methods, section 4.3.
adenine via the minor groove in the crystal structure of Konig et al. (1996). In order to test the significance of these interactions, we removed the lysine side chain, replacing it with alanine at residue 446 or at residue 360. These substitutions are modeled in Figure 23 and Figure 24. Figure 23A shows the position of wild-type Lys-446. This residue lies at the start of the HTH2 subdomain (shown at right of Figure 23A). Replacement with an alanine residue (modeled in Figure 23B) would remove the positive charge and hydrogen bonding via the minor groove observed in the crystal structure (Konig et al., 1996), but should otherwise not disrupt the protein’s three-dimensional structure.

A similar replacement was carried out for residue Lys-360 (Figure 24). To display this residue in the computer generated model, the DBD backbone shown in Figure 23 was rotated 40° right to left on the Y axis. Lys-360 is shown protruding from the backbone into the minor groove (Figure 24A). Ala-360 is modeled in Figure 24B, removing possible hydrogen bonds to the minor groove. Since Lys-360 is not involved in stabilization of the three dimensional structure (Konig et al., 1996) we would not expect any disruption.

4.1. $K_d$ determination for mutants K446A and K360A

A sample $K_d$ determination for K446A is shown in Figure 25A. K446A gives a distinct shift at $6 \times 10^{-11}$ M on a telomeric site for both full length and DBD protein. This represents 6-fold lower affinity than the wild-type protein. A roughly 2-fold decrease in affinity is observed at HML-E sites compared to telomere sites for both full length and DBD protein.

Dissociation constants and competition assays were also performed on the K360A mutant. A sample $K_d$ determination is shown in Figure 26A. A gradual shift of label is observed with these mutations. $K_d$ values for binding to a telomeric site were identical to wild-type protein for both DBD and full length mutant. HML-E gave affinities 6-fold less for full length and 10-fold less for DBD.

4.2. K446A and K360A competition assays

Competition assays with K446A revealed comparable results to wild-type Rap1. A sample assay is shown in Figure 25B. A 2-fold increase of IC$_{50}$ values from $1 \times 10^{-8}$ M for telomere to $2 \times 10^{-8}$ M for HMR-E was observed. HML-E gave a 7-fold increase for IC$_{50}$. Values
Figure 23. Three dimensional model of the K446A mutation

(A) Model of Rap1 DBD bound to a telomeric site. Residue Lys-446 is shown in blue. This residue makes a direct hydrogen bond contact to adenine at base 10 via the minor groove.
(B) Putative position of alanine 446 in the site-directed mutant. Residue 446 is shown in blue. This mutant would be unable to make any direct base contacts, yet is unlikely to disrupt the DBD three-dimensional structure.

Figure 24. Three dimensional model of the K360A mutation

(A) Model of Rap1 DBD bound to a telomeric site. Residue Lys-360 is shown in blue (arrow). This residue makes a direct hydrogen bond contact to adenine via the minor groove.
(B) Putative position of alanine 360 shown in blue in the site-directed mutant (arrow). This mutant would be unable to make any direct base contacts, and is unlikely to disrupt the DBD three dimensional structure.
Figure 25. Interactions of K446A protein with telomeric and silencer binding sites under various conditions

Full length K446A and a K446A DBD mutant were subjected to various studies. (A) Kd determination of the protein. A sample mobility shift used to determine the Kd of full length K446A is shown at left. Values were obtained for binding to both a telomere and HML-E site. See Materials and Methods, section 4.2. (B) Competition analysis of K446A binding. K446A protein and labeled telomeric binding site were subjected to mobility shift reactions with increasing concentration (10^{-8}M, 3x10^{-7}M, and 5x10^{-7}M) of cold telomeric or silencer site. A sample K446A DBD competition assay is shown at left. IC_{50} values were calculated from the mean of three assays by determining the molarity of unlabeled duplex required to halve the control amount of shifted label. See Materials and Methods, section 4.3. (C) CAT assay analysis of K446A. A K446A expression plasmid was co-transfected with telomeric or silencer reporter plasmids into COS-7 cells, and the level of acetylated product was determined from three assays. Mean percent acetylation was calculated, along with standard deviation values (shown as error bars). Control assays were used to zero mean percent acetylation, giving a control value of zero. See Materials and Methods, section 5.
Results

<table>
<thead>
<tr>
<th></th>
<th>IC_{50} (M)</th>
<th>K446A</th>
<th>K446A DBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEL</td>
<td>1X10^{-8}</td>
<td>2X10^{-8}</td>
<td></td>
</tr>
<tr>
<td>HMR-E</td>
<td>2X10^{-8}</td>
<td>3X10^{-8}</td>
<td></td>
</tr>
<tr>
<td>HML-E</td>
<td>7X10^{-8}</td>
<td>1X10^{-7}</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Kd (M)</th>
<th>K446A</th>
<th>K446A DBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEL</td>
<td>6X10^{-11}</td>
<td>6X10^{-11}</td>
<td></td>
</tr>
<tr>
<td>HML-E</td>
<td>1X10^{-10}</td>
<td>1X10^{-10}</td>
<td></td>
</tr>
</tbody>
</table>
Results

Figure 26. K360A protein interactions with telomeric and silencer binding sites under various conditions

Full length K360A and a K360A DBD mutant were subjected to various studies. (A) Kd determination of the protein. A sample mobility shift used to determine the full length K360A Kd is shown at left. Values were obtained for binding to both a telomere and HML-E site. (B) competition analysis of K360A affinity. K360A protein and labeled telomeric binding site were subjected to mobility shift reactions with increasing concentration (10^{-4}M, 3x10^{-7}M, and 5x10^{-7}M) of cold telomeric or silencer site. A sample full length K360A competition assay is shown at left. IC_{50}’s were calculated from the mean of three assays by determining the molarity of cold oligo required to halve the control amount of shifted label. See Materials and Methods sections 4.2 and 4.3 for procedures.
Results

<table>
<thead>
<tr>
<th>Kd (M)</th>
<th>K360A</th>
<th>K360A DBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEL</td>
<td>1X10^{-11}</td>
<td>1X10^{-11}</td>
</tr>
<tr>
<td>HML-E</td>
<td>6X10^{-11}</td>
<td>1X10^{-10}</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IC_{50} (M)</th>
<th>K360A</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEL</td>
<td>1X10^{-8}</td>
</tr>
<tr>
<td>HMR-E</td>
<td>2X10^{-8}</td>
</tr>
<tr>
<td>HML-E</td>
<td>3X10^{-8}</td>
</tr>
</tbody>
</table>
Results

for the DNA binding domain were comparable, with a 1.5-fold increase for HMR-E and a 5-fold increase for HML-E binding site.

Competition assays were also performed on the full-length version of Rap1 K360A (Figure 26B). Values were comparable to those obtained with the wild-type protein (Figure 17). HMR-E gave 2-fold less affinity, while HML-E gave 3-fold less affinity when compared to a telomeric site.

4.3. K446A CAT assay

As shown in Figure 25C, interactions with the telomeric binding site was similar to wild-type protein, though *in vivo* activity as determined by CAT assay was roughly 30% less than wild-type when using a telomeric binding site. Full length K446A gave comparable values using HMR-E and HML-E reporter plasmids to those of wild-type Rap1. Results were also obtained using a DNA binding deletion of K446A. In this instance, telomeric percent acetylation values were the same as those for HMR-E and HML-E.

CAT assays were not performed using the K360A mutant.

5. Mutants deleting the residue 588-594 region and residue 418 to 438 region affect DNA binding

Deletion analysis to determine the Rap1 DNA binding domain (Henry et al., 1990) as well as Rap1 crystal structure data (Konig et al., 1996) suggest that residues 588 to 594 play a role in Rap1 DNA binding. These residues form a C-terminal “tail” which interacts with the “core” region of the telomeric site via the major groove. To test whether this interaction is essential to the functioning of Rap1, these residues were removed from the full-length protein. This “tail” region is highlighted in red in Figure 27B. These residues make base-specific contacts via the major groove, and also make protein-protein contacts with HTH1.

A region conserved between *S. cerevisiae* and *K. lactis* Rap1 lies between HTH1 and HTH2. This region contains an omega loop, but has no known function. To test the importance of this region, residues 418 to 438 were removed (shown in red in Figure 27A). This “loop” region was deleted from the DBD and replaced by a glycine. According to the crystal structure, the angle between the backbone of residues 418 and 438 is approximately 20° with a distance between the residues similar to that of an amino acid insertion. This suggests that a deletion of
Figure 27. Three-dimensional model of deletions 418 to 438 and 588 to 594.

Model of Rap1 DBD bound to a telomeric site. (A) Residues 418 to 438, comprising the “loop” region, are shown in red (arrow). These residues were removed by site directed mutagenesis, and replaced by a glycine used to link residues 417 and 439. This mutant should maintain the three dimensional structure of the DBD. (B) Residues 588 to 594, comprising the “tail” region, are shown in red (arrow). These residues were removed by site directed mutagenesis from a full length version of the RAP1 protein. This would likely remove all protein-DNA interactions in this region of RAP1.
residues 418 to 438 with the insertion of a glycine would not significantly distort the three dimensional structure of the DBD.

5.1. Determination of $K_d$ for 588 to 594 deletion

Dissociation constant values were obtained for full length $\Delta588$-594 deletion mutants to both telomere and HML-E sites. A sample mobility shift is shown in Figure 28A. Binding to a telomeric site displayed a $K_d$ of $2 \times 10^{-11}$ M, 2-fold weaker than the affinity of the wild-type protein. Binding to a HML-E site was further reduced, showing at 10-fold less affinity when compared to wild-type protein.

5.2. Competition assay

IC$_{50}$ values were obtained for full length Rap1 with a deletion of residues 588 to 594 and full length Rap1 with deletion of residues 418 to 438 (Figure 28B). The $\Delta588$-594 deletion showed IC$_{50}$ values comparable to wild-type protein. Both HMR-E and HML-E telomeres gave values of $3 \times 10^{-8}$ M, 3-fold less than unlabelled telomere.

$\Delta418$-438 deletions gave 2-fold less affinity for HMR-E and 10-fold less affinity for HML-E as determined by IC$_{50}$.

5.3. CAT assay

The ability of these two mutants to activate a reporter gene in vivo was also assayed. Figure 28C details the result of these experiments. The $\Delta588$-594 deletion of full length Rap1 results in low levels of acetylation using all three binding sites. Similar low levels are also seen for “loop” deletions, in this case tested using the DBD version of the mutant protein. While this region is not essential for in vitro binding, it does effect the ability of Rap1 to bind at wild-type levels in vivo.

6. The effect on binding of a G590V, R404K Rap1 double mutant

The affinity of the R404K mutant for DNA suggests that this mutation, while it disrupts hydrogen bonding, does not significantly reduce the affinity of Rap1 for a telomeric site. One possibility is that the HTH1 interaction is stabilized by the residues 588 to 594, which interacts
**Figure 28. Interactions between “Tail” and “loop” deletions and Rap1 binding sites under various conditions**

Full length Rap1 was mutated such that amino acids 588 to 594 of the protein (the “tail” region) were removed. In another set of mutations, residues 418 to 438 (the “loop” region) were removed from either full length or the DBD fragment of Rap1. These mutations were then subjected to various studies. (A) Kd determination of the Δ588-594 protein. A sample mobility shift used to determine the full length Δ588-594 Kd is shown at left. Values were obtained for binding to both a telomere and HML-E site. See Materials and Methods, section 4.2. (B) competition analysis of full length Δ588-594 and full length Δ418-438. Both proteins were incubated with labeled telomeric binding site and subjected to mobility shift reactions with increasing concentration of unlabeled telomeric or silencer site. A sample competition assay with Δ588-594 is shown at left. IC$_{50}$'s were calculated from the mean of three assays by determining the molarity of unlabeled competitor DNA required to halve the control amount of shifted label. See Materials and Methods, section 4.3. (C) CAT assay analysis of full length Δ588-594 and the DBD deletion of Δ418-438. Expression plasmids were co-transfected with telomeric or silencer reporter plasmids into COS-7 cells, and the level of acetylated product was determined from three assays. Mean percent acetylation was calculated, along with standard deviation values (shown as error bars).
Results

<table>
<thead>
<tr>
<th>Kd (M)</th>
<th>Δ588-594</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEL</td>
<td>2X10⁻¹¹</td>
</tr>
<tr>
<td>HML-E</td>
<td>1X10⁻¹⁰</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IC₅₀ (M)</th>
<th>Δ588-594</th>
<th>Δ418-438</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEL</td>
<td>1X10⁻⁸</td>
<td>5X10⁻⁹</td>
</tr>
<tr>
<td>HMR-E</td>
<td>3X10⁻⁸</td>
<td>1X10⁻⁸</td>
</tr>
<tr>
<td>HML-E</td>
<td>3X10⁻⁸</td>
<td>5X10⁻⁸</td>
</tr>
</tbody>
</table>
with the same sequence via the major groove, as well as forming a protein-protein contact to HTH1.

In order to test the possibility of cooperative association, a double mutant disrupting the “tail” region and R404K was designed. Replacing the glycine at residue 590 with a valine disrupts the “tail” region. This alteration is modeled in Figure 29. Gly-590 allows the protein backbone to protrude into the major groove, where it makes a base-specific hydrogen bond (Figure 29A). Valine, a bulkier hydrophobic residue, would disrupt this interaction (see Figure 29B). Valine should push the residue 588-594 region away from the major groove, disrupting DNA binding as well as interaction with HTH1.

6.1. G590V \(K_d\) determination

At first, \(K_d\) values for a single mutant containing only G590V were determined (see Figure 30A). A sample mobility shift is shown at left. A gradual shift was observed. These results appear to correspond with \(\Delta 588-594\) studies. While there were problems of weak signal with this and other assays, all trials show identical shift patterns. \(K_d\) values for full length G590V are comparable to \(\Delta 588-594\) \(K_d\) values, with a 2-fold decrease in affinity for a telomeric site compared to wild-type protein. A 6-fold (for full length) and 7-fold (for DBD) decrease in affinity are observed when compared to wild-type for binding to a HML-E site. These values are comparable to that of \(\Delta 588-594\), suggesting that G590V does disrupt interactions with DNA involving residues in the 588 to 594 region.

6.2. G590V competition assay

The IC\(_{50}\) was determined on full length G590V. A sample assay is shown at left (Figure 30B). Molarity was the same using all three binding sites, at \(2\times10^{-8}\) M, suggesting similar affinity to all three sites. Again, these values are comparable to \(\Delta 588-594\) deletions.

6.3. Determination of \(K_d\) for the R404K/G590V double mutant

\(K_d\) values for the G590V R404K double mutant binding to a telomeric site and wild-type Rap1 are shown in Figure 30C. The double mutant maintains high affinity binding to a telomeric site, at \(2\times10^{-11}\) M for full length protein. Binding of the DBD is at \(3\times10^{-11}\) M, a 3-fold difference compared to wild-type DBD dissociation. The difference in binding affinity is not
Figure 29. **Three dimensional model of the G590V mutation**

(A) Model of Rap1 DBD bound to a telomeric site. Residue Gly-590 is indicated by an arrow. This residue makes a direct hydrogen bond contact to cytosine via the major groove. (B) Putative position of valine 590 in the site-directed mutant. Residue 590 is shown in blue. This mutant would disrupt hydrogen bonding and, as a bulky hydrophobic residue, valine would likely disrupt the "tail" regions three dimensional structure.
Results

A

B
Figure 30. G590V and G590V/R404K double mutant protein interactions with telomeric and silencer binding sites under various conditions

Full length G590V and a G590V DBD mutant were subjected to Kd and IC₅₀ analysis. (A) Kd determination. A sample mobility shift used to determine the Kd of full length G590V is shown at left. Values were obtained for binding to both a telomere and HML-E site. See Materials and Methods, section 4.2. (B) competition analysis of G590V. G590V protein and labeled telomeric binding site were subjected to mobility shift reactions with increasing concentration (10⁻⁸M, 3x10⁻⁷M, and 5x10⁻⁷M) of cold telomeric or silencer site. A sample full length competition assay is shown at left. IC₅₀'s were calculated from the mean of three assays by determining the molarity of unlabeled duplex required to halve the control amount of shifted label. See Materials and Methods, section 4.3. (C) Kd determination of the double mutant. The DBD deletion of G590V was subjected to a second round of site directed mutagenesis, creating a G590V/R404K double mutant protein. Dissociation constants were determined with both a telomeric and HML-E site. For comparison values obtained for wildtype Rap1 and its DBD are also shown.
Results

**A**

<table>
<thead>
<tr>
<th>Kd (M)</th>
<th>G590V</th>
<th>G590V DBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEL</td>
<td>2X10^{-11}</td>
<td>2X10^{-11}</td>
</tr>
<tr>
<td>HML-E</td>
<td>6X10^{-11}</td>
<td>7X10^{-11}</td>
</tr>
</tbody>
</table>

**B**

<table>
<thead>
<tr>
<th>IC_{50} (M)</th>
<th>G590V</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEL</td>
<td>2X10^{-8}</td>
</tr>
<tr>
<td>HMR-E</td>
<td>2X10^{-8}</td>
</tr>
<tr>
<td>HML-E</td>
<td>2X10^{-8}</td>
</tr>
</tbody>
</table>

**C**

<table>
<thead>
<tr>
<th>Kd (M)</th>
<th>R404K G590V</th>
<th>R404K G590V DBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEL</td>
<td>2X10^{-11}</td>
<td>2X10^{-11}</td>
</tr>
<tr>
<td>HML-E</td>
<td>6X10^{-11}</td>
<td>7X10^{-11}</td>
</tr>
</tbody>
</table>
significant, but was consistent between experiments. The two mutations appear to have an
additive rather than synergistic effect on the DNA binding activity of Rap1. Since R404K was
already shown to disrupt DNA binding to a silencer, no estimates of $K_d$ values were attempted
for binding to HMR-E and HML-E sites.

7. Affinity of Rap1 for TEL$^{A3}$ and HMR-E$^{A10/C11}$ binding sites

Our results indicate that various mutants, especially R404K, have reduced affinity for
silencer, but not telomeric sites. In order to determine the DNA base sequence responsible for
differences in Rap1 affinity, two altered binding sites were assayed in inhibition/competition
experiments (Figure 31). TEL$^{A3}$ replaces the third base of the telomere sequence with an
adenine (Figure 31A). This mutation alters the 5' end of the telomere site to resemble the
HML-E and HMR-E binding sites.

A second altered site, HMR-E$^{A10/C11}$, changes bases 10 and 11 of HMR-E, so that the 3'
end of the site resembles a telomeric site (Figure 31A). These alterations compensate for
differences between HMR-E and telomere sites.

Wild-type Rap1 and R404K proteins were assayed (Figure 31B) for IC$_{50}$. The wild-type
protein gave IC$_{50}$ values for TEL$^{A3}$ and HMR-E$^{A10/C11}$ identical to values for TEL and HMR-E.
R404K protein gave values of 3x10$^{-8}$ M for both TEL$^{A3}$ and HMR-E$^{A10/C11}$. The IC$_{50}$ measured
for R404K TEL$^{A3}$ is 3-fold higher than that of R404K TEL, while the IC$_{50}$ measured for HMR-
E$^{A10/C11}$ is roughly 3-fold lower than that for R404K HMR-E. These results indicate that
alterations at the HTH2 binding site of HMR-E result in higher affinity by the R404K mutant.
Conversely disruption of the HTH1 binding site of TEL results in a lowered affinity, suggesting
that binding of Rap1 HTH1 and HTH2 are linked in some manner.

SECTION C. STUDIES OF THE HTH1 BINDING MOTIF

1. Expression of the HTH1 motif

Previous studies of HTH-variant DNA binding proteins suggest that often two HTH
domains are required for DNA binding at high affinity (Verrijzer et al., 1990; Ogata et al., 1994;
Li et al., 1995). Diverse HTH motifs (including c-Myb, the MATa1/MATa2 heterodimer and
the oct-1 Homeodomain) require two HTH motifs to bind DNA. A single motif often binds only
A

<table>
<thead>
<tr>
<th></th>
<th>TEL</th>
<th>TELA³</th>
<th>HMR-E</th>
<th>HMR-EA¹⁰/C¹¹</th>
</tr>
</thead>
</table>

CONSENSUS: T A/G C A C C C A N N C A/C C C

B

<table>
<thead>
<tr>
<th></th>
<th>IC₅₀ (M)</th>
<th>Rap1</th>
<th>R404K</th>
</tr>
</thead>
<tbody>
<tr>
<td>TELA³</td>
<td></td>
<td>1X10⁻⁸</td>
<td>3X10⁻⁸</td>
</tr>
<tr>
<td>HMR-EA¹⁰/C¹¹</td>
<td>1X10⁻⁸</td>
<td></td>
<td>3X10⁻⁸</td>
</tr>
</tbody>
</table>

Figure 31. Competition analysis of altered TEL and HMR-E binding sites

(A) Comparison of the various sequences. The C at base three on a telomeric site was altered to an A, converting the 5' end of a TEL site to a HMR-E site. As well, C at base 10, and A at base 11 of a HMR-E site were altered to A and C respectively, converting the 3' end of a HMR-E site to a TEL site. (B) IC₅₀ values for Rap1 and R404K were determined as described in section 4.2 of Materials and Methods.
at low affinity, or in a non-specific manner. Data presented in this thesis show that deletion of residues 588-594, previously thought to be essential for DNA binding, results in a protein still able to bind telomeric DNA \textit{in vitro}. It was decided to determine if a single Rap1 HTH motif is also capable of binding DNA specifically.

In order to determine the DNA binding affinity of Rap1 HTH1, this motif was expressed in \textit{E. coli} cells and the crude extract partially purified by Heparin Sepharose chromatography (see Materials and Methods). The presence of HTH1 was determined by western blot (Figure 32A), showing as a 6 kDa band. Yield was estimated at less than 1% in the crude extract, centrifuged to remove any insoluble material. We were unable to assay folding of the fragment.

2. Assaying HTH1 DNA binding affinity by cross-link mobility shift analysis

Standard mobility shift assays of HTH1 with a telomeric binding site did not produce a complex, so a laser cross-linking experiment was carried out to determine if any specific interaction between HTH1 and a telomeric site could be observed. After incubation of probe and HTH1 protein, the assay mix was cross-linked (see Materials and Methods). The mixture was then separated on SDS-PAGE and exposed. Results are shown in Figure 32B. In the presence of poly dIdC, a non-specific DNA competitor (lanes 1-3), no shifting of radioactive probe was observed. However, in the absence of poly dIdC (lanes 4-6), a band at 29 kDa was observed (lane 4), which was both shifted in the presence of Rap1-specific antibodies (lane 5), and removed by the addition of excess unlabelled telomeric site (lane 6). There is a hint of non-specific trapping of probe by antibody (lane 2), but this would not remove the band observed at 29 kDa.

3. \textit{In vitro} PKC phosphorylation of DBD and HTH1 deletions

Previous work in the Roberge lab (Yamazaki, 1995) indicates that the Rap1 DBD contains two heart muscle kinase (HMK) phosphorylation sites. These sites affect the DNA binding activity of DBD to a TEF2 binding site. To determine if these phosphorylation sites are located within HTH1, a partially purified HTH1 motif was subjected to phosphorylation by HMK. Figure 33 shows an SDS-PAGE of both DBD and HTH1 after phosphorylation by HMK and exposure to a Phosphor Imager screen. Both lanes contain equivalent amounts of the respective Rap1 protein, as determined by Coomassie blue staining (data not shown). Lane 1,
Results

Figure 32. Analysis of HTH1 DNA binding affinity

Residues 359 to 415 of Rap1 were expressed in *E. coli* and partially purified by heparin sepharose chromatography. See Materials and Methods, section 3.3. (A) Crude fractions from the heparin sepharose purification were separated on a 15% SDS-PAGE and HTH1 was detected by western blotting with Rap1 antibody. Lanes 5 and 6 contain a band at 6 kd corresponding to HTH1. (B) Cross linking assay. The fractions containing HTH1 were combined and 1µl placed in a 10µl mobility shift reaction with 10^{-10} M labeled telomeric binding site. After incubation, reactions were laser cross-linked (see section 4.4, Materials and Methods) and separated on a 10% SDS-PAGE gel. The gel was then dried and exposed to film, revealing any protein-DNA complexes. Lanes 1-3 contain poly dl-dC, which was absent in lanes 4-6. Lanes 2 and 5 contain 1 µl of Rap1 antibody, lanes 3 and 6 contain 10^{-8} M unlabeled telomeric oligo duplex.
Results

Figure 33. Rapl DBD and HTH1 Heart Muscle Kinase Reactions.

Partially purified Rapl DBD and Rapl HTH1 (both proteins at approximately $10^{-10}$ M) were subjected to PKC phosphorylation with bovine Heart Muscle Kinase in the presence of radiolabeled ATP. Equivalent amounts of DBD and HTH1 were assayed, as measured by western blot. See Materials and Methods, section 8. Lane 1, DBD phosphorylation: A band at 29 kd corresponding to the molecular weight of DBD was visible after exposure to film. Lane 2, HTH1 phosphorylation: A weak band at 6 kd, corresponding to the molecular weight of HTH1 was visible after exposure.
Results

containing DBD, displays a strong signal at 29 kDa, corresponding to the DBD protein. In lane 2, only a weak signal is visible at 6 kDa corresponding to HTH1.
IV. DISCUSSION

1. PRIMARY SEQUENCE OF THE RAP1 DBD AND STUDIES OF RAP1 DNA BINDING CAN BE USED TO MODEL A BIPARTITE HELIX-TURN-HELIX DNA BINDING DOMAIN

Even in the absence of three-dimensional data, it is often possible to identify the structural motif of a particular DNA binding domain through primary sequence comparison. However, if no regions of amino acid conservation are observed, biochemical and genetic analysis of the protein can still play a role in pinpointing key residues and regions of the protein involved in DNA interaction. Since the crystal structure of the Rap1 DBD-Tel Complex was not known at the beginning of my studies, I first used a comparative approach to elucidate 2 key residues involved in DNA binding.

I hypothesized that Arg-404 and Arg-408 play a role in DNA binding through a Myb-like HTH. As well, Arg-542 and Arg-546 possibly play a role in a second helix-turn-helix motif. The conservation of key residues between Rap1 HTH1 and the Myb HTH motif was crucial in determining whether to proceed with site-directed mutagenesis studies to test the model. This model suggested that Rap1 DBD contains a bipartite DNA binding domain, consistent with biochemical data (Gilson et al., 1993). The separation of the HTH1 and the HTH2 recognition helices suggested that other motifs may be present between HTH1 and HTH2. Alternatively, structures may loop from the HTH2 ‘turns’, forming a structure similar to the HNF-3γ DBD (Brennan, 1993).

Site-directed mutants confirmed that these residues play a role in DNA binding at a TEF2 binding site. Mutations of arginine to an aspartate (R404D and R408D) reduced binding, while mutant R408N was still able to bind the TEF2 site. These data were consistent with the model: mutation from wild-type to the corresponding Myb residue results in weak changes in affinity, while mutation from a basic to an acidic residue results in strong changes to binding affinity. Dissociation constants of the mutants bound to a telomeric site suggest that these mutations do not affect TEL DNA binding in the manner they do TEF2. $K_a$ values do follow the trend of TEF2 binding, with R408N exhibiting the strongest affinity, R404D the weakest. The ability of all mutants to bind at high affinity (from $1 \times 10^{-11}$ to $3 \times 10^{-10}$ M) to a TEL site
suggests other DNA binding motifs are involved. If HTH1 (the Myb-like domain) was the only domain required, $K_d$ values should have risen to the Rap1 non-specific levels of $8.7 \times 10^{-6}$ M (Vignais et al., 1990). These observations, as well as biochemical data, led to the proposal of the HTH2 domain.

The crystal structure of the DBD bound to a telomeric site validated this model. HTH1 followed the prediction from the Myb motif. The HTH1 hydrophobic cluster includes Phe-363, Phe-370, Ile-371, Val-375, Tyr-388, Ile-391, Val-395, Ile-403, and Phe-407 (Konig et al., 1996). All of these residues were predicted from Myb. Leu-411, predicted to form the hydrophobic cluster from the Myb structure, is located outside of the HTH1 $\alpha$-helix. This residue still interacts with the hydrophobic cluster, however.

The residues for the HTH2 recognition helix also followed predictions based on the sequence of Myb. Trp-541, Phe-545 and Leu-550 all form the hydrophobic core of HTH2 (Konig et al., 1996). An additional core component is Leu-551, which compensates for the absence of a predicted hydrophobic residue at Arg-544.

A number of novel structures were not evident by sequence analysis, and not predicted by our model, notably the “tail” and “loop” structures. As well, base-specific hydrogen bond interactions by Asn-401, which are more significant than bonding by Arg-408, were not predicted by the model. Konig et al. (1996) suggest that residues His-405 and Val-409 could affect the protein-DNA interface of Arg-408, since on HTH2 the corresponding residue Arg-546 makes two base-specific hydrogen bonds.

2. MODELING OTHER HTH DNA BINDING DOMAINS - VARIOUS ISSUES

The sequence and spacing of hydrophilic residues is often widely varied in HTH motifs. All residues are capable of hydrogen bonding with DNA bases via the major groove (Suzuki, 1994). As we note with the Rap1 “tail” motif, as long as the region is not formed into a helix, even glycine and proline residues can make base-specific DNA contacts via the major groove. This variability makes Helix-Turn-Helix model predictions from hydrophilic residues extremely difficult, and predictions of binding site sequence from amino acid data, or vice-versa, highly speculative.
In contrast, conservation of the spacing and function of hydrophobic residues involved in HTH1 stability (residues forming a hydrophobic cluster linking the helices of HTH1) provides a quantifiable means of determining whether a sequence under scrutiny contains a HTH motif and where a putative recognition helix would be located. Currently there is no computer algorithm capable of finding de novo HTH motifs (aside from alignment of related proteins), and considering the importance of biochemical and genetic data to verify any models, matches would likely be more dependent on detailed study of DNA binding domains determined by deletion analysis.

Despite recent advances in determining protein three dimensional structure, including heteronuclear multidimensional NMR, the majority of known DNA binding proteins has not been subjected to three dimensional analysis. Many of these proteins show identity to a family of DNA binding motifs, allowing assumptions to be made about their DNA binding mechanism from the structure of related proteins. This leaves a number of DNA binding proteins without recognizable DNA binding motifs as well as proteins of unknown function, which may interact with DNA. The ability to determine key residues located in the recognition structure of a DNA binding domain, before determining the three dimensional structure of the domain, would allow researchers to study the DNA binding mechanism of the protein. Research could include the creation of site-directed mutants to determine key amino acid residues involved in DNA recognition.

3. WILD-TYPE RAP1 HAS SIMILAR AFFINITY TO TELOMERIC AND SILENCER BINDING SITES IN VITRO

Differences in the apparent binding affinity of Rap1 to a telomeric site compared to a TEF2 binding site led us to study differences in affinity between telomeric and silencer binding sites. A summary of the data from these various experiments is shown in Figure 34.

Studies comparing affinity of Rap1 to telomeres and silencers were first performed using the wild-type protein. Very little difference in affinity was seen in vitro, with Rap1 binding at high affinity to all sites. A 3-fold increase in competition values using HML-E binding site was the only change observed.
A general summary of experimental results is shown above. Following analysis of the Kd, IC\textsubscript{50} and CAT activity data, a value was assigned to each Rap1 species bound to a particular Rap1 binding site. Values are based on a qualitative comparison of results for particular mutants with results for wildtype protein. Only R404 and R408 mutants were assayed using the TEF2 DNA binding site.

### Figure 34. Summary of Rap1 binding affinity data

<table>
<thead>
<tr>
<th></th>
<th>TEL</th>
<th>HMR-E</th>
<th>HML-E</th>
<th>TEF2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wildtype</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td>R404K</td>
<td>++</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>R404D</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>R408N</td>
<td>+++</td>
<td>+++</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>R408D</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>K360A</td>
<td>+++</td>
<td>+++</td>
<td>++</td>
<td>N/A</td>
</tr>
<tr>
<td>K446A</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>N/A</td>
</tr>
<tr>
<td>Δ588-594</td>
<td>+++</td>
<td>++</td>
<td>+</td>
<td>N/A</td>
</tr>
<tr>
<td>Δ418-438</td>
<td>+++</td>
<td>++</td>
<td>+</td>
<td>N/A</td>
</tr>
<tr>
<td>G590V</td>
<td>+++</td>
<td>++</td>
<td>++</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Considering there is a total of six base alterations between HML-E and TEL DNA, one would expect more dramatic differences in their affinity. However, of these six bases only the 3’ T and A are non-consensus at HML-E. Thus differences should be confined to this site. These 3’ base changes would be expected to alter binding of HTH2, specifically through contacts by Arg-542 and Arg-546. These residues provide three of five direct base-specific hydrogen bonds. Thus our results suggest that disruption of HTH2 base-specific contacts does not significantly alter the binding affinity of wild-type Rap1, consistent with the results of Graham and Chambers (1994).

The DBD deletion product provided equivalent data for both \( K_d \) and competition assays. This suggests that wild-type Rap1 has the ability to bind to a variety of sites with a similar affinity, with or without the N- and C-terminal portions of the protein. Thus DNA bending by the N-terminal portion of Rap1 is not involved in enhancing its affinity for DNA, but is more likely a non-specific protein-DNA interaction.

4. \textit{IN VIVO STUDIES SUGGEST RAP1 HAS VARIED AFFINITIES FOR ITS BINDING SITES WITHIN THE YEAST CELL}

The use of the CAT assay system has allowed us to compare the affinity of Rap1 to telomeric and silencer sites in an \textit{in vivo} environment. Preliminary experiments to set up optimal CAT assay conditions suggested that this assay is quite sensitive to changes in DNA binding affinity. Transfection of 10-fold less expression/reporter plasmid than in our studies resulted in only background levels of acetylation, and a minimum of five Rap1 binding sites was required.

Using a telomeric site and a wild-type Rap1-VP16 hybrid, CAT assays give high levels of activity. This suggests that Rap1 would bind to telomeres under similar conditions in the yeast cell. Silencer binding sites at HMR-E and HML-E gave noticeably less activity. If conditions in yeast are comparable to those in Cos-7 cells, one would expect lower affinity binding of Rap1 to silencers. This is not necessarily likely, and should be tested in a yeast system.

Fluorescence microscopy experiments indicate that Rap1 is primarily found in the telomeric region on the nuclear periphery. Telomeres sequester Rap1 from other areas of the genome, reducing the pool of Rap1 outside of chromosome ends. The findings of Buck and
Shore (1995) suggest that telomeres compete with HMR-E binding sites for available Rap1 protein. This competition was predicted to result from the large number of Rap1 sites on telomeres (Gilson and Gasser, 1995). Our in vivo data suggest another possibility. Sequestering of Rap1 by telomeres may also be the result of their higher affinity for Rap1. The ability of telomeres to bind Rap1 strongly, combined with the large number of Rap1 sites would counteract the tendency of Rap1 to diffuse throughout the nucleus. There are roughly 8,000 Rap1 molecules per cell (Gilson and Gasser, 1995) and 20% of these could be bound at telomeres (given the length of the telomeres and one Rap1 molecule bound per 18bp). Therefore a low telomeric $K_d$ would help maintain a sharp concentration gradient, keeping the majority of Rap1 molecules around the telomeres.

For the various Rap1 DNA binding mutants, in vivo studies gave consistently low levels of CAT activity at all three binding sites. One exception was the K446A mutant, which had reduced levels of TEL binding when compared to wild-type protein. These results suggest that the CAT assay was not able to quantify differences in binding to the three sites. CAT assays would need to be less sensitive to changes in affinity. Unfortunately, experiments using more Rap1 binding sites on the CAT promoter were unable to clarify the issue. Use of a different mammalian system may be required to characterize these mutants in more detail.

A number of differences can be expected between the CAT assay and in vitro assays to determine $K_d$. In the in vitro system, Rap1 was expressed in E. coli, and would be the unphosphorylated form of the protein. Rap1 expressed within COS cells would be phosphorylated depending on the possible kinase sites on the protein. As well, the ionic environment within COS cells differs from the controlled environment used for the various mobility shift experiments. These and other differences, including non-specific DNA binding proteins and DNA supercoiling, may have affected the binding of Rap1 between the two experimental systems.

5. MUTATION OF HTH1 AFFECTS BINDING TO RAP1 SITES WITH ALTERED 3' BINDING SEQUENCES

The various mutations at Arg-404 and Arg-408 displayed moderately reduced affinity to a telomeric site (up to $3 \times 10^{-10} \text{ M}$) but minimal binding to a HML-E site. This effect was seen
for both $K_d$ and competition data. Qualitative data from TEF2 binding suggest that this site also binds Arg-404 and Arg-408 mutants very weakly. Deletion of the "tail" region and G590V mutants show that this effect is independent of the C-terminal portion of DBD.

The most significant difference between the HML-E and telomere sites is at the 3' end of the sequence. HML-E contains a T A -3' while TEL, HMR-E and the consensus site contain C C -3'. Tests with TEL$^3$ and HMR-E$^{A10C11}$ binding sites confirm only these two 3' bases are able to affect binding. That TEF2 is unable to bind mutants and contains T T at its 3' end is consistent with this model. These results suggest that HTH1 binding is affected by changes at the 3' end, i.e. HTH1 mutations coupled with disruption of HTH2 binding are required to alter the high affinity binding of the Rap1 protein. This is discussed below.

6. MINOR GROOVE INTERACTIONS HELP STABILIZE RAP1-DNA BINDING IN VITRO AND IN VIVO

Mutation K446A, which affects N-terminal minor groove interaction by HTH2, has only a moderate effect on DNA binding. This was seen with both in vivo and in vitro studies of the protein. HML-E binding was reduced only 10-fold by the mutation, compared to a 100-fold reduction with Arg-404 mutants. K446A reduces TEL binding 6-fold as well, giving less than a 2-fold difference in affinity between the TEL and HML-E sites.

Thus K446A, which is a HTH2 mutation, does not produce a significant effect on changes at the 3' end of the binding site. This suggests that at the HML-E site, binding is already sufficiently disrupted that the K446A mutation can exert only minimal effect.

Mutation K360A, affecting N-terminal minor groove interaction by HTH1, has no effect on TEL binding, but does affect binding to a HML-E site. Competition assays show moderate differences between TEL and HML-E site IC$_{50}$ levels. These results suggest that mutation of the N-terminal site does not dramatically alter binding. Overall, the results suggest that these sites help stabilize association of the HTH motif and its binding site, but are not essential for binding. The consistent use of N-terminal minor groove interactions in HTH variant DBDs (Engrailed, MATa2, POU, and Paired DBD all contain N-terminal arms) confirm that this interaction is integral to the binding stability of these HTH motifs, even if they are not essential for binding.
7. THE DBD “TAIL” REGION IS NOT ESSENTIAL FOR RAP1 DNA BINDING

While it may play a role in the optimal binding of Rap1 to its sites in vivo, the C-terminal tail region of the Rap1 DBD (residues 588-594) can be removed without significantly affecting binding to telomerases and silencers in vitro. When determining the minimal binding domain of Rap1, Henry et al. (1990) tested Rap1 binding with the Z+ DNA fragment from the PGK promoter (sequences -473 to -409 of the promoter containing the Rap1 site C A C C C A C A C A T T). This fragment appears to be a relatively weak binding site when compared to HMR-E or TEF2 sites in vitro (Henry et al., 1990). This would suggest that the minimal DNA binding domain of Rap1 can be reduced to residue 580, or even residue 562, and still maintain high affinity binding to a telomeric site in vitro.

At the same time, conservation of the “core” ACC is found at all known in vitro Rap1 binding sites. This sequence, which interacts with both HTH1 and the “tail”, is much more invariant than the HTH2 binding sequence. If the “tail” was not involved in binding, one would expect Rap1 to bind sites with either a degenerate “core” ACC or a degenerate 3’ ACC, but not both. Instead variation occurs only at the 3’ ACC (Graham and Chambers, 1994).

8. THE DBD “LOOP” REGION AFFECTS DNA BINDING AFFINITY

Removal of the “loop” region (residues 418-438) results in competition values comparable to that of wild-type protein (5x10^-9 M to 5x10^-8 M) but in vivo values are reduced for a telomeric site, from the range of 75% for wildtype to the range of 25% for the loop deletion mutant. This suggests that the loop region does exert some stabilizing effects on the protein, though it is not essential to binding in vitro. While the in vivo effect may be an artifact of mutagenesis, conservation of this loop between S. cerevisiae and K. lactis Rap1 points to a structural role for the motif.

Another HTH variant with extended loop between HTH1 and HTH2 is the POU domain (Klemm et al., 1994). In this case the loop is only partially structured, allowing the two Helix-Turn-Helix motifs to clamp around the DNA helix. The two Rap1 HTH motifs lie in tandem, and do not require a flexible linker because wrapping around the DNA helix is not required. Instead the Rap1 DBD has a highly structured loop. What role might this structured loop play...
in the DBD? One possibility is that the loop maintains precise spacing between HTH1 and HTH2, reducing the Gibbs free energy of binding by a chelate effect. This is discussed in more detail below.

In addition, recent studies of Rap1 deletion mutants in vivo (Graham et al., 1999) indicate that deletion of the loop region affects Rap1 activity. Deletion of the loop results in inability to grow in the presence of glucose when wild-type Rap1 is under the control of a GAL promoter. In vitro analysis of this deletion binding to a TEF2 site indicates that the mutant does bind DNA, but with lower affinity than the wild-type protein.

9. RAP1 DNA BINDING INVOLVES COOPERATIVE INTERACTION BETWEEN HTH1 AND HTH2, INDEPENDENT OF THE “TAIL” REGION

In relation to DNA binding domains, cooperativity means that the affinity of a domain for its binding site is enhanced when another domain is bound to an adjacent site (Senear et al., 1998). Where the two domains are part of a single macromolecule, bound by a flexible linker region, contribution to cooperative binding is enhanced by the local concentration effect. Binding of the first domain anchors the second domain in close proximity to its site, increasing the probability of the second binding reaction.

From the research presented in this thesis it is clear that mutations in HTH1 significantly alter binding to silencer, but not telomeric, sites. A summary of the various DNA affinity experiments performed is shown in Figure 34. This result is unusual considering that telomeres and HML-E have identical HTH1 binding sequences. Significant differences are only seen at the 3' end of the sites. This suggests that binding of HTH1 and HTH2 is cooperative in some manner: weak binding at one site being significant only in the context of altered binding at the second site. This effect is independent of the “tail” region. The inability of HTH1 to bind specifically to a telomeric site supports a model where HTH1 and HTH2 bind in a cooperative manner: some form of interaction occurs between the two motifs, providing more than additive affinity of HTH1-HTH2 to DNA. This particular form of cooperativity is seen with the Oct-1 POU domain. In the studies of Klemm and Pabo (1996), the two DNA binding subdomains, POU specific and POU homeodomain, bind in a cooperative manner when the linker between them is removed. \( \Delta G_{\text{cooperativity}} \) gives a value of -1.6 kcal/mole. This cooperative interaction may
occur due to overlapping contacts of the two subdomains at base pair 5 of the Oct-1 binding site and by overlapping phosphate contacts. No protein-protein interactions are observed, discounting any classical protein-protein cooperative interactions. However, comparison of $K_d$ values for the unlinked subdomains ($1.5 \times 10^{-7}$ M for POU homeodomain and $1.7 \times 10^{-6}$ M for POU specific domain) to the $K_d$ value for the intact Oct-1 POU domain ($7.1 \times 10^{-11}$ M), shows that a significant contribution to binding is made by the flexible linker between the two subdomains. This contribution is through the chelate effect, increasing the local concentration of the second subdomain when the first is bound to its site.

Cooperative binding of HTH DNA binding domains has been observed for c-Myb (Ogata et al., 1994) and the MATα1/MATα2 heterodimer (Li et al., 1995). In the case of c-Myb, protein-protein contacts between the two HTH motifs explain the cooperativity observed. For the POU domain, overlapping phosphate contacts are surmised to provide a thermodynamic coupling of the two HTH motifs. In addition, there are overlapping contacts with base pair 5 (A/T) of the Oct-1 binding site. A leucine from the POU-specific domain contacts the A, while an arginine from the POU homeodomain contacts the T. The two subdomains also contact overlapping phosphates that flank base pair 5.

In the case of Rap1 DBD, overlapping phosphate contacts occur at the phosphates between bases 8 and 9 on the backbone. HTH1 residues Thr-386 and His-385 make hydrogen bond contacts to the backbone. At the overlapping site, HTH2 residues Thr-578, Leu-577, Asn-576, Lys-575 and Phe-548 all make hydrogen bond contacts to the backbone. This overlapping of phosphate contacts (bonding to phosphates that are directly opposite across the double helix) could provide the thermodynamic coupling required to enhance binding of the two motifs.

A “connection Gibbs energy” as described by Jencks (1981) may also enhance binding. Connection of the two HTH motifs, especially by the structured linker “loop” region, would provide a chelate effect, increasing the observed energy of binding. Rap1 binding sites typically contain tandem repeats, which mirror the tandem repeat of HTH1-HTH2. Proper spacing of these two motifs would be required for the optimal binding of Rap1. In their study of the MATα1/MATα2 heterodimer, Li et al. (1995) surmise that precise binding site spacing imposed by the heterodimer interface increases the cooperativity of binding by the dimer. Like the Rap1 DBD, MATα1/MATα2 HTH motifs bind in tandem to a relatively large binding site (18 bp span for MATα1/MATα2, 14 bp for Rap1). As a single molecule the Rap1 DBD encompasses
many of the characteristics of this heterodimer, including a structured linker. It can therefore be suggested that the precise spacing of HTH1-HTH2 may also contribute to cooperative HTH1-HTH2 DNA binding.

Another explanation for differences in telomeric and silencer binding affinity would be that Rapl does not make identical protein-DNA contacts in all Rapl sites. Rapl would in fact “shift” its protein-DNA interactions, allowing it to maintain optimal affinity to a wide variety of sites. Shifting of the recognition surface of the protein relative to DNA has been observed with Helix-Turn-Helix motifs (Brennan and Matthews, 1989). However, studies on the KMnO₄ sensitivity of telomeric and HMR-E sites when Rapl is bound (Gilson et al., 1993) indicate that the two sites protect the same nucleotides, suggesting that the protein-DNA interactions are similar. However, OH radical reactivity studies might have been more sensitive. Indrissi and Pina (1998) were able to distinguish binding to a telomere and UAS<sup>g</sup> Rapl binding site by chemical probes in vitro. This supports the theory that the recognition surface does shift for HTH2 depending on the binding site Rapl complexes. A crystal structure of Rapl in complex with a silencer and/or UAS<sub>rg</sub> site would indicate any shifts in the interaction of Rapl recognition helix residues and the DNA.

10. RELATION OF RAP1 STRUCTURE/FUNCTION TO CHROMATIN ORGANIZATION

Rapl protein plays a role in numerous cell functions. Chromatin organization of telomeres is dependent on Rapl binding. Gene silencing involves a similar set of heterochromatin inducing gene products, including Rapl. Transcriptional activation is known to be Rapl-dependent for some genes, and is assumed to be a factor for many more. Various regions of Rapl are involved in the particular protein-protein interactions required for each function. As well, the intracellular environment can regulate these functions. For example, phosphorylation of Rapl by PKA can increase transcriptional activation (Klein and Struhl, 1994).

Despite the multiple roles for Rap1, its DNA binding involves only a single domain. How does the structure of this DBD aid function of the protein? For one, the tandem orientation of the two Helix Turn Helices allows optimal binding to the telomeric repeat sequences [(C)<sub>i</sub>,
Spacing of HTH1-HTH2, as well as the rotation of HTH2 in relation to HTH1 are optimal for binding at these repeats.

Secondly, because binding to the tandem repeats is cooperative, degenerate DNA binding sites are able to bind the protein at high affinity. This allows for genetic drift to occur since it is not essential that Rap1 binding sites be completely conserved for cell viability. Mutation of a Rap1 binding site in the UAS of a “glycolytic” gene, for instance, could occur at all but the highly conserved A C C sequence. Rap1 dependent gene activation would still take place.

The ability of Rap1 to bind degenerate sites is also consistent with the large number of Rap1 binding sites in the cell. If Rap1 were specific to one function, for instance binding only at telomeres, a highly specific DNA binding sequence would keep random Rap1 sites from appearing outside of the telomeres. At the same time, if Rap1 binding affinity were less specific, the protein could form cryptic silencing regions, disrupting transcriptional activation by a single point mutation. Rap1 is able to balance its varied functions with its need for specificity using the HTH1-HTH2 domain. A cryptic site would require both the “core” AC ACC and some conservation of the 3’ sequence.

The recent publication of Indrissi and Pina (1999) suggests another role for the two HTH domains. The HTH2 changes its structure depending on the Rap1 site to which it complexes. Their results show that Rap1 bound to a UAS^ site induces KmnO4-hypersensitive sites not present when Rap1 is bound to a telomeric site. They conclude that Rap1 and its binding site undergo structural rearrangement depending on the binding site. This allows Rap1 to adopt a different structure between telomere and activator binding sites. Rap1 placed in tandem is able to activate transcription at a higher level when bound to a transcriptional activator site when compared to a telomeric site (Indrissi and Pina, 1999).

Though not essential for in vitro binding, the “tail” region may also play a role in the structure/function of Rap1 in vivo. Interactions via both HTH1 and the “tail” with the A C C sequence may explain why this sequence is invariant in all high affinity binding sites. Any mutation at these bases would disrupt both HTH1 and the “tail”, allowing only optimal HTH2 interactions. This would be more destabilizing than a similar mutation affecting HTH2. The tail can also increase cooperativity, via a protein-protein contact (Konig et al., 1996).
11. PHOSPHORYLATION AND DBD BINDING AFFINITY

Phosphorylation of HTH1 with heart muscle kinase in the presence of labeled phosphate shows that HTH1 does not contain any in vitro PKA phosphorylation sites. This is consistent with the predictions of Yamazaki (1995) that phosphorylation sites are located at Ser-450 and Ser-516 on the DBD of Rap1. Both of these sites are outside of HTH1. Ser-450 is located between the minor groove interacting Lys-446 and the first helix of the HTH2 motif. Ser-516 is located 8 residues N-terminal to the second helix of the HTH2 motif. Thus both of these potential phosphorylation sites might alter the three dimensional structure of HTH2.

This altering of HTH2 may in part explain some contradictory results concerning Rap1 phosphorylation. In the work of Tsang et al. (1990), one Rap1 binding site (the Z+ fragment of the PGK promoter) bound phosphorylated Rap1 at higher affinity compared to unphosphorylated protein. The same Z+ fragment, with an additional 5' sequence from the host plasmid added, bound phosphorylated Rap1 at lower affinity than unphosphorylated protein. A TEF2 site was used in Yamazaki's (1995) study, and displayed higher affinity for phosphorylated protein. If the three dimensional structure of Rap1 HTH2 is altered by phosphorylation, then spacing and rotation of HTH1 relative to HTH2 could be altered, allowing Rap1 to interact at higher affinity with a TEF2 site. This same three dimensional shift may decrease the number of interactions on a Z+ fragment with additional 5' sequence.

Rap1 is able to bind to a variety of divergent sites. Phosphorylation may be used by the cell to modulate this binding. This would give the cell more control over which Rap1 sites are to be occupied.

12. CONCLUSIONS

The experiments outlined in this thesis identify a number of key aspects of Rap1 function: 1. HTH1 and HTH2 interact in a cooperative manner, allowing high affinity binding to a range of DNA binding sites. 2. The DBD "tail" region is not essential for binding, but still may play a role in vivo. 3. The "loop" region is likely involved in some aspect of Rap1 DNA binding. 4. In vivo, Rap1 protein may bind different Rap1 sites, including telomeres and silencer sites, at different affinities. 5. Minor groove interactions help stabilize DNA binding, but are not
6. All of the structures mentioned above aid the function of Rap1 in the yeast cell, providing a means for Rap1 to regulate varied cell processes in a specific manner.

13. FUTURE PROSPECTS

Two studies, one involving yeast genetics, the other X-ray crystallography, would complement the research described above. Various mutants, especially “tail” and “loop” deletions, could be expressed in yeast, replacing the endogenous Rap1. If these mutants were viable, phenotypic and genotypic studies could then be undertaken to understand the effect of these changes. The technique of Graham et al. (1999) would be an appropriate experimental method to employ.

A second experiment would involve the X-ray crystallography of Rap1 DBD bound to a HML-E binding site. Determining the actual protein-DNA contacts made by DBD-HML-E would confirm the model of cooperative binding and would provide evidence for how Rap1 binds such disparate sites with high affinity.
V. REFERENCES


127
References


Gilson, E., Roberge, M., Giraldo, R., Rhodes, D. and Gasser, S. M. Distortion of the DNA double helix by RAP1 at silencers and multiple telomeric binding sites. *J. Mol. Biol.* 231:293-310, 1993

Giraldo, R. and Rhodes, D. The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. *EMBO J.* 13:2411-2420, 1994


Goutte, C. and Johnson, A. D. Yeast a1 and alpha 2 homeodomain proteins form a DNA-binding activity with properties distinct from those of either protein. *J. Mol. Biol.* 233:359-371, 1993


References


References


Konig, P., Giraldo, R., Chapman, L. and Rhodes, D. The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA. *Cell* 85:125-136, 1996


References

Li, T., Stark, M. R., Johnson, A. D. and Wolberger, C. Crystal structure of the MATa1/MATα2 homeodomain heterodimer bound to DNA. *Science* 270:262-269, 1995


132


Renauld, H., Aparicio, O. M., Zierath, P. D., Billington, B. L., Chhablani, S. K. and Gottschling, D. E. Silent domains are assembled continuously from the telomere and are defined by promoter distance and strength, and by SIR3 dosage. *Genes Dev.* 7:1133-1145, 1993


Snyder, M., Sapolsky, R. J., Davis, R. W. Transcription interferes with elements important for chromosome maintenance in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 8:2184-2194, 1988


Turner, B. M., Birley, A. J. and Lavender, J. Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in Drosophila polytene nuclei. *Cell* 69:375-384, 1992

References


Vignais, M. L. and Sentenac, A. Asymmetric DNA bending induced by the yeast multifunctional factor TUF. *J. Biol. Chem.* 264:8463-8466, 1989


