SPECIFICITY AT THE *b* MATING TYPE LOCUS OF *USTILAGO MAYDIS*

by

ARTHUR RAYMOND YEE

B.Sc., The University of British Columbia
M.Sc., The University of British Columbia

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
in
THE FACULTY OF GRADUATE STUDIES

Biotechnology Laboratory and
Department of Plant Science

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

March 1998

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 Plant Science and Biotechnology Laboratory

The University of British Columbia
Vancouver, Canada

Date 2nd April 1998

Abstract

The ability to discriminate between self and nonself allows *Ustilago maydis* to maintain genetic diversity through a multiallelic mating type system governed by the *b* locus. There are approximately 25 different alleles of the *b* locus and a given individual is incompatible with "self", but compatible with all "nonself" mating types. The *b* locus consists of two divergently transcribed genes, designated *bE* and *bW*, and recognition of "self" occurs because bE and bW proteins from the same allele do not dimerize. Conversely, bE and bW proteins encoded by two different *b* alleles do form dimers, and these dimers are active as homeodomain transcription factors, regulating the processes of sexual reproduction and pathogenicity. The N-terminal domains of both bE and bW have been shown to contain the determinants of dimerization specificity. A large set of closely spaced chimeric alleles was constructed for each of the interacting genes, one set of *bE1-bE2* chimeras, and another of *bW1-bW2* chimeras. The specificity of each chimeric allele was determined and they fell into three classes; *b1* (class I), *b2* (class III), and specificity different from either *b1* or *b2* (class II). The close spacing of the chimeric alleles allowed specificity changes at class boundaries to be attributed to single amino acid differences. An examination of substitutions affecting specificity revealed that 3 out of 8 involved an amino acid with an aromatic side chain (Tyr or Phe). To gain further insight into the interaction between bE and bW, strains carrying chimeric alleles of *bE* and *bW* were crossed in all pairwise combinations. The patterns of compatible and incompatible crosses identified intermolecular interactions between two pairs of positions on bE and bW. Site directed mutagenesis was employed to confirm the interaction between one of these pairs of positions. The results indicated that Val at position 90 of bE1 and Tyr at position 74 of bW1 interact, either directly or indirectly, to inhibit dimerization between bE1 and bW1. These data provide evidence for an "inhibitory determinant" model, whereby specificity between bE and bW proteins is mediated by residues that prevent interaction between self combinations.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I thank my supervisor Dr. James W. Kronstad for the opportunity to work and train in his laboratory, and for his guidance during my graduate studies. Appreciation is extended to the members of my supervisory committee Dr. Brian Ellis, Dr. Lawrence McIntosh, Dr. Ivan Sadowski, and Dr. David Theilmann for their advice and assistance during my research. Thanks is given to George Athwal, Alan Au, and Jeanette Beatty for their help on parts of this research. I also wish to acknowledge the numerous postdoctoral fellows who have passed through the laboratory during my stay, especially Dr. Scott Gold. The diversity of their knowledge and background has enriched the laboratory immensely.

Finally, I thank my wife Lorraine, for her help, support, and perseverance through my years of graduate studies. This accomplishment is her's as well as mine.

1. INTRODUCTION

Specific interactions between proteins play a critical role in many biological processes such as self/nonself recognition, transcriptional activation, and signal transduction. The antibody-antigen complex is a well known example of specific protein-protein interactions that mediate self/nonself recognition (Davies and Cohen 1996; Wilson and Stanfield 1993). Transcription factors such as the Fos and Myc oncoproteins interact specifically with another partner, but not with themselves, leading either to the activation or repression of genes involved in oncogenesis (Amati et al. 1993; O'Shea et al. 1992). The specificity of signal transduction pathways largely results from protein-protein interactions which act to target or anchor the signaling components (Cohen et al. 1995; Faux and Scott 1996).

Protein-protein interactions are also involved in a classic self/nonself recognition phenomenon, that of fungal mating type. Recent molecular genetic studies of the yeast *Saccharomyces cerevisiae* and other yeast-like and filamentous fungi have revealed that one mechanism of mating type recognition involves heterodimerization of transcription factors (Dolan and Fields 1991; Kahmann and Bolker 1996). This formation of heterodimers is a key event that results in the recognition of nonself and allows the transcription factors to then repress or activate the genes involved in the mating process. For example, in *S. cerevisiae*, there are two alternate mating types and two proteins, called **a1** and α2, involved in formation of a transcriptional repressor during the nonself recognition event.

The dimorphic yeast *Ustilago maydis*, presents an added dimension to the recognition phenomenon, because it has multiple mating types, estimated to be 25 or more. Any combination of two haploid strains having different mating types is sexually compatible but combinations of the same mating type are not. A multiallelic mating type system presumably increases the number of potentially compatible mating partners for a given individual. Again, two proteins designated bE[1]

[1] Names of proteins are given in non-italics and the corresponding genes are italicized

and bW, are involved in heterodimer formation, but they exist in multiple forms, corresponding to each of the 25 or so mating types. That is, bE exists as bE1, bE2, bE3, etc., and bW exists as bW1, bW2, bW3, etc. Heterodimers form between bE and bW only if they are derived from different alleles. Thus, bE1 heterodimerizes with each of the bW proteins except for bW1. The reason for this specificity in protein-protein interactions is the subject of this thesis.

Since *U. maydis* and the *b* locus are amenable to both classical and molecular genetic studies, they provide a convenient system to study the phenomenon of multiallelic recognition in a relatively simple eukaryotic organism. The elucidation of the mechanism involved would provide an answer to this interesting self/nonself recognition problem and likely provide insight to similar recognition phenomenon in other organisms.

1.1 *Ustilago maydis*

Ustilago maydis has several desirable characteristics as an experimental organism. The fungus is easily cultured in the laboratory and is capable of unicellular yeast-like growth, so it can be conveniently subjected to standard microbiological manipulations. The classical genetics of *U. maydis* and its mating type system have been well characterized (Holliday 1974). It is the organism that Robin Holliday used to develop his model for DNA recombination in eukaryotes (Holliday 1962; Holliday 1964). Furthermore, it is amenable to efficient DNA transformation and targeted gene replacement (Tsukuda et al. 1988; Wang et al. 1988). Thus, *U. maydis* provides a convenient system for studying recognition in a relatively simple eukaryotic organism.

U. maydis is a fungus which is classified in the phylum *Basidiomycota*, order *Ustilaginales*, the smut fungi (Alexopoulos et al. 1996). The smut fungi share characteristics such as the ability to cause plant disease and the production of dark teliospores within the host. The corn smut disease caused by *U. maydis* (Agrios 1988) is characterized in early stages by chlorosis and purpling on stems and leaves around the area of infection (Figure 1-1). Then, neoplastic growth produces swollen galls or tumors which eventually fill with dark sooty teliospores. These can form on the stems, leaves, or ears of the corn plant and can vary in size from a few millimeters

to 1-2 cm, depending on the location and the time of infection. The infection spreads within the plant by hyphal growth in meristematic tissues, but it is not a systemic disease.

Infection of the corn plant occurs during the growing season, with the inoculum coming from overwintering teliospores in the soil or plant residue (Agrios 1988). The teliospores can germinate on the growing corn plant or on plant residue to give haploid sporidia which grow saprophytically. Compatible haploid strains will fuse to form a dikaryon which can then infect and grow parasitically within the meristematic regions of the plant. The parasitic growth leads to tumor and teliospore formation. The teliospores are survival propagules, allowing the fungus to overwinter in the soil or plant residue.

Adequate control of the disease is presently achieved by using resistant corn varieties, and to a limited extent, by sanitation. Resistance to corn smut is multigenic and is associated with field corn varieties (Christensen 1963), but no varieties are completely resistant. Crop damage from corn smut is highly variable. Yield losses may average about 2% for resistant varieties (Agrios 1988) while susceptible varieties, such as those of sweet corn, may suffer considerable damage. A good review of the disease is given by Christensen (1963).

Figure 1-1. Colony morphology and disease symptoms of *U. maydis* on corn. (a) Haploid strain 518 *(a2b2)* and (b) diploid strain d132 *(a1/a2 b1/b2)* of *U. maydis* grown on charcoal media. (c) *U. maydis* tumor at the base of a corn seedling. (d) Disease symptoms of *U. maydis* on corn leaves and stems; note tumors at base of stem. (e) deformation and incipient tumors on a corn leaf.

1.2 The Life Cycle of *Ustilago maydis*

The life cycle of *U. maydis* (Banuett and Herskowitz 1988; Christensen 1963) consists of three stages: haploid, dikaryotic, and diploid (Figure 1-2). The haploid cells (sporidia) are normally yeast-like, grow vegetatively by budding and can be supported by artificial media. This stage is nonpathogenic. Haploid cells that are compatible at the mating type loci are capable of fusing with each other to form dikaryotic cells. The dikaryotic cells (possessing two individual haploid nuclei) are mycelial, grow by hyphal extension and can be supported by charcoal containing media (Day and Anagnostakis 1971). This stage is pathogenic. Growth of dikaryotic hyphae in the plant eventually leads to nuclear fusion (karyogamy) and the formation of the diploid teliospore. To date, teliospore formation has only been successful in the plant or in cell culture (Christensen 1963). It has not been demonstrated to occur on artificial media. The teliospore is a resting stage, but given the correct temperature and moisture conditions, it will germinate, undergo meiosis, and produce 4 haploid products. These haploid sporidia can again grow vegetatively and repeat the cycle. Occasionally, during teliospore germination, meiosis fails and diploid cells are formed. These cells are capable of both yeast-like or mycelial growth and they are also pathogenic.

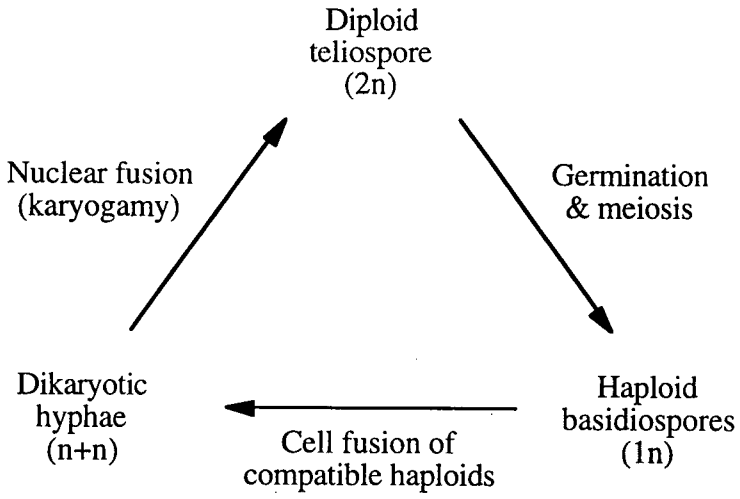

Figure 1-2. The life cycle of *Ustilago maydis*.

1.3 The Mating Type System of *Ustilago maydis*

Many aspects of biological recognition are played out in the mating type system of *U. maydis*. This mating type system determines the sexual compatibility of two haploid strains of *U. maydis*; that is, whether or not two haploid strains can fuse and proceed through the sexual cycle, eventually leading to teliospore (sexual spore) formation

Mating type studies in *U. maydis* date as far back as the 1920's (Hanna 1929; Stakman and Christensen 1927), and extensive classical genetics studies of the mating type loci were carried out in the 1950's, 60's, and 70's. These led to the characterization two unlinked mating type loci designated *a* and *b* (Rowell and DeVay 1954), which regulate different stages of sexual compatibility in *U. maydis*. Haploid strains of *U. maydis* must possess different alleles at both *a* and *b* loci in order to be sexually compatible.

The *a* locus regulates initial cell fusion between haploid cells (Holliday 1961; Puhalla 1969; Rowell 1955). It has two alternative alleles *a1* and *a2*, and fusion only occurs between *a1* and *a2* cells, not between cells containing the same *a* allele.

The *b* locus regulates the events after cell fusion such as mycelial phenotype (Puhalla 1969; Rowell 1955), pathogenicity, and teliospore formation (Day et al. 1971; Holliday 1961). The *b* locus is multiallelic, and compatibility at the *b* locus requires the combination of any two different alleles.

The actual number of *b* alleles is not known. Rowell (1954) isolated 15 *b* alleles from teliospores collected in the United States and Brazil. In an independent study, Puhalla (1970) isolated 18 different *b* alleles from teliospores collected from various parts of the United States and Canada. Using a statistical formula, Puhalla estimated the total number of *b* alleles in North America to be a maximum of 25. Silva (1972) isolated 12 *b* alleles from a collection of teliospores from Poland; six of which were different from 17 of the 18 alleles isolated by Puhalla (1970), giving a total of 23 known alleles. These alleles were assigned letters from A to X (allele J was missing). An alternative naming system of independently isolated *b* alleles was used by R. Holliday (Commonwealth Scientific and Industrial Research Organization, Laboratory of

Molecular Biology, Sydney, Australia). Holliday designated *b* alleles using numbers *b1* to *b7* and some of their corresponding letter designations were reported in Kronstad and Leong (1990).

There are few classical mutagenesis studies on the *a* or *b* loci but one significant result was reported by Day et al. (1971) concerning mutagenesis of the *b* locus. In this study, they mutagenized with UV light, a diploid *U. maydis* strain that was *a1/a2 bG/bG* (heterozygous at the *a* locus, but homozygous at the *b* locus). This parent diploid strain was derived from a product of teliospore germination that failed to undergo meiosis. It formed yeast-like colonies on double complete medium but upon mutagenesis produced diploid mutants that were mycelial. These diploid mutants were solopathogenic and produced teliospores when injected into corn plants on their own. The meiotic products of these teliospores showed a 1:1 segregation of *bG wild type:bG mutant* alleles. Thus, the mutant diploids were presumably *a1/a2 bG/bGmut*. Haploid *bGmut* meiotic products tended to be mycelial and led to teliospore formation when injected alone into corn plants (they were solopathogenic). Only *bGmut* meiotic products could be isolated from these second generation teliospores; and no *bG* wild-types were found. This was strong evidence that the *bG* mutation was truly at the *b* locus. In an independent but possibly related finding, the isolation of weakly self compatible but not solopathogenic natural strains from Poland was reported by Silva (1972). An explanation for the self compatible *b* alleles found by Day et al. (1971) and Silva (1972) came 20 years later with the application of molecular techniques to the *b* locus puzzle. The work of Kamper et al. (1995) described single codon changes in one of the *b* locus genes, resulting in self compatibility (see Section 1.5 Molecular Studies of the *b* Mating Type Locus).

Earlier studies with *U. maydis* used corn seedling inoculations with mixtures of two haploids to determine compatibility. Compatible strains produced teliospores on corn plants, while incompatible mixtures did not. This method was accurate but results took one to two weeks to obtain. The development of faster *in vitro* assays grew out of observations on growth morphology

between compatible and incompatible cell mixtures on agar media. *U. maydis* normally produces smooth yeast-like growth on agar plates. Rowell (1955) reported the use of agar media containing corn coleoptile extract to promote vigorous growth of infection hyphae resulting from compatible matings but concluded that corn inoculations were still a more accurate test of compatibility. Puhalla (1968) found that the accuracy of the plate mating assays was markedly improved with the use of rich agar medium with a high concentration of amino acids. The hyphae formation along a contact line between two compatible strains was most reliable on double strength complete agar, less reliable on complete agar, and nonexistent on minimal. Temperature was less critical, but he reported that 25°C was better than 30°C, and sealing of the petri plates improved the differentiation of yeast-like and hyphal growth. A small number of haploid pairs that were capable of teliospore formation would show little or no hyphae formation on his plate assay, but all haploid pairs that show a positive plate mating would form teliospores. Thus, Puhalla (1968) felt that his plate mating assay was accurate enough to be of "inestimable value in genetic studies of the incompatibility loci."

The next improvement to the plate mating test was the addition of activated carbon to the mating media by Day and Anagnostakis (1971). They were attempting to force dikaryon formation by growing compatible haploids having complementary auxotrophic markers on minimal media. Unfortunately, the minimal media would not support the dikaryotic hyphal growth. Their observation that the nutrient rich medium that supports hyphal growth was often caramelized after autoclaving prompted them to add activated carbon to minimal media. This proved to be successful in supporting hyphal growth. They hypothesized that the activated carbon may have removed a substance produced by the cells that favors yeast-like growth over hyphal growth. Presently, the standard mating media for *U. maydis* is double complete with 1% activated charcoal (Holliday 1974). The plate mating test on charcoal media is illustrated in Figure 1-3.

Compatible
521 x 518
(a1b1) (a2b2)

Incompatible
521 x 032
(a1b1) (a2b1)

Figure 1-3. Mating tests between compatible and incompatible strains of *U. maydis*. Haploid strains of *U maydis* are grown overnight in liquid culture, then co-spotted on charcoal media. The compatible combination of 521*(a1b1)* x 518*(a2b2)*, with different alleles at both *a* and *b* loci, show a white mycelial reactions (fuz+) in the overlap area. The incompatible combination of 521*(a1b1)* x 032*(a2b1)*, which have different *a* alleles but the same *b* alleles, show yeast-like growth (fuz-).

The mating type system of *U. maydis* not only illustrates the fundamental process of self/nonself recognition, but it provides two versions of this phenomenon, recognition between two alternate cell types and recognition between multiple cell types. Only within the last 10 years, with the application of molecular genetics and recombinant DNA techniques to the study of these mating type genes, have we been able to determine the mechanisms behind these recognition systems. As it turns out, the mechanism involved in *a* locus recognition is quite different from that of the *b* locus. The *a* locus codes for a pheromone and receptor complex which is coupled to a signal transduction pathway. In contrast to this, the *b* locus codes for a pair of transcription factors which mediate recognition through heterodimer formation. The stories behind the elucidation of these mechanisms follow in the next two sections.

1.4 Molecular Studies of the *a* Mating Type Locus

The *a* locus controls initial cell fusion of opposite mating type partners and it does so though a pheromone-receptor signaling system which is triggered only when *a1* and *a2* mating types are brought together. The two allelic forms of the *a* locus each code for a small peptide pheromone plus a receptor for the pheromone of the opposite mating type. Thus, cells of a single mating type cannot activate the pheromone response pathway, but when *a1* and *a2* cell are combined, the pheromone produced by the *a1* cell binds to the receptor carried on the surface of the *a2* cell. The reciprocal event also occurs, resulting in activation of the pheromone response pathway in both mating types. This pheromone response pathway has the classic components of a MAP kinase signal transduction cascade analogous to the *S. cerevisiae* pheromone response pathway (Marsh 1991).

Cloning of the *a* locus was first accomplished (Froeliger and Leong 1991) by exploiting two genetic observations. First, the *a* locus was tightly linked to the auxotrophic marker *pan1-1* (Holliday 1961). Second, diploid *U. maydis* strains heterozygous at *a* have a dual mating capability, mating with haploids carrying either *a1* or *a2* alleles (Banuett and Herskowitz 1989; Holliday 1961). These two observations suggested that the *a2* allele could be isolated by transformation of a *a1b1 pan1-1* strain with a genomic cosmid library from a *a2b2* wild type strain, followed by selection for transformants that were Pan+ and dual maters. These transformants would survive on minimal medium without pantothenic acid supplementation and would be compatible with both *a2b2* and *a1b2* testers. When this experiment was carried out by Froeliger and Leong (1991), they obtained six transformants which appeared to contain both a functional *pan1* gene and a functional *a2* allele. The transforming cosmids were isolated from two of these transformants by extraction and phage packaging of the cosmid DNA, and transfection of *E. coli* with the rescued cosmids. Subsequent subcloning of the cosmid, retransformation into *U. maydis*, and mating tests identified a 6.0 kb *Bam*HI fragment which contained the *a2* allele activity. This fragment with *a2* activity was used to probe a genomic library from an *a1* strain of *U. maydis* to isolate a 10 kb *Bam*HI fragment which contained *a1* allele activity.

Detailed restriction mapping of the genomic region containing the *a1* and *a2* activity indicated that these alleles were embedded within unique sections of DNA, even though they were located at the same chromosomal location and distantly flanked by homologous sequences (Froeliger and Leong 1991). It was suggested that the term "idiomorphs" (Metzenberg and Glass 1990) may be a more appropriate term to describe the *a* mating type alleles because they map to the same location but are not related by structure. No DNA sequence was presented by Froeliger and Leong (1991) to confirm this idea.

The cloning of the *Bam*HI fragments containing the *a1* and *a2* mating type alleles was also accomplished by Bolker et al. (1992) using a strategy similar to that of Froeliger and Leong (1991). Additional subcloning, transposon, and deletion analysis revealed that each *Bam*HI fragment contained two open reading frames, coding for a precursor lipopeptide pheromone and a pheromone receptor. The pheromone genes were designated *mfa1* (mating factor *a1*) and *mfa2* (mating factor *a2*) and the receptor genes were *pra1* (pheromone receptor encoded by *a1* allele) and *pra2* (pheromone receptor encoded by *a2* allele) (Bolker et al. 1992). Both these genes showed homology to the corresponding *a* factor pheromone (including a CAAX prenylation motif) and *a* factor receptor gene *(STE3)* of *S. cerevisiae*.

Assays for pheromone and receptor gene function were performed by transforming the pheromone or receptor genes by themselves into recipients having either *a1* or *a2* genotypes (Bolker et al. 1992). These experiments led to a model for *a* locus function whereby each *a* locus allele possesses the gene for its own mating factor pheromone plus the receptor gene for the opposite mating factor.

These data show that the *a* mating type alleles are more accurately described as alleles than as idiomorphs, even though they are embedded in nonhomologous DNA. The role of the nonhomologous DNA may be to prevent recombination in the vicinity of the *a* locus, thus preserving the correct pairing of pheromone and receptor or preventing the switching of mating types.

A number of the downstream signal transduction components have been identified in *U. maydis* and they appear to show homology to the classic pheromone response pathway of *S. cerevisiae*. A *STE7* homolog, named *fuz7* has been cloned and sequenced (Banuett and Herskowitz 1994). It is required for filamentous growth in *U. maydis* and belongs to the MAP kinase kinase (MAPKK/MEK) family of kinases . Multimeric G-proteins in *U. maydis* that mediate signal transduction between the pheromone receptor to the MAP kinase cascade have also been reported (Regenfelder et al. 1997). It is likely that the *U. maydis* pheromone response pathway possesses additional *S. cerevisiae* homologs such as STE11 (a MAP kinase kinase kinase), FUS3 (a MAP kinase) and perhaps even STE5, the "scaffolding" protein (Herskowitz 1995).

Thus, the *a* locus of *U. maydis* illustrates one of the classic recognition mechanisms, namely, a hormone/receptor/G-protein/MAP kinase cascade. But there are more lessons to be learned from *Ustilago*, as will be illustrated by the *b* locus story.

1.5 Molecular Studies of the *b* Mating Type Locus

The *b* locus controls both pathogenicity and sexual development through a mechanism involving heterodimer formation between the two b proteins. These two proteins, designated bE and bW, are transcription factors with N-terminal dimerization domains and DNA binding homeodomains. This recognition system has many parallels to the **a1**/α2 mating type system of *S. cerevisiae*.

The *b* locus was first cloned by Kronstad and Leong (1989). They used a nonmycelial diploid strain of *U. maydis*, homozygous at the *b* locus *(a1/a2 b2/b2),* and transformed this with a cosmid library from a *b1* haploid. Transformants receiving the *b1* gene were expected to be mycelial and pathogenic. This strategy was successful in isolating one cosmid clone containing *b1* activity. Further subcloning and transformation narrowed the *b1* activity to a 8.5 kb *Bam*HI fragment. This fragment was used as a probe to identify homologous clones in a cosmid library

prepared from a *b2* haploid strain. Eventually, a 8.5 kb *Bam*HI fragment was isolated containing *b2* activity.

Sequencing of several *b* alleles revealed an open reading frame containing a carboxy (C) portion that is conserved between alleles and an amino (N) portion that is variable (Kronstad and Leong 1990; Schulz et al. 1990). Also, the conserved portion contained a homeodomain related motif suggesting that the *b* locus coded for a DNA binding protein (Schulz et al. 1990). This *b* locus gene was first reported to be 410 amino acids in length, but subsequent discovery of an intron and an additional exon close to the C-terminus led to a revision of this to 473 amino acids (Gillissen et al. 1992).

A second open reading frame at the *b* locus was discovered by Gillissen et al. (1992) and initially reported to be 626 amino acids in length. This second gene, designated *bW*, was transcribed divergently from the first, now designated *bE*, and the two were separated by a 200 bp shared promoter region (Figure 1-6). The size of the *bW* ORF was later revised to 644 amino acids for *bW1* and 645 amino acids for *bW2*, after determining that the correct translational start site was 18 to 19 codons upstream of the original ATG start site (Kamper et al. 1995) . Although *bE* and *bW* were not homologous, they did share a common structure, with allelic variability in the N-terminal domains, a conserved C-terminal domain, and a homeodomain separating the two. The N-terminal variable region of *bE* occurs from amino acid positions 1 to 110. Within this variable region, the sequences of *bE1* and *bE2* are 66% identical, whereas the remainder of the *bE* open reading frame is 93% identical between *bE1* and *bE2*. Similarly, the *bW* variable region from amino acid positions 1-150 is 54% identical between *bW1* and *bW2*, while the remainder of the open reading frame is 96% identical.

A comparison of amino acid sequences between *bE1* and *bE2* is given in Figure 1-4 and between *bW1* and *bW2* in Figure 1-5.

```
                .         .         .         .         .
bE1   MSSDPNFSLISFLECLNEIEHEFLRDKGENYPVLVRKLRELQQKIPNDIA  50
      ||  |||||  || ||||||||||||||| || ||||||| ||  | ||  |
bE2   MSNYPNFSLTSFVECLNEIEHEFLRDKLENRPVLVRKLQELRRKTPNNVA  50

                .         .         .         .         .
bE1   NLPRDPETIQQIHQTTHRIRAVAQAFIRFDQKFVSLCSEVVHGTSKVMQE 100
       |  || || |||||||||   | |||| || |||| |  |  |||
bE2   SLSYDPGTIHQIHQTTHRIKVAAKAFIRIDQSFVSLHSDAVEDTSKALKK 100

                .         .         .         .         .
bE1   FNVVSPDVGCRNLSEDLPAYHMRKHFLLTLDNPTPRQEEKETLVRLTNES 150
          || |||| |||||||||||||| |||||||||||| ||||||||
bE2   ADASSPVVGCRDLSEDLPAYHMRKHFLHTLDNPTPTQEEKEGLVRLTNES 150
```

Figure 1-4. Amino acid sequence alignment of the N-terminal portions of bE1 and bE2. Vertical bars indicate identity. The *bE* variable region lies from amino acids 1 to 110. Note that the entire bE protein is 473 amino acids long.

```
                .         .         .         .         .
bW1   MKDFEYFSKILSLASQIRMTLPPLPRISQTAPRPTCFLPLSLEGPNQQAL  50
      | | | || || | ||||  ||||||||||||||||| | ||||| |||  |
bW2   MTDLECFSEILYLTSQIRAMLPPLPRISQTAPRPTRFFPLSLESPNQLTL  50

          .         .         .         .         .
bW1   SRKLSKLGIGSVCRDTLEEIFIEYLRKLRRVYEAQYENAFVTWQQENLYE 100
      |  ||  |        |   |   |  |     ||||  |  ||||||||
bW2   SHELSGHGVNGSYHEALIKLFLGHLNELHIGCQAQYERVFAIWQQENLYE 100

                .         .         .         .         . 150
bW1   EAYDQAFRKLLNRLFAMHSQETWHMVLDEVSKVFRTDSSLTVTQRDNASY
      |||||||| ||   |   |  | | | | ||     ||       |  ||
bW2   EAYDQAFRNLLKHVFSTRSTKMWYMLLKEASKYTQRGSSPPEPHDDDVSY 150

                .         .         .         .         .
bW1   EGAPLKTGRGHDSEAVRILEQAFKHSPNITPAEKFRLSEVTGLKPKQVTI 200
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
bW2   EGAPLKTGRGHDSEAVRILEQAFKHSPNITPAEKFRLSEVTGLKPKQVTI 200
```

Figure 1-5. Amino acid sequence alignment of the N-terminal portions of bW1 and bW2. Vertical bars indicate identity. The *bW* variable region lies from amino acids 1 to 150. Note that the entire bW protein was initially reported to be 626 amino acids long and was later revised to 644 or 645 amino acids (see text).

Gillissen et al. (1992) presented genetic evidence, through individual gene disruptions of *bE* or *bW*, indicating that the combination of at least one *bE* and one *bW* gene product was necessary for recognition (Gillissen et al. 1992). This led to a proposed model whereby the activity of a bE-bW heterodimer depended upon whether bE and bW proteins came from the same allele or from different alleles.

Since variability between *b* alleles was found in the N-terminal portions of both *bE* and *bW*, it was logical to carry out N-terminal swapping experiments to identify the regions responsible for allelic identity. This was carried out for the *bE* gene by two separate groups using different strategies.

An *in vitro* strategy using common restriction sites allowed Dahl et al. (1991) to construct *bE2-bE3* chimeric alleles, which were then transformed into *U. maydis* to replace the resident *b* allele. The chimeric allele strains were then mated with *bE2* and *bE3* testers to determine the specificity of the chimeric alleles. They found that a chimeric allele with the first 115 amino acids derived from bE2 and the remaining amino acids derived from bE3 showed bE2 specificity. A chimeric allele with only 56 amino acids from bE2 was unchanged. This allowed them to define a specificity domain from positions 56 to 115 for the bE2 and bE3 polypeptides.

An *in vivo* strategy was used by Yee and Kronstad (1993) to construct *bE1-bE2* chimeric alleles. A linear DNA construct containing *bE1* variable region sequences linked to a selectable marker was transformed into a *bE2* haploid. One step gene replacement of the resident *bE2* sequences with the transforming *bE1* variable region produced a *bE1-bE2* chimeric allele. This process occurs at a reasonably high frequency (5%-50% of transformants) by homologous integration of the transforming DNA. A series of alleles with recombination points scanning the entire variable region were constructed by varying the length of transforming *bE1* sequence. This strategy allowed more precise definition of the specificity region for *bE1* versus *bE2* alleles, which was located between positions 39 and 87. Surprisingly, alleles with recombination points falling within this region had a specificity that was different from either parental allele.

The proposed bE-bW heterodimer model (Gillissen et al. 1992) and the significance of the specificity domain studies (Dahl et al. 1991; Yee and Kronstad 1993) were confirmed by two-hybrid studies that demonstrated that bE and bW dimerize but only if they are derived from different alleles (Kamper et al. 1995). This allele specific dimerization was mediated by the N-terminal variable domains of both bE and bW. The N-terminal domains were thus dimerization domains and these domains were unable to dimerize when derived from the same allele. A bE-bW heterodimer would presumably be active (either as a transcriptional repressor or activator) in controlling genes involved in pathogenicity and sporulation. This work provided some important insight into self incompatibility but it still did not provide a mechanism for the observed dimerization specificity.

Kamper et al. (1995) also reported mutagenesis of bE2 to generate derivatives with altered specificity. They used PCR under conditions resulting in high misincorporation rates to produce random mutations in bE2. There were five single amino acid substitutions which made bE2 compatible with bW2 and allowed dimerization between the two proteins. All five changes involved an increase in hydrophobicity suggesting that hydrophobicity plays an important role in bE-bW dimerization. These changes though, were not the same substitutions found between corresponding positions in naturally occurring *bE* alleles, so they did not necessarily reflect differences that are relevant to allelic specificity.

The *bE2* mutations of Kamper et al. (1995) do provide a likely explanation for the *bG* locus mutations reported by Day et al. (1971) that resulted in pathogenicity. Although the *bG* allele[2] is a different allele from *b2*, the mechanism of bE-bW protein dimerization still applies. The *bGmut* allele was likely a mutation that increased the hydrophobicity of one or more side chains in the variable region of either bEG or bWG, causing them to form active heterodimers. This would result in activation of genes involved in pathogenicity.

---

[2] The corresponding number designation for the *bG* allele is not known

Further support for the bE-bW heterodimer model was provided by Romeis et al. (1997) who reported that C to N-terminal fusions of bE to bW were active in triggering compatibility and pathogenicity in *U. maydis*. The fusions were active regardless of whether bE and bW originated from the same allele or different alleles, and fusions missing both N-terminal variable/dimerization domains were also active but much attenuated. The presence of both homeodomains were found to be essential. This work emphasizes that compatibility between mating types of *U. maydis* is determined primarily by dimer formation between bE and bW homeodomains.

A primary question left to be answered is: why does dimerization not occur between bE and bW proteins from the same allele? Kahmann and Bolker (1996) suggest it may be due to interfering side chains on both proteins at corresponding sites. In combinations of bE and bW from different alleles, the interfering residue on one partner would not be positioned opposite to an interfering residue on the other partner. No data supporting this proposal have been published.

Surprisingly, the same model for self-incompatibility in the higher fungi *Schizophyllum* spp. was proposed by Kuhn and Parag (1972). Quoting from this paper, "protein subunits that are specified by alleles Aα1 and Aα2 of gene Aα aggregate only when both allelic subunits are present and only with one another. This aggregation leads to a protein of increased molecular weight which unlike its unaggregated subunits, is active. Most likely, identical subunits have a tendency to form aggregates, but opposing forces (electrostatic, etc.) prevent expression of this tendency" (Kuhn and Parag 1972). As will be presented in the next Section, the mechanism of self incompatibility in *Schizophyllum* shares the same elements of protein heterodimer formation as the *b* mating type system.

Thus, the goal for further research into the *b* locus is to identify the regions and specific amino acids that prevent dimerization between self combinations of bE and bW. This thesis presents the results of experiments leading towards this goal.

1.6 The Mating Type Systems of Related Basidiomycetes

The related basidiomycete mushroom fungi *Schizophyllum commune* and *Coprinus cinereus*, have mating type systems similar to that of *U. maydis*, except they have added dimensions of complexity. Instead of a single *b* locus containing one pair of multiallelic genes, these fungi contain two loci, each containing one pair of multiallelic genes (Figure 1-6). Each of these gene pairs behave the same way as *bE* and *bW*; they code for homeodomain transcription factors that are capable of forming heterodimers when they are derived from different alleles.

S. commune contains two linked loci designated as $A\alpha$ and $A\beta$ which each function in a similar manner to the *b* locus of *U. maydis*. The $A\alpha$ locus has 9 alleles and the $A\beta$ has 32 (Stankis et al. 1990). Two *S. commune* haploids must possess different alleles at $A\alpha$ and/or $A\beta$ in order to be sexually compatible, thus these two loci are functionally redundant.

The $A\alpha$ locus is the best studied of the two and it contains two divergently transcribed open reading frames designated *Y* and *Z* (Stankis et al. 1992). The *Y* and *Z* genes are not homologous but they both contain homeodomain. In contrast to the *U. maydis b* genes, the variability between alleles of both the *Y* and *Z* genes is not limited to their respective N-termini (Stankis et al. 1992). Genetic evidence (Specht et al. 1992) and two-hybrid studies (Luo et al. 1994; Magae et al. 1995) of the $A\alpha$ *Y* and *Z* genes indicates that the proteins derived from *Y* and *Z* form active heterodimers when they are derived from different alleles, but not when they are derived from the same allele. Through the construction of *Z5-Z4* chimeric alleles, the N-terminal amino acids between positions 16 and 60 were found to be responsible for allelic specificity (Wu et al. 1996). Also, by two-hybrid analysis, the first 110 amino acids of the *Z* protein were required for interaction with the *Y* protein (Wu et al. 1996). Similar results were obtained by Yue et al. (1997) for chimeric alleles and two-hybrid studies on the *Y* gene. Here, a series of chimeric alleles between *Y4* and *Y3* were constructed with sufficiently close spacing to delineate three classes of alleles. One class, with recombination points at codons 1 and 16, had unaltered *Y3* specificity. Another class, with recombination points at codons 73, 118, and 224 had *Y4* specificity. A novel specificity class, with recombination points at codons 26, 41, and 62, had specificities different from either parental

Y3 or *Y4* alleles. The results presented by Yue et al. (1997) for *Schizophyllum* are incredibly similar to those reported for chimeric alleles of the *Ustilago b* locus (Dahl et al. 1991; Yee and Kronstad 1993), suggesting that these two mating type systems share a common mechanism.

The *Aβ* locus is likely to have the same structure and function as *Aα*. It contains a gene designated as *V*, which also codes for a homeodomain protein (Shen et al. 1996). Although gene *V* does not show overall homology to *Y* or *Z* sequences, it does share with *Y* alleles a WFQN recognition helix sequence in the homeodomain, and a 20 amino acid sequence just C-terminal of the homeodomain. The identification of the expected partner to gene *V*, tentatively called *W*, has not yet been reported.

C. cinereus has a mating type system with the same organization as that of *S. commune*. It also has two closely linked, functionally redundant loci, also designated *Aα* and *Aβ*. In the archetypal situation, each locus consists of a pair of divergently transcribed genes coding for homeodomain proteins (Kues et al. 1992). Some *C. cinereus A* loci alleles show deletions or duplications of one member of a gene pair (Figure 1-6).

The analysis of the homeodomain gene pairs in *C. cinereus* as well as *S. commune* and *U. maydis* revealed that the homeodomains were of two types, HD1 and HD2, and each gene pair had one of each (Kues and Casselton 1992). The HD1 proteins are characterized by a recognition helix with a WFI sequence, and the HD2 proteins by a WFQ sequence (Kues and Casselton 1992). The *S. commune* AαZ and the *Ustilago* bE proteins fall under the HD1 class while AαY and bW falls in the HD2 class (Kues et al. 1994b).

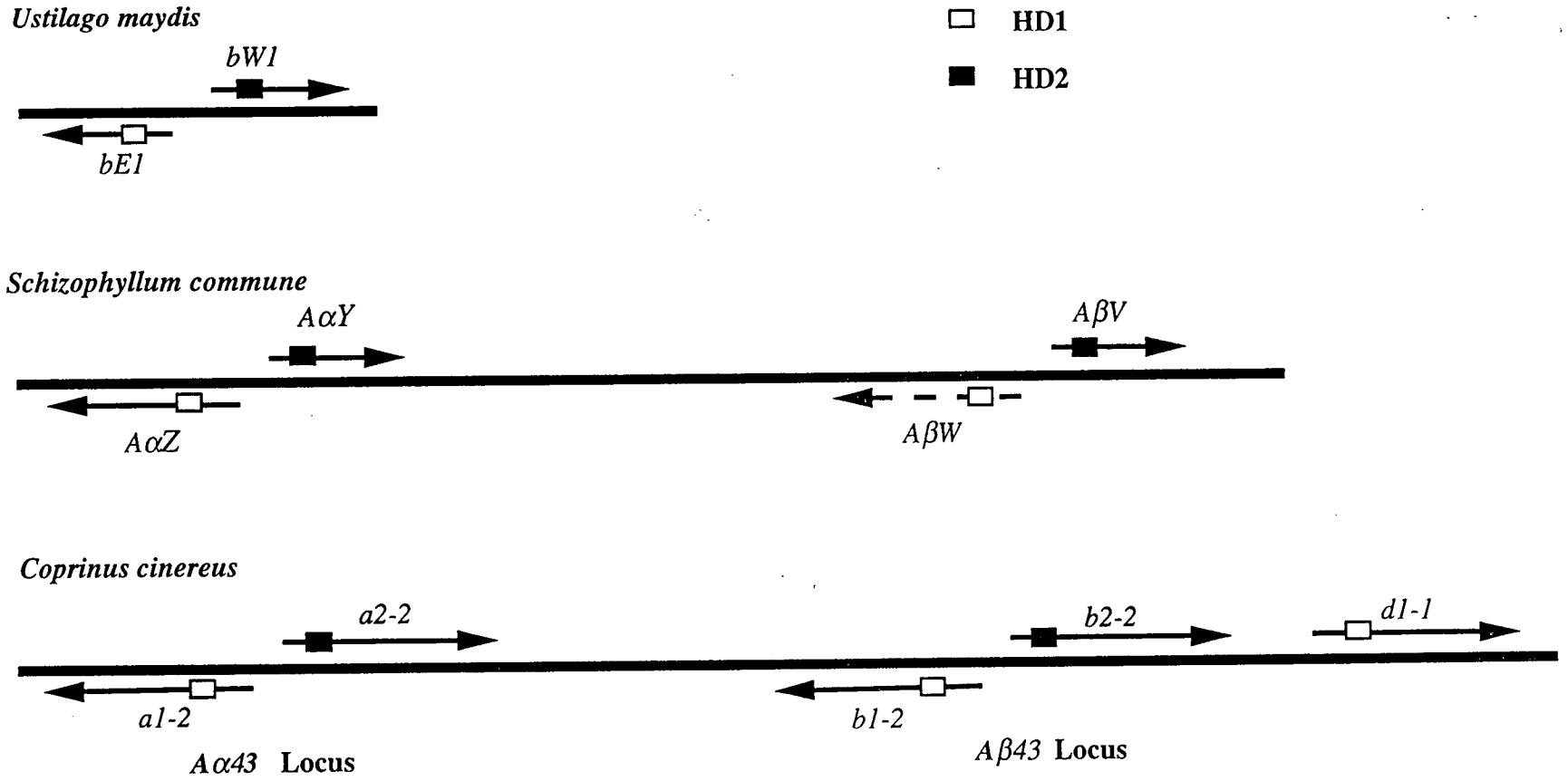

Figure 1-6. Organization of basidiomycete mating type loci.
The open boxes denote HD1 type homeodomains and the closed boxes, HD2. The *AβW* gene from *S. commune* has not yet been isolated. The *d1-1* gene of *C. cinereus* is an extra HD1 gene without a partner HD2.

The *C. cinereus* homeodomain genes have the same domain structure as those of *S. commune* and *U. maydis*, with variable N-terminal domains responsible for specificity and heterodimerization (Banham et al. 1995). Heterodimer formation occurs between HD1 and HD2 type proteins from different alleles and this allows compatibility at the *A* locus. In fact, a rare natural mutation that constitutively activates *A*-regulated development has been shown to result from an inframe fusion of HD1 and HD2 type proteins (Kues et al. 1994a). This C to N-terminal fusion of homeodomain proteins is analogous to the *Ustilago bE-bW* fusion recently reported by Romeis et al. (1997) except the *Coprinus* fusions were active even with one of the homeodomains of the fusion missing.

These related Basidiomycetes also have a separate pheromone-receptor based mating type system analogous to the *a* locus of *Ustilago*, except that they too have additional dimensions of complexity. In both *S. commune* and *C. cinereus,* there exists the two linked *Bα* and *Bβ* loci, each of which is multiallelic. An initial study has revealed that *Bα1* of *S. commune* codes for pheromones and a pheromone receptor (Wendland et al. 1995). It is likely that the *Bα* and *Bβ* loci for both *S. commune* and *C. cinereus* are complex versions of the *a* locus of *Ustilago*.

1.7 The *MAT* Locus of *Saccharomyces cerevisiae*

S. cerevisiae (the budding yeast) has a mating type system consisting of a single locus called the *MAT* locus with two alleles, *MATa* and *MATα*. The *MAT* alleles code for homeodomain proteins **a1** and α2 respectively (Laughon and Scott 1984; Shepherd et al. 1984), which upon fusion of opposite yeast mating types, produce an **a1**-α2 heterodimer (Dranginis 1990). This **a1**/α2 heterodimer acts as a transcriptional repressor of haploid-specific genes (Dranginis 1990) and allows meiosis and sporulation, through repression of a repressor of meiosis *RME1* (Covitz et al. 1991).

There are some similarities between the **a1** and α2 mating type proteins of *S. cerevisiae* and the mating type proteins coded for by the *b* locus of *U. maydis* and the *A* loci of *S. commune* and *C. cinereus*. Firstly, a comparison of the **a1** and α2 homeodomain sequences suggests they

are of two types, with the α2 homeodomain falling into the HD1 class and the **a1** homeodomain into the HD2 class (Asante-Owusu et al. 1996; Kues and Casselton 1992). Secondly, the N-terminal domains of both **a1** and α2 are involved in **a1**-α2 heterodimer formation (Ho et al. 1994; Sauer et al. 1988), just as with mating type proteins of the Basidiomycetes.

It appears that *S. cerevisiae* and the basidiomycete fungi share common elements of their mating type systems. This may extend further to the function of the homeodomain heterodimer. In *S. cerevisiae*, the **a1**-α2 heterodimer functions as a transcriptional repressor. Unfortunately, the targets of the bE-bW or the AαY-AαZ heterodimers are not yet known.

1.8 Protein-Protein Interactions

As discussed in the previous sections, the process of recognition between fungal mating types is governed by the specificity of interactions between two proteins. In order to understand the basis of this specificity, it is important to understand the forces involved in protein dimerization.

A good deal of insight into protein-protein interactions can be gained from studying the characteristics of interfaces between proteins (Creighton 1993). Of fundamental importance is that the protein interfaces exhibit a high degree of shape complementarity, with close packing of atoms of each protein along the interface. This is a primary requirement in order for the stabilizing interactions between proteins to take effect. Secondly, the centers of the interface are dominated by nonpolar side chains involved in hydrophobic interactions. Thirdly, all the polar groups in the interface participate in hydrogen bonds (H-bonds), and ionized side chains are generally neutralized by opposite charges. These polar and charged groups tend to be common on the periphery and rare in the center of the protein-protein interface. The characteristics of a protein-protein interface are essentially the same as those of the interior of a globular protein and the forces involved in bringing together two proteins are essentially the same as those involved in the folding of a single protein (Creighton 1993).

The forces which affect how two proteins interact in an aqueous environment can be categorized as electrostatic, van der Waals, H-bond, and hydrophobic. Electrostatic interactions, involve attraction or repulsion between charged atoms. The van der Waals interactions are close range forces which are repulsive when atoms are closer than normal contact distances, yet weakly attractive beyond this. H-bonds originate from a partial positive charge on a hydrogen atom interacting with a partial negative charge on an electronegative atom, commonly an oxygen. Hydrophobic interactions are due to the preference of nonpolar atoms for nonpolar environments. Superimposed on these three general forces is the effect of water molecules, ions, and other solutes within the aqueous environment which moderate the strength of these interactions. Ionized or polar groups interact with water almost as favorably as they interact with other ionized or polar groups (Creighton 1993). Nonpolar groups prefer to interact with each other rather than with water, and it is this hydrophobic interaction that is significant in the folding and dimerization of proteins. As a generalization, in most protein-protein interactions in an aqueous environment, the hydrophobic interaction is a major if not dominant driving force for dimerization.

1.8.1 Electrostatic Interactions

Electrostatic interactions involve attraction between oppositely charged particles or repulsion between like charged particles. Indeed, the origin of all intermolecular forces is fundamentally electrostatic. In addition to the classic charge-to-charge electrostatic interaction, the interactions discussed in the following sections, such as van der Waals, hydrogen bonds, and even hydrophobic interactions, all result from either permanent or induced electrostatic effects of charged electrons and charged nuclei.

In proteins, electrostatic interactions typically involve a pair of closely positioned, oppositely charged side chains. This is commonly referred to as a salt bridge. For example, an electrostatic attraction would exist between a negatively charged side chain, either Glu or Asp, paired with a positively charged side chain such as Lys or Arg. The strength of the interaction is quite variable, depending upon distance between ion pairs, and if the ion pair is buried within the

protein or exposed to the aqueous environment. Reported values for the strength of single salt bridges are in the order of 2 to 5 kcal/mol for buried salt bridges (Anderson et al. 1990; Finer-Moore et al. 1996; Waldburger et al. 1995) but close to zero for those that are solvent exposed (Fersht and Serrano 1993; Waldburger et al. 1995). Additional ions in a solvent environment interfere with attractive forces between ion pairs.

1.8.2 The van der Waals Interactions - Short Range Attractions and Repulsions

The van der Waals repulsions and attractions occur between all atoms, whether charged or uncharged, and they originate from electrostatic forces involving electrons and protons. The attractive interactions are weak, close-range forces that exist between all atoms (Creighton 1993). These ubiquitous interactions are the result of mutually induced polarization effects, known as London or dispersion forces. The repulsive interactions are created when atoms are so close that their electron orbitals overlap, resulting in repulsion between electrons.

A greatly simplified explanation of the attractive force can be derived from the classical picture of an atom with electrons orbiting the nucleus. Such a spherical atom has no net dipole moment, but it can have a transient dipole moment from a temporary asymmetric distribution of electrons. This transient dipole induces a complementary dipole in any atom nearby, creating an attraction between them. The transient dipole of the first atom rapidly changes but the dipole of the other atom tends to follow it. This dispersion force is basically electrostatic in nature and it decreases as $1/r^6$ (r=radius) with distance because the strength of the mutual induction decreases. Thus, the van der Waals interactions is strictly a short-range force. The strength of a van der Waals interaction is in the order of 0.2 kcal/mol.

The repulsive van der Waals forces are very important in terms of the energetics and structure of atoms and molecules. This repulsion occurs because the electron orbitals on different molecules cannot occupy the same space at the same time. These repulsions are nonexistent at long range but as two atoms approach each other, they reach a threshold distance where the repulsive energy rises very steeply with the inverse of the distance (Creighton 1993). Due to this steep rise

in repulsive energy, the atoms and molecules are considered to have definite dimensions and occupy volumes that are effectively impenetrable to other atoms or molecules at ordinary temperatures. This is the basis for the physical surface of all materials.

The "impenetrable" volume occupied by an atom is defined by a measure called the van der Waals radius. The van der Waals radius is usually obtained from the smallest distance that can exist between neighboring, but not covalently bonded nuclei of the same atom in a crystalline state. This distance is considered the sum of their van der Waals radii. A value for the van der Waals volume of a molecule such as an amino acid side chain can be obtained by summing up the volumes of the constituent atoms and subtracting the volume overlap due to covalent bonding (Table 1-1). This van der Waals volume will give a measure of the space that a side chain occupies. This measure of volume is useful but has limitations because it does not factor in the volume occupied by water molecules that are associated with the side chain in an aqueous environment (this varies with the hydrophilicity/hydrophobicity of the side chain), and it does not factor in differing conformational effects of the side chain.

A unification of the attractive and repulsive van der Waal forces can be described by an interaction energy profile of two atoms as a function of distance. This is commonly represented by the Lennard-Jones potential (Creighton 1993). This predicts the net energy state to be zero when the atoms are at a distance. Then the energy state decreases as the distance between atoms decreases (a net attractive force develops), reaching an energy minimum at a distance where van der Waal interactions are optimum. This is usually at 0.3-0.5 Å beyond the van der Waals contact distance. As the atomic distance decreases from this point, the energy state increases rapidly to zero and continues to infinity (a net repulsion develops) at an atomic distance representing the sum of the van der Waals radii. Thus, the van der Waals interactions are weak attractive forces at close range and strong repulsive forces at distances representing atomic contact.

1.8.3 Hydrogen Bonds

A hydrogen bond consists of an electropositive hydrogen atom "electrostatically bonded" to an electronegative atom such as oxygen or nitrogen, while being covalently bonded to a second electronegative atom such as oxygen, nitrogen, or sulfur (Creighton 1993).

$$- O^{\delta-} - H^{\delta+} \cdots {}^{\delta-}N -$$

Figure 1-7. A representation of a hydrogen bond.
Dashes represent covalent bonds, dots represent a hydrogen bond.

The electronegative atom to which the hydrogen is covalently bonded, tends to withdraw the shared pair of electrons from the hydrogen, leaving the hydrogen nucleus with a partial positive charge. This positive charge on the hydrogen interacts favorably with the partial negative charge of the hydrogen bonded electronegative atom. The hydrogen atom is unique in being able to interact with one electronegative atom while being covalently bonded to another. This is due to its relatively large charge compared to its small size.

Although there are a large variety of hydrogen bonding arrangements in proteins, hydrogen bonds typically occur between a carbonyl oxygen (>C=O) and an amide hydrogen (>N-H). Two common occurrences are between peptide backbone atoms in an alpha-helix, or between a carbonyl oxygen and a hydroxy hydrogen (-O-H), such as in a Ser side chain. There is a marked tendency for the three atoms of a hydrogen bond to be close to linear. This presumably reflects favorable energetics for this arrangement but there is considerable uncertainty over how the strength of the hydrogen bond varies with departure from linearity.

The strength of hydrogen bonds are in the range of 2-10 Kcal/mol (Creighton 1993; Janin and Chothia 1990). This variability is due in part to the variety of hydrogen bonds that can form, and also due to uncertainty in measurement. For comparison, the strength of a covalent O-H bond is 110 Kcal/mol (Rawn 1989).

1.8.4 Hydrophobic Interactions

The term hydrophobic interaction or effect seems to imply the existence of an attractive force between hydrophobic molecules. This is actually not the case; there is no such attractive force specific to hydrophobic molecules. Rather, the hydrophobic interaction results from the unfavorable interactions between polar and nonpolar molecules, driving the nonpolar molecules to seek a nonpolar (hydrophobic) environment.

When a nonpolar molecule is placed in a polar environment, the water molecules surrounding the solute, being unable to form hydrogen bonds with the nonpolar molecule, compensate by forming additional hydrogen bonds among themselves, resulting in a hydrogen bonded "iceberg" network around the nonpolar surface. This "water ordering effect" decreases the randomness (entropy) of the system. Since Thermodynamics dictates that favored processes proceed towards maximum randomness (second law of thermodynamics), this ordering effect is not favored. To reduce this unfavored water ordering effect, the nonpolar molecules will tend to minimize the surface contact with water by seeking a nonpolar environment. Thus, the nonpolar side chains of proteins tend to associate with other nonpolar side chains, rather than polar molecules of an aqueous environment. The hydrogen bonded "iceberg" network may also be less favored energetically, because these hydrogen bonds may deviate from the favored linear arrangement.

The hydrophobicities of amino acid residues vary enormously, depending on whether polar groups are present. Polar side chains tend not to interact with nonpolar solvents. They have lower solubilities in nonpolar solvents because of the unfavorable energetics of placing a polar molecule in a nonpolar environment. The magnitude of this effect varies not only with the side chain but also with the solvent. For this reason, the experimental determination of hydrophobicities for a given amino acid residue depends upon the method of measurement.

Table 1-1 gives relative hydrophobicities of the amino acid side chains, with Gly normalized to 0. Hydrophobic side chains have negative values as rated on this hydrophobic scale. These hydrophobicity values are a function of the presence of polar or charged groups on the side chains, and on the number of nonpolar hydrocarbon molecules

Side chains such as Arg, Lys, Glu, and Asp that have positively or negatively charged groups are the least hydrophobic. Side chains such as Asn, Gln, Ser, and Thr have groups capable of hydrogen bonding, so these are also not hydrophobic. The His side chain has positive charge depending on the pH of the environment and it can form hydrogen bonds; therefore it is not hydrophobic.

The term amphiphile is applied to side chains with both polar and nonpolar segments (Creighton 1993). For example, Tyr has a mixed character because it has a hydroxyl group capable of hydrogen bonding yet the bulk of the structure is a hydrophobic aromatic ring. Tyr is rated as more hydrophobic than Ala using 3 out of the 4 measurements of hydrophobicity in Table 1-1. Even Lys, which is strongly hydrophilic is capable of participating in hydrophobic interactions via its long aliphatic side chain (Wells 1996).

The side chains that have no capability of hydrogen bonding are clearly hydrophobic and their hydrophobicities are related to the size of the nonpolar side chain. Ala with its single β carbon is weakly hydrophobic and Val, Met, Phe, Ile, Leu, and Trp are moderately to strongly hydrophobic.

Table 1-1. Volume and relative hydrophobicities of amino acid side chains

Residue	van der Waals Volume (Å^3)	Hydrophobicity (kcal/mol)			
		Side chain Analogues (cyclohexane)	Amino Acids (ethanol)	N-acetyl amides (octanol)	Calculated (octanol)
Arg (R)	148	15.86	3.0	1.01	3.95
Asp (D)	91	9.66	2.5	0.77	3.81
Glu (E)	109	7.75	2.5	0.64	2.91
Asn (N)	96	7.58	0.2	0.60	1.91
Lys (K)	135	6.49	3.0	0.99	2.77
Gln (Q)	114	6.48	0.2	0.22	1.30
His (H)	118	5.60	-0.5	-0.13	0.64
Ser (S)	73	4.34	0.3	0.04	1.24
Thr (T)	93	3.51	-0.4	-0.26	1.00
Tyr (Y)	141	1.08	-2.3	-0.96	-1.47
Gly (G)	48	0	0	0	0
Pro (P)	90	na	-1.4	-0.72	-0.99
Cys (C)	86	-0.34	-1.0	-1.54	-0.25
Ala (A)	67	-0.87	-0.5	-0.31	-0.39
Trp (W)	163	-1.39	-3.4	-2.25	-2.13
Met (M)	124	-1.41	-1.3	-1.23	-0.96
Phe (F)	135	-2.04	-2.5	-1.79	-2.27
Val (V)	105	-3.10	-1.5	-1.22	-1.30
Ile (I)	124	-3.98	-1.8	-1.80	-1.82
Leu (L)	124	-3.98	-1.8	-1.70	-1.82

Data reproduced from Creighton (1993)

Four different methods of measuring or calculating relative hydrophobicities are presented, differing in the form of the side chain (side chain alone, or free amino acid, or N-acetyl amino acid amide) and the solvent used (ethanol, octanol, cyclohexane) for the measurement. In a nonaqueous solvent, the polar groups of the side chains can have varying interactions with other polar groups of the peptide backbone, and any polar groups of the solvent, including any water that is present. With hydrophobicity measurements of side chains alone, molecules that have polar groups can appear more hydrophobic than they really are. For this reason, the most extreme values of hydrophobicity have been measured for the side chain alone, without the amino acid or peptide backbone, and with the least polar solvent, cyclohexane. The moderate values of hydrophobicity, and likely the more relevant values for proteins in an aqueous environment, are derived from N-acetyl amino acid amides (which partially mimics a peptide backbone) in a less polar solvent like octanol.

1.8.5 The Thermodynamics of Protein-Protein Interactions

The process of dimerization opposes the thermodynamic tendency for an increase in entropy (disorder) of a system. For dimerization to occur spontaneously, the resultant reduction in entropy must be offset by an increase in entropy in some other part of the system, or by a decrease in enthalpy (energy state) of the system. Without presenting the details of the Gibbs Free Energy formula, it turns out that the offsetting increase in entropy comes largely from the hydrophobic interaction (Creighton 1993).

When separate hydrophobic surfaces are present in an aqueous environment, the water molecules in contact with the hydrophobic surfaces are unable to hydrogen bond with the nonpolar surface. To compensate, they form a greater number of hydrogen bonds with neighboring water molecules, resulting in a more rigid network of water on the hydrophobic surface. As described in Section 1.8.4, this "water ordering effect" decreases the entropy of the system. If the hydrophobic surfaces come together, excluding the water, then these water molecules are freed from the rigid network, and the entropy of the system increases. This increase in entropy more than offsets the

decrease in entropy resulting from dimerization. Thus, there is a net increase in entropy of the system and the dimerization process proceeds spontaneously. The hydrophobic interaction is the dominant factor in protein folding and stability (Dyson and Wright 1993; Fersht and Serrano 1993; Munson et al. 1996; Spassov et al. 1995). For proteins in an aqueous environment, the hydrophobic interaction is a significant if not dominant driving force in dimerization (Jones and Thornton 1996; Young et al. 1994). Other factors such as hydrogen bonds and electrostatic interactions also contribute enthalpically to offset the loss in entropy from dimerization (Janin and Chothia 1990; Weng et al. 1996).

1.8.6 Protein-Protein Interfaces and Specificity

In order for protein interfaces to interact specifically, there must be forces that favor interaction with one surface but not another. For van der Waal forces to favor interaction, the atoms of the two proteins must be close together because the these forces act only at close range. Thus there must be shape complementarity to the two surfaces. For hydrogen bonding, there must be polar groups adjacent to each other and in the correct orientation. For electrostatic interactions to occur, there must be charged groups in close proximity. For hydrophobic interactions to occur, there must be hydrophobic residues on both interfaces. These requirements lead to the following generalized picture of a specific protein dimer interface.

The two surfaces of the interface are complementary in shape and their atoms pack closely together (Jones and Thornton 1996). Shape complementarity is of fundamental importance because it is prerequisite to the stabilizing interactions between proteins. Polar groups on the interface are paired in hydrogen bonds and charged groups on one protein are neutralized by opposite charges on the other protein. The central core of the interface is dominated by hydrophobic side chains and polar side chains tend to occur on the periphery (Korn and Burnett 1991). This hydrophobic core provides the dominant hydrophobic interaction to bring about dimerization. Finally, the backbone structure of the two proteins is likely to be relatively rigid (Creighton 1993; Janin and Chothia 1990). In order for van der Waals, hydrogen bonding, and

electrostatic interactions to occur between one pair of proteins and not another, a defined protein structure is more likely. A flexible protein structure would tend to lack specificity because it would adopt the shape of the dimer interface and positioning of polar and charged groups to complement many different proteins (Creighton 1993; Malby et al. 1994).

An additional aspect of specificity should be discussed, namely that of positive and negative determinants of specificity. This is especially relevant to specificity among related proteins. A family of related proteins would tend to possess many common positive specificity determinants such as a similar backbone structure and similarly shaped dimerization interface, conserved hydrophobic residues, and even conserved charged residues. One mechanism of differentiating against several closely related proteins may have been the evolution of 'negative' or "inhibitory" determinants to actively prevent unwanted interactions. This model has been proposed for the mechanism of specificity between related hormone-receptor combinations, whereby binding between a family of evolutionary related protein hormones and receptors is restricted by steric hindrance domains that only allow the desired combinations to interact (Moyle et al. 1994; Tian et al. 1996). An inhibitory determinant model may be applicable to the problem is specificity among allelic bE and bW proteins

1.9 Objectives

I have presented *U. maydis* as an organism with: (1) an interesting multiallelic recognition system encoded by the *b* locus, which allows cells to mate with multiple "nonself" mating types but prevents mating with "self" , and (2) relatively simple genetics that are amenable to study and manipulation. Thus, the potential exists in this system to determine how multiple cell types recognize self from nonself through the use of well designed genetic experiments. Knowledge gained about the mechanism of *b* locus recognition is likely to provide insight into how other fungi, and even higher eukaryotes such as plants and animals might also carry out similar multiple recognition processes.

Our present knowledge of the mechanism of the *b* locus indicates that recognition is the result of dimerization between the two *b* locus proteins, bE and bW. The N-terminal domains of both proteins are involved in this dimer formation; they somehow allow dimerization between bE and bW proteins derived from different alleles, but prevent dimerization when bE and bW are from the same allele. The question remaining to be answered is, what is the mechanism of this specificity or "anti-specificity" in dimerization?

We know that the answer to this question lies in the nature of the protein-protein interactions between bE and bW; thus we need information on the amino acid residues on the dimer interface. Obviously, the determination of the three-dimensional structures of the bE and bW monomers and of the heterodimer would provide many of the answers. Unfortunately, these structures are not available at this time, and even if they were, functional data is still necessary to corroborate the structural data. Obtaining functional data on the amino acid residues of the dimer interface should be possible using genetic techniques.

With this introduction to the problem of specificity at *b* locus, I would like to outline my objectives in this body of research:

1. Construct chimeric alleles of *bE* and *bW*.
2. Determine through analysis of these chimeras, which amino acid positions in bE and bW are involved in dimerization specificity.
3. Obtain functional data as to the nature of amino acid changes that affect allelic specificity.
4. Develop a model for dimerization specificity at the *b* locus.

2. MATERIALS AND METHODS

2.1 Strains and Media

Escherichia coli strain DH5α (F-, *end*A1, *hsd*R17(r_k^-,m_k^+), supE44, thi$^-$, *rec*A1, ϕ80d*lac* ZM15) from Bethesda Research Laboratories was used for routine DNA manipulation work and was grown in LB medium (Sambrook et al. 1989). *U. maydis* wild-type prototrophic strains were 518 *(a2b2),* 521 *(a1b1),* 031 *(a1b2),* and 032 *(a2b1)* and ATCC 18604 *(a1b3).* The first four *b1* and *b2* strains originated from R. Holliday (Commonwealth Scientific and Industrial Research Organization, Laboratory of Molecular Biology, Sydney, Australia) and the *b3* strain was obtained from the American Type Culture Collection. An additional *b3* strain designated 023 *(a2b3)* was isolated as haploid progeny of a cross between ATCC 18604 *(a1b3)* and 032 *(a2b1).* *U. maydis* strains carrying chimeric and mutant alleles are listed in Table 3-1 and Table 3-2. Throughout this thesis, the gene and allele names will be given in italics and the corresponding proteins will be nonitalicized. *U. maydis* cultures were grown in potato dextrose medium (Difco Laboratories) or complete medium (Holliday 1974) and mating reactions were carried out on double complete charcoal plates (Holliday 1974). Media recipes are given in Appendix 6.1.

2.2 Recombinant DNA Techniques

Standard protocols were used for DNA manipulations (Ausubel 1995; Sambrook et al. 1989). Most of the DNA subcloning was accomplished using low melting point agarose ligation (Kalvakolanu and Livingston 1991) and small scale alkaline plasmid preparations were done by the "ten minute" protocol of Zhou et al. (1990). Restriction and DNA modifying enzymes were obtained from Bethesda Research Laboratories, Boehringer Mannheim, or New England Biolabs, and they were used according to the manufacturer's instructions.

2.2.1 Construction of the Plasmid Series pAR50 to pAR69

An 800 bp fragment consisting of 200 bp of the *bW1* promoter and 600 bp encoding the N-terminal portion of the *bW1* open reading frame (containing the *bW1* variable region) was

amplified by PCR using the primers BW1 and BW6 (Table 2-2) on template pUC18b1 (Figure 2-1). These primers have *Bam*HI recognition sites at their 5' ends. This PCR product was cloned into the *Bam*HI site of vector pUC9 and sequenced by the dideoxy termination method (Sanger et al. 1977) using the Pharmacia T7 Sequencing Kit, and this plasmid was called pbW1v. The 800 bp *Bam*HI insert from pbW1v was inserted into plasmid pBAD8-2 (Kronstad and Leong 1990) replacing the 1.5 kb *Bam*HI-*Bgl*II fragment carrying *bW1* sequences, giving plasmid pBED. Then, the 2.9 kb *Bam*HI-*Sal*I fragment from pBED, containing the *bW1* promoter and variable region, the Hsp70/HygB marker, and part of the *bE1* constant region, was inserted between the *Bam*HI and *Sal*I sites of pGEM-3Zf+ (Promega, Madison, WI) and this construct was designated pAR30. Plasmid pAR30 was digested with *Kpn*I and *Bam*HI and progressive deletions were carried out from the *Bam*HI site using Exonuclease I and S1 nuclease (Erase-a-Base System, Promega, Madison, WI). The deletion products were recircularized and transformed into *E. coli.* The resulting clones were screened for *bW1* variable region deletions of up to 750 bp, at intervals of approximately 30 bp. A set of 19 plasmids were selected and designated pAR31 to pAR49.

Plasmids pAR30 to pAR49 were further manipulated to add a greater length of *bE1* sequence by removal of the 2.5 kb *Xba*I-*Sac*I fragments containing deletions of the *bW1* variable region and the Hsp70/Hygromycin marker, and combining them individually into a construct consisting of the 6 kb *Bam*HI-*Xba*I fragment from pUC18b1 (Figure 2-1) inserted between the *Bam*HI and *Xba*I sites of pBluescript KS+. These plasmids were designated pAR 50 to pAR69 (Figure 2-2 and Figure 2-3). They were used for *in vivo* construction of chimeric *bW1-bW2* alleles by transformation into *U. maydis* and homologous integration into the *b* locus. The *bE* ORF was disrupted in these plasmids by the hygromycin marker, therefore, transformants with chimeric *bW* alleles did not have functional *bE* ORFs.

Large scale preparations of these plasmids were produced using the Promega Wizard Maxi-Prep columns. These plasmids were linearized for transformation of *U. maydis* by digesting with *Bam*HI and *Sac*I. The exact location of the deletion endpoint of each construct was determined by double-stranded DNA sequencing described in the next section.

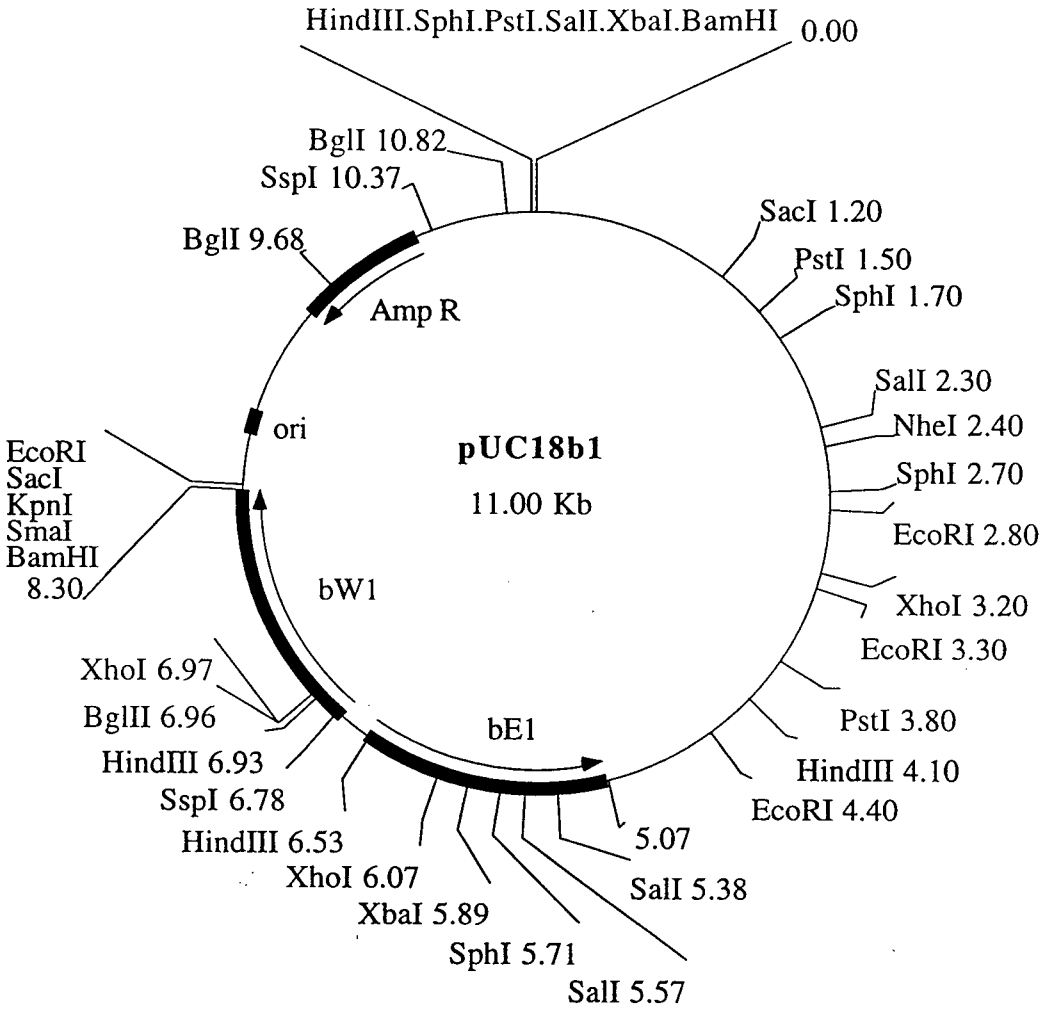

Figure 2-1. Plasmid map of pUC18b1.
This plasmid consists of 8.3 kb (reported as 8.5 kb) *Bam*HI fragment containing the *b1* locus inserted into the *Bam*HI site of pUC18 (Kronstad and Leong 1989).

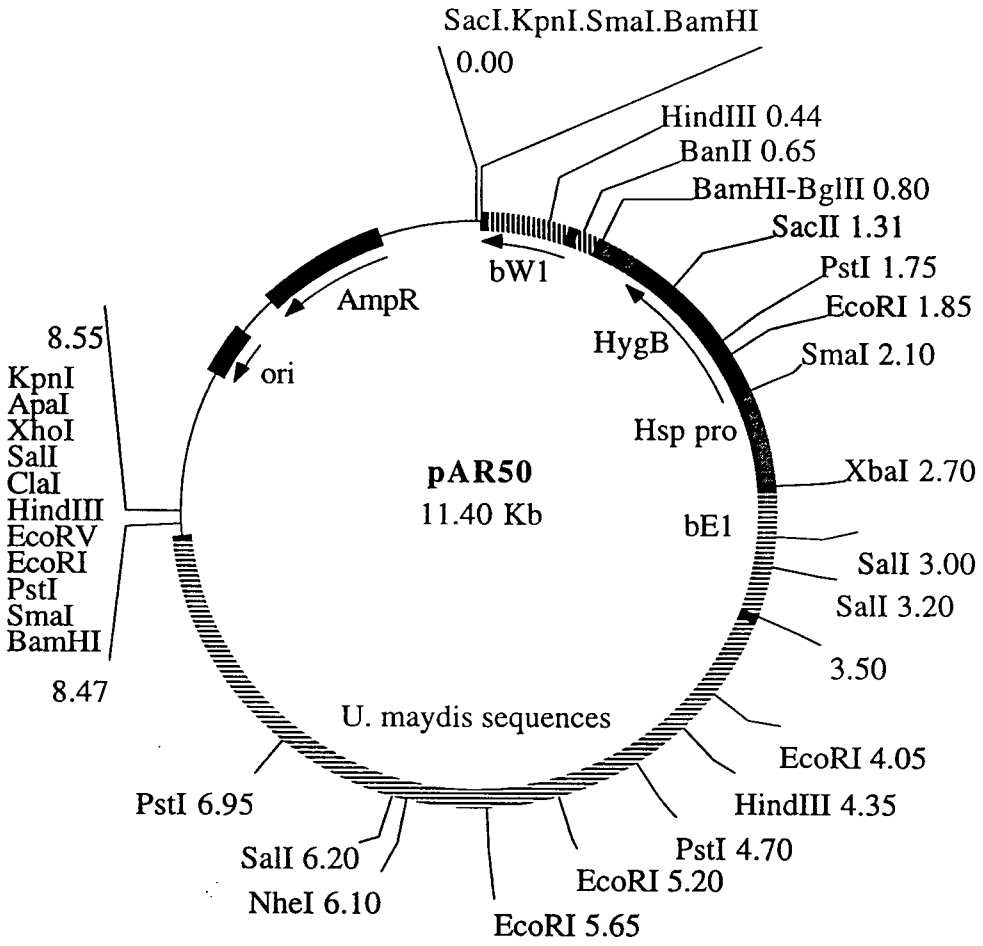

Figure 2-2. Plasmid map of pAR50.
Parent vector for pAR50 was pBluescript KS+. The *bE1* sequences originated from the 6 kb *Bam*HI-*Xba*I fragment from pUC18b1 (Figure 2-1). The *Xba*I-*Bgl*II HygB/Hsp cassette originated from pHL1 (Wang et al. 1988) via plasmid pBAD8-2 (Kronstad and Leong 1990). The bW1 sequences were obtained by PCR with primers BW1 and BW6 (Table 2-2) on template pUC18b1 (Figure 2-1).

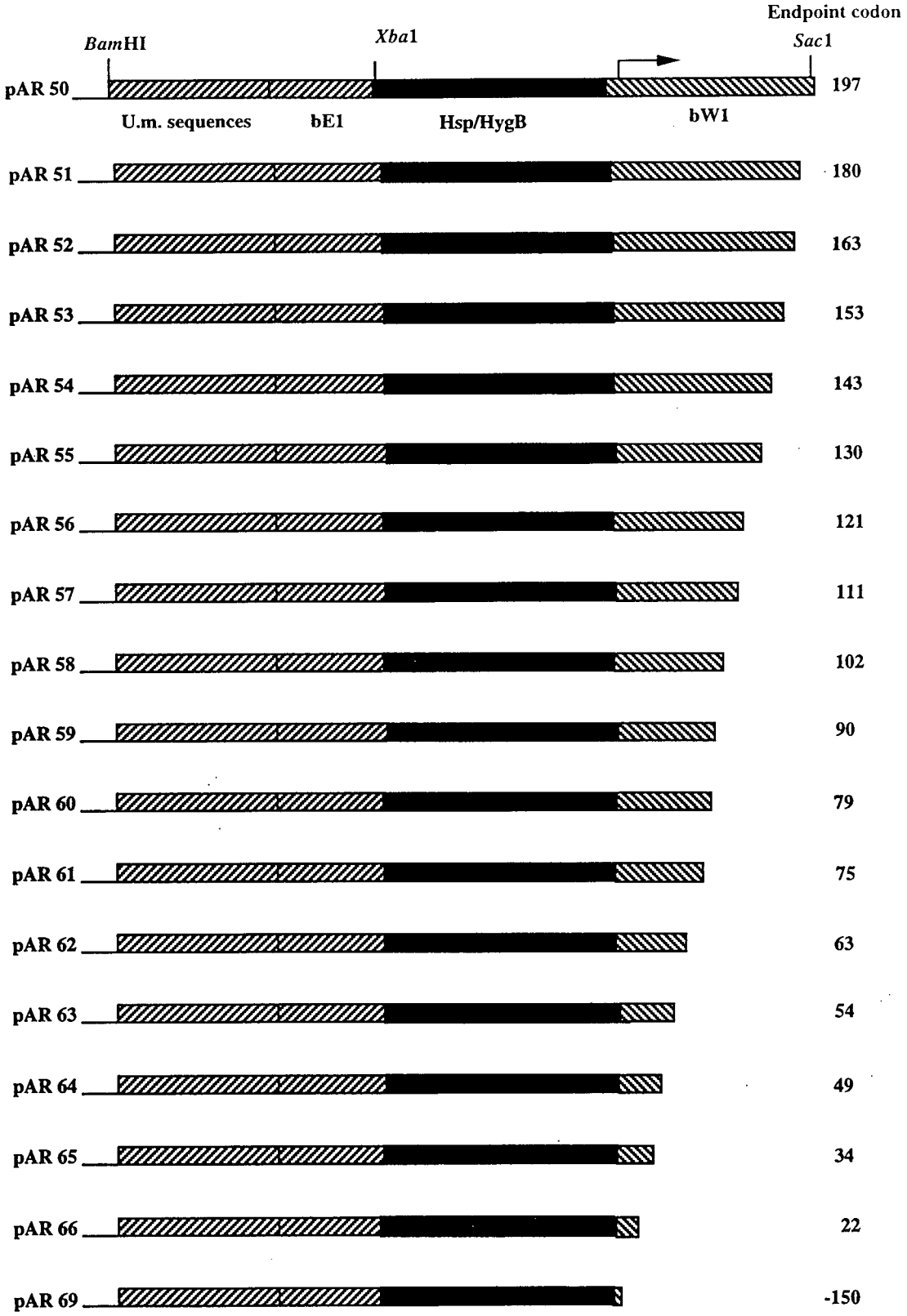

Figure 2-3. Partial Linear Maps of pAR 50 to pAR69.
These were obtained by Exonuclease III-S1 nuclease digestion from the *Bam*HI site of pAR30, a predecessor to pAR50 (Figure 2-2; see description in Section 2.2.1).

2.2.2 PCR of Genomic DNA and DNA Sequencing

PCR of *U. maydis* genomic DNA was performed with a Perkin-Elmer Cetus DNA Thermal Cycler in 0.5 ml GeneAmp tubes. Cycling parameters were as follows; initial 5 min denaturation at 94°C, followed by 30 cycles of 1 min at 94°C, 1 min at 55°C, and 1 min at 74°C, and a final extension step of 10 min at 74°C. PCR products were held at 4°C for overnight storage or at -20°C for longer periods. Reactions were carried out in volumes of 50 to 100 ul containing 10 mM Tris-HCl pH 8.3, 50 mM KCl, 2.0 mM MgCl, 0.001% gelatin, 100 uM each of dATP, dCTP, dGTP, dTTP, 0.2 uM of each primer, 2.5 U of Taq Polymerase (Perkin Elmer Cetus) per 100 ul reaction, and about 1 ug of genomic DNA from small scale preparations (Elder et al. 1993).

Nucleotide sequences of chimeric *bW* alleles were obtained by PCR amplification of the genomic small scale preparations, followed by direct sequencing of the PCR product by dideoxynucleotide chain termination sequencing (Sanger et al. 1977), either by thermal cycle sequencing using Taq Polymerase with the AmpliTaq Cycle Sequencing Kit (Perkin Elmer, Norwalk, Connecticut) and ^{32}P labeling by incorporation, or by using T7 polymerase with the Sequenase PCR Product Sequencing Kit (United States Biochemical, Cleveland, Ohio) and ^{35}S labeling. For AmpliTaq Cycle Sequencing, the PCR products were first purified by gel electrophoresis through low melting point agarose (Ultra Pure Gibco BRL, Gaithersburg, MD) and isolated by a "melt-freeze" procedure (Qian and Wilkinson 1991). For T7 Sequenase PCR Product Sequencing, no purification of the PCR product was necessary. This procedure uses Shrimp Alkaline Phosphatase and Exonuclease III to inactivate or remove the dNTP's and primers before standard T7 dideoxy termination sequencing.

Sequencing of double stranded plasmid DNA was done from Promega Wizard plasmid preparations or from purified small-scale plasmid preparations. Dimethyl sulfoxide (DMSO) denaturation was employed, adapted from Drebot and Lee (1994) followed by dideoxy termination sequencing (Sanger et al. 1977) using the T7 Sequencing Kit (Pharmacia LKB Biotechnology) with ^{35}S labeling.

Prior to electrophoresis, all sequencing reactions were denatured at 80°C for 2 min and kept on ice until loaded. Electrophoresis of reactions was done through 5% Long Ranger (J.T. Baker Inc. Phillipsburg, NJ, USA) acrylamide-50% urea using a BRL Model S2 sequencing apparatus at 65 Watts with 0.6% TBE running buffer.

2.2.3 *In vitro* Construction of Chimeric Alleles and Site-Directed Mutagenesis

In vitro gene manipulation was carried out using PCR techniques and these chimeric or mutant gene products were cloned into a *U. maydis* transformation construct consisting of a Hsp70 promoter and terminator, and a Hygromycin B resistance gene, flanked by *b* locus sequences to allow targeted gene replacement. This mutagenesis process is diagrammed for *bE* in Figure 2-4 and for *bW* in Figure 2-5.

The PCR technique called megaprimer PCR (Sarkar and Sommer 1990) was used for the *in vitro* construction of chimeric alleles and for site-directed mutagenesis. This technique requires only one mutagenesis primer, plus two flanking primers, and it uses two rounds of PCR. The first round of PCR involves the mutagenesis primer and a flanking second primer to produce a PCR product called the "megaprimer". This megaprimer is used in a second round of PCR along with a flanking third primer to produce the final mutant product. The construction of chimeric alleles was done in the same way as mutagenesis except template from a different allele was used for the second round of PCR.

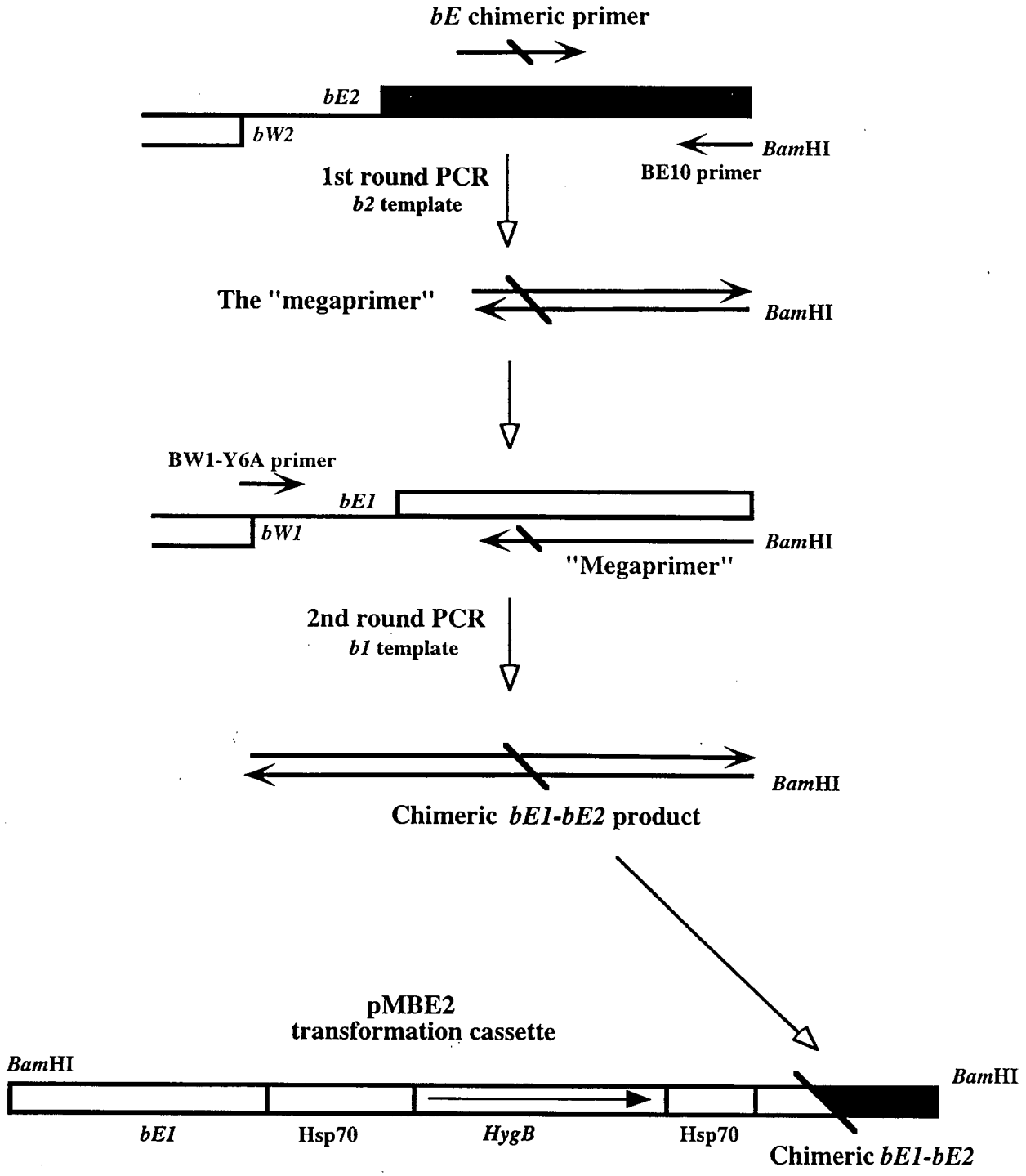

Figure 2-4. *In vitro bE* gene manipulation and gene replacement.

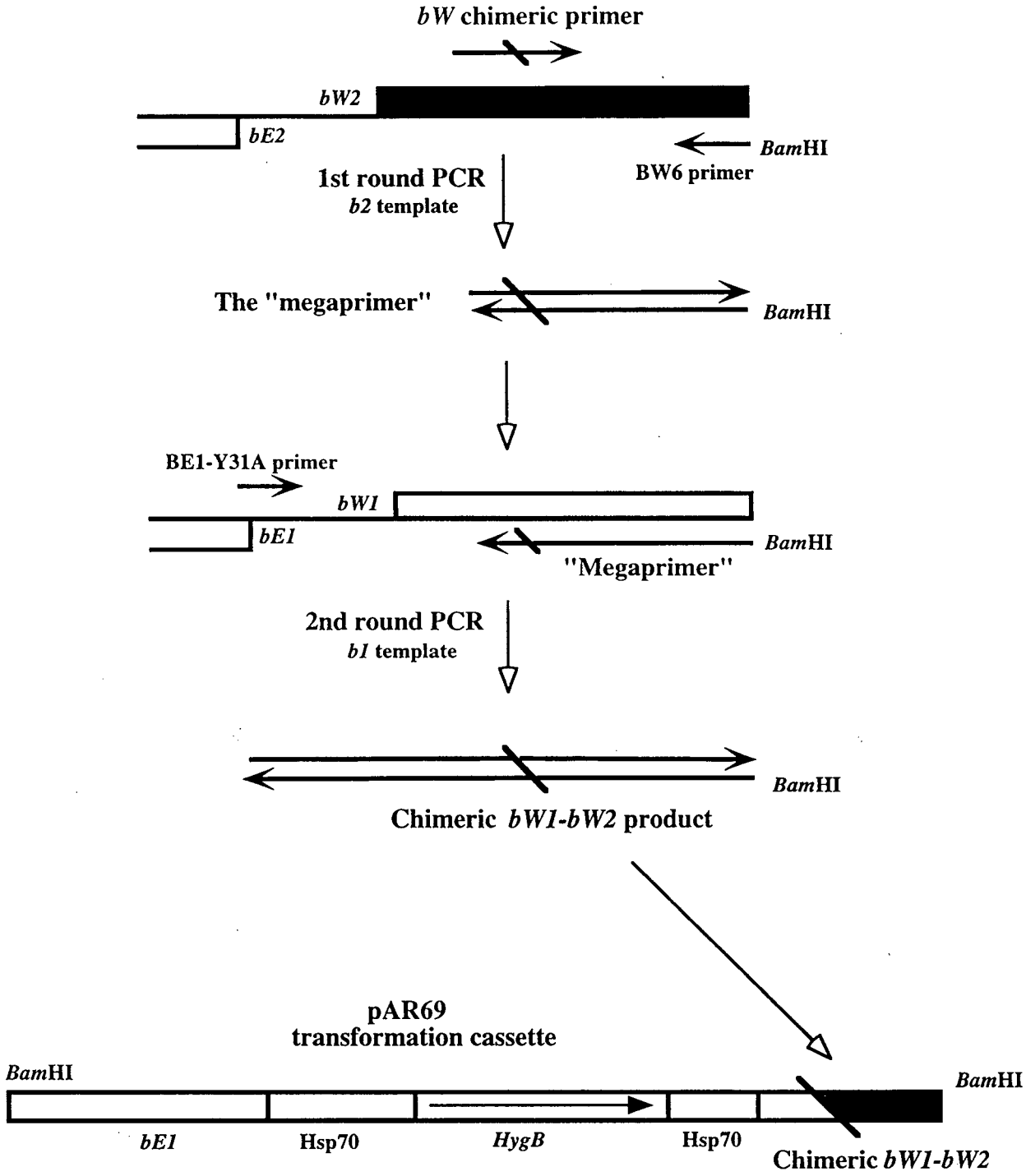

Figure 2-5. *In vitro bW* gene manipulation and gene replacement.

The protocol used was that of Sarkar and Sommer (1990) except that Vent Polymerase (New England BioLabs, Beverly, MA) was substituted for Taq Polymerase. Vent polymerase has a 3'-5' proofreading capability which increases the accuracy of nucleotide incorporation and prevents nontemplate derived terminal additions of nucleotides to the 3' end of the PCR product. These nontemplate derived terminal additions to the first round PCR product can interfere with the extension from this product during the second round of PCR. Conditions for PCR with Vent polymerase, as recommended by Cease et al. (1994), were 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris-HCl pH 8.8 at 25°C, 3 mM $MgSO_4$, 500 uM each dNTP, 0.1% Triton X-100, 1 uM primers, 1 femtomole of DNA template and 1 U of Vent polymerase per 100 ul reaction. Thermal cycling was done on a Perkin-Elmer Cetus DNA Thermal Cycler with an initial 3 min time delay at 94°C, then a step cycle of 1 min at 94°C, 1 min at 52°C, and 1 min at 72°C, and a 10 min final extension at 72°C. PCR products were stored at 4°C overnight or at -20°C for periods longer than 24 hr.

The chimer/mutagenesis primers (Table 2-1) used for *bE* were MBEG28 (BEX31), BEX45, BEX90-A90Y, BEX90-A90F, BEX90-A90L, BEX90-A90M, each used individually in combination with the flanking second and third primers, BW1-Y6A and BE10 respectively. For *bW*, the chimer/mutagenesis primers (Table 2-2) were MBWL2 (BWX4), BWX6, BWX68, BWX72, BWX73, BWX74, BWX76, BWX77, BWX79, BWX80, BWX81, BWX82, BWX83, BWX91, BWX74-H74A, BWX74-H74V, BWX-H74F, BWX74-H74F, BWX740H74L, BWX74-H74M , each used individually with the flanking second and third primers BE1-Y31A and BW6 respectively. A map showing the location of these primers within the *b* locus is given in Figure 2-6.

The chimeric or mutant PCR product was first cloned as a blunt end fragment into the *Eco*RI site of pBluescript KS+ . This cloned chimera or mutant product was then subcloned from pBluescript KS+ into a *Ustilago* transformation plasmid, designed to integrate homologously into the *b* locus. For *bE* mutagenesis, the cloned *bE* chimeric or mutant PCR product was inserted as a blunt end fragment into pMBE2 (Figure 2-7) linearized with *Hind*III, made blunt with T4

polymerase, and dephosphorylated. The plasmid pMBE2 was derived from plasmid pAR (Yee and Kronstad 1993) by deleting the *Bgl*II to *Sal*I fragment containing the Hsp70/HygB cassette and *bE1* sequences, and replacing this with a 1.9 kb *Bgl*II to *Sal*I fragment from pIC19HHL1 (Figure 2-8) containing only the Hsp70/HygB cassette. For *bW* mutagenesis, the cloned *bW* chimeric or mutant PCR product was inserted as a blunt end fragment into the *Sac*I site of pAR69 (described in section 2.2.1 and Figure 2-9). Chimeric/mutagenesis derivatives of pMBE2 and pAR69 containing the correct orientation of insert were chosen and small scale plasmid preparations were made of these constructs for transformation into *U. maydis*. Thus, the final organization of the chimeric/mutagenesis plasmids consisted of a Hsp70/HygB cassette flanked on one side by a disrupted *bW* or *bE* ORF, and on the other side by the promoter and mutated N-terminus of the other *b* locus ORF. The result of homologous integration at the *b* locus would be a mutated *bE* or *bW* gene, a Hsp70/HygB cassette, and a disrupted partner *b* locus gene.

Both the *bE* and *bW* chimeric/mutagenesis constructs were linearized with *Bam*HI, extracted once with phenol:chloroform:isoamyl alcohol (24:24:1) and once with chloroform:isoamyl alcohol (24:1), ethanol precipitated and redissolved in TE. This DNA was then used to transform *U. maydis* protoplasts according to the protocol described in section 2.3.1. Transformants were then selected for homologous integration and one-step gene replacement by the PCR technique described in Section 2.2.4.

Table 2-1. Oligonucleotide primers for *bE*.

Primer Name	Primer Sequence (5' TO 3')	Size
BE1	CTG GAT CCG GTT TGT GAG ATT CTT GG	26
BE1A	AAC GGA TGC TCA TAA GCC TCC TCG TAT	27
BE1B	CAT TCA TTT CCT TGC GCC GAC ACC G	25
BE1C	TTC CTT GCG CCG ACA CCG	18
BE2	AAG GAT CCT CTC AGC AAC ATC ATC AC	26
BE2-1	ACG AAT TCA CAA TGT CCA GCG ACC CG	26
BE2-2	ACG AAT TCA CAA TGT CTA ACT ACC CG	26
BE3	AGA CTA CAC ACA GGA TTA	18
BE3A	TCA AAT ACA CCA GAC TAC TCA CAG G	25
BE3-1	CAC ATA GGA TTG AAG TCG	18
BE4	CTC GAG GTT CAT CAG CTC A	19
BE5	CCT GTT GAC AAG CAG CA	17
BE6	GCT AGA GTC GAC TCC AGA	18
BE6A	GCG GTC AGA CCA TCT CGT	18
BE6B	TGG GAT CCT GAT TGA GAA CAA CAG AC	26
BE6C	CGG ATC CCT GAA ATC AGC CAA ACG C	25
BE7	ATG GAT CCT GTT TGT CTG TGT GGC AG	26
BE7-1	GGG ATT GTA GTG AGA CAT	18
BE8	GCT GAG TTC TGG AGT CG	17
BE9	CAT CTT CGA ACT GCT GC	17
BE10	AAG GAT CCA TAG CGT GAG CTG ATG A	25
BE10A	CGA GGA TCC GTG AAC GAT CCT CGC GTG C	28
BE11	CTG TGT GTA GTC TGG TGA	18
BE11-1	GAA ACT CGT GCT CGA TCT CGT	21
BE11-2	ATC CTG TGT GCA GTC TGG	18
BE11-3	AGC CTA TGA GCA ATC TGG	18
BE12	GAC ATT GTG ATG ATG TTG	18
MBEG28	ACG GGG TAA TTT TCC CCT TTA TCT CGC	27
BEX42-Q39E	GTC TTT CGT CGC AGC TCT TCG AGT TTA CG	29
BEX42a	GTC TTT CGT CGC AGC TCT CGG AG	23
BEX43	GTT GTT GGG TGT CTT TCG TTG CAG CTC TCG GAG	33
BEX45	AAC GTT GTT GGG TGT CTT TTG TTG CAG CTC TCG	33
BEX90-A90Y	GTC TCT CTG TGT TCC GAG TAC GTT GAA GAC ACA	36
BEX90-A90F	GTC TCT CTG TGT TCC GAG TTC GTT GAA GAC ACA	36
BEX90-A90L	GTC TCT CTG TGT TCC GAG CTT GTT GAA GAC ACA	36
BEX90-A90M	GTC TCT CTG TGT TCC GAG ATG GTT GAA GAC ACA	36
BE1-Y31A	ACG CAC GAG AAC GGG GGC ATT CTC CCC	27
BE1-V90A	TCT CTG TGT TCC GAG GCC GTT CAC GGC	27

Table 2-2. Oligonucleotide primers for *bW*.

Primer Name	Primer Sequence (5' TO 3')	Size
BW1	CGG GAT CCT GGA CAT TGT GAT GAT G	25
BW1A	CTC AGC GCA GGT TCT TTG	18
BW1B	AAC CGC ACC AAG ACC GAC	18
BW1C	GGA GGC TTA TGA TCA GGC	18
BW2	CCG GTA CCA AGC TTC GAC AAT TTG C	25
BW3	ATG GAT CCT CAT ACA CTC GTC GTA G	25
BW4	TCG GAT CCA AAG ACT TTG CTG ACC TC	26
BW5	GGG GAT CCT TCA TAG GAA GCA TT	23
BW6	TTG GAT CCA GTG ACC TCT GAA AG	23
BW7	GCA CTC ACC CAG ATA GTG ACT TGC	24
BWX140	CAG GGG GAG AAG AAT CAG TTC G	22
MBWL2 (BWX4)	AAA CAT TCA AGA TCT TTC ATG TTG GG	26
BWX6	ATC TCG GAG AAA CAT TCA AAA TCT TTC ATG TTG	33
BWX68	GTC TGT CGA GAT ACA CTC ATC AAG CTT	27
BWX72	CTC GAG GAG ATC TTC TTG GGA CAT TTG	27
BWX79	GAG TAT CTG AGG AAG CTA CAC ATA GGG	27
BWX88	CGA CGA GTG TAT GAG GCT CAG TAC GAG	27
BWX74	GAG GAG ATC TTC ATA GAG CAT TTG AAC GAA CTA	33
BWX77	TTC ATA GAG TAT CTG AGG GAA CTA CAC ATA GGG	33
BWX80	TAT CTG AGG AAG CTA CGA ATA GGG TGC CAA GCT	33
BWX81	CTG AGG AAG CTA CGA CGA GGG TGC CAA GCT CAG	33
BWX82	AGG AAG CTA CGA CGA GTG TGC CAA GCT CAG TAC	33
BWX73	CTC GAG GAG ATC TTC ATA GGA CAT TTG AAC GAA	33
BWX76	ATC TTC ATA GAG TAT CTG AAC GAA CTA CAC ATA	33
BWX83	AAG CTA CGA CGA GTG TAT CAA GCT CAG TAC GAG	33
BWX91	CAA TAC GAA AAT GCG TTC GCG ATA TGG	30
BWX74-H74A	GAG GAG ATC TTC ATA GAG GCT TTG AAC GAA CTA	36
BWX74-H74V	GAG GAG ATC TTC ATA GAG GTT TTG AAC GAA CTA	36
BWX74-H74F	GAG GAG ATC TTC ATA GAG TTC TTG AAC GAA CTA	36
BWX74-H74L	GAG GAG ATC TTC ATA GAG CTT TTG AAC GAA CTA	36
BWX74-H74M	GAG GAG ATC TTC ATA GAG ATG TTG AAC GAA CTA	36
BW1-Y6A	CAA AAT CTT GGA GAA AGC TTC AAA ATC TTT CAT	33
BW1-Y74A	GAG ATC TTC ATA GAG GCT CTG AGG AAG	27

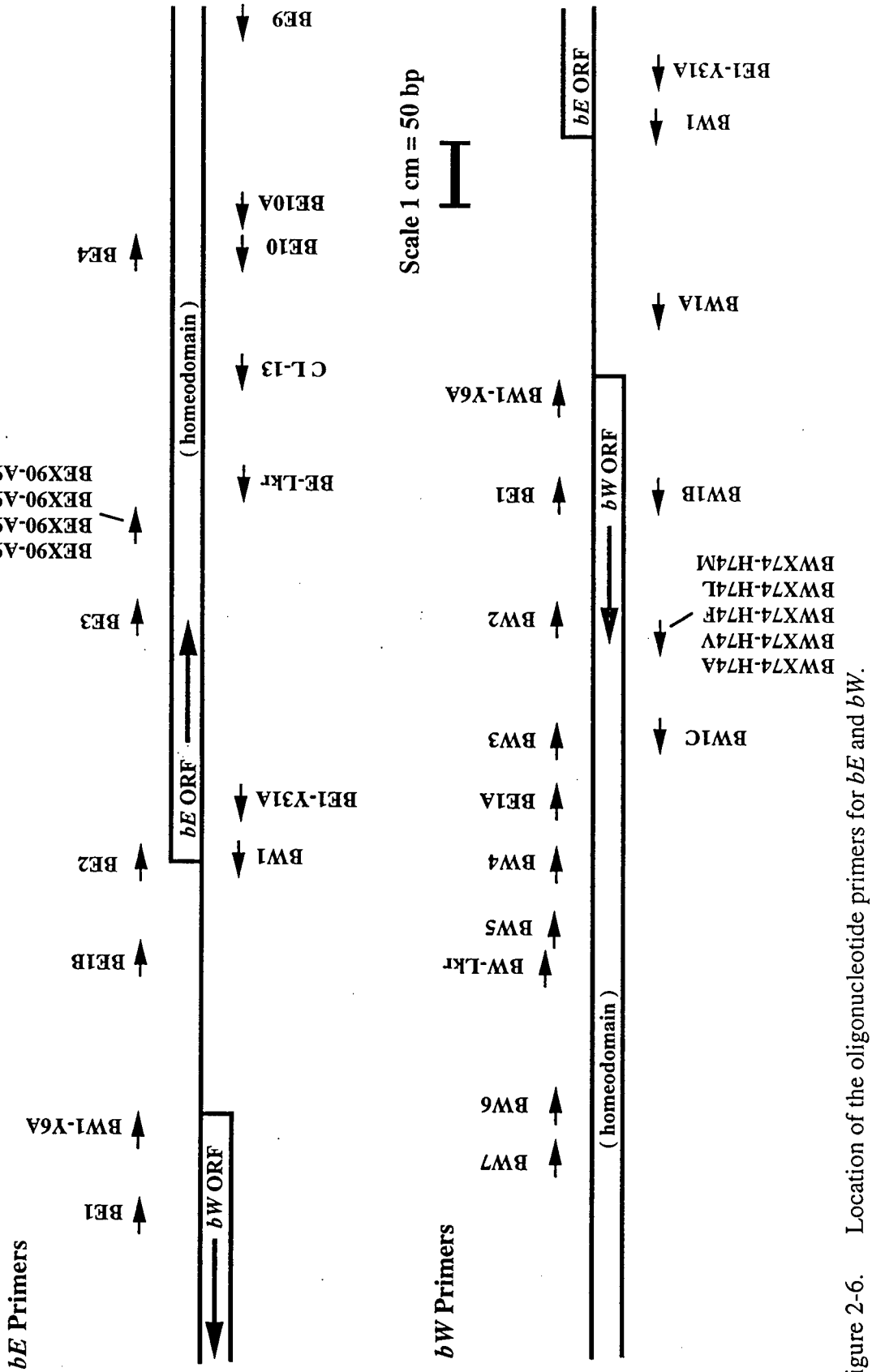

Figure 2-6. Location of the oligonucleotide primers for *bE* and *bW*.

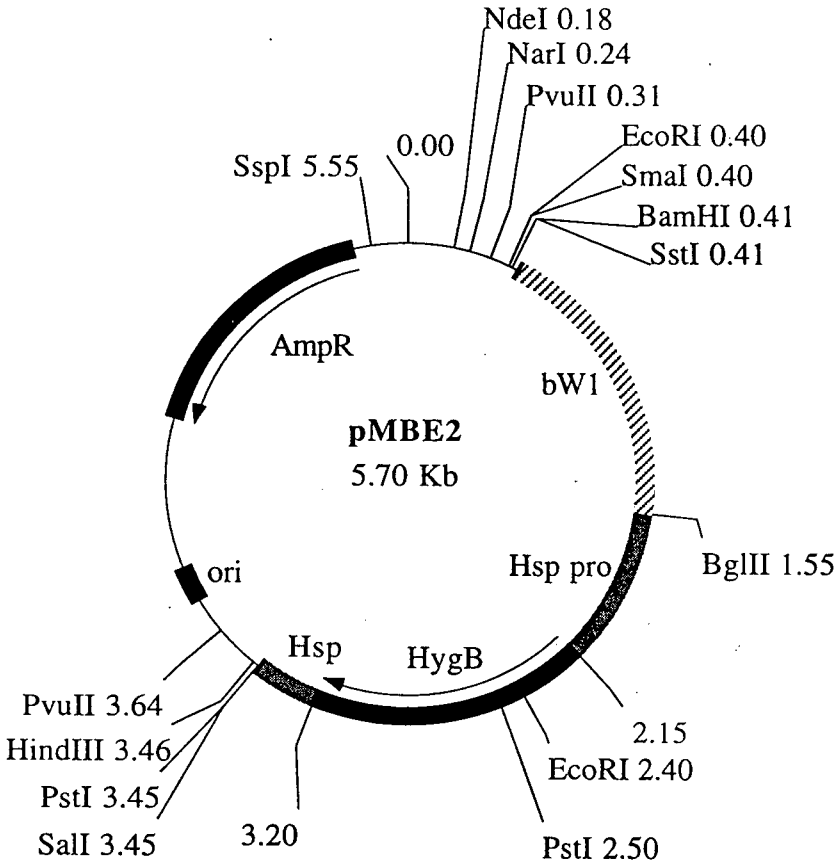

Figure 2-7. Plasmid map of pMBE2.
This plasmid was used for the construction of mutagenesis plasmids for *bE*. Derived from plasmid pAR (Yee and Kronstad 1993) by digesting with *Bgl*II and *Sal*I, then inserting the 1.8 kb *Bgl*II to *Sal*I fragment containing the HygB/Hsp cassette from pIC19HHL2 (see Figure 2-8).

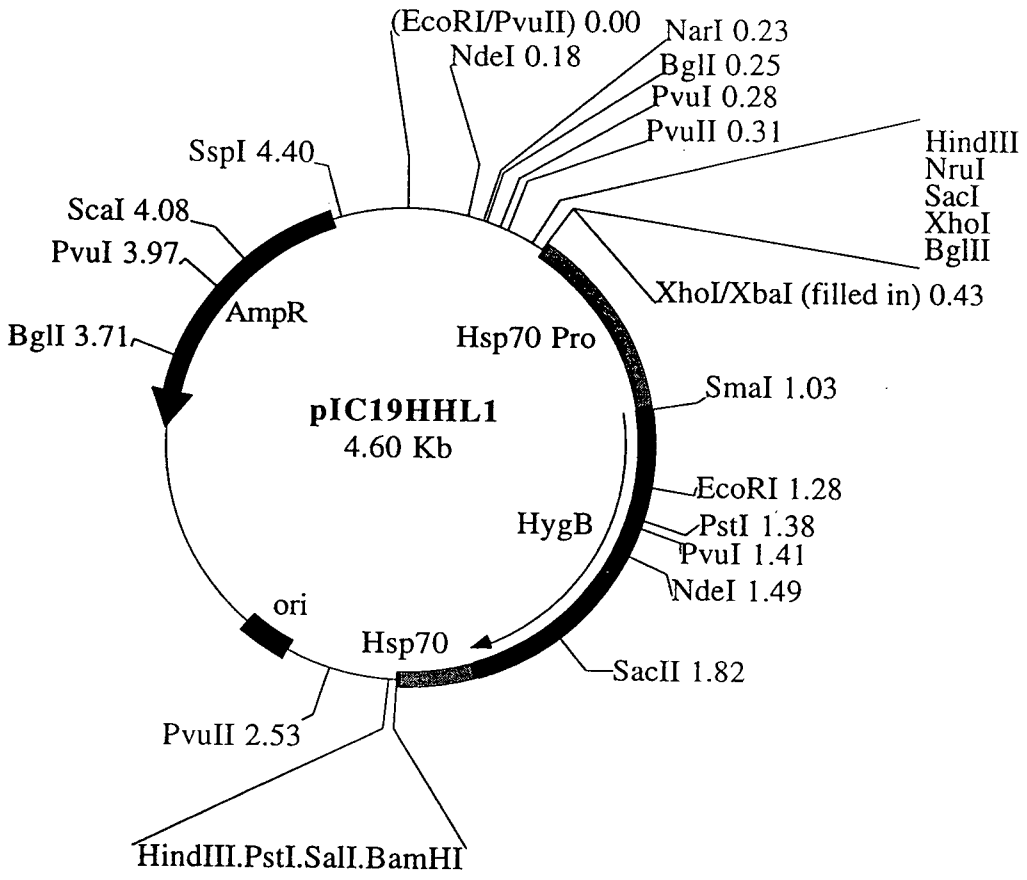

Figure 2-8. Plasmid map of pIC19HHL1.
The 2.7 kb *Hind*III to *Xho*I fragment of pHL1 (Wang et al. 1988) containing the Hsp70/HygB cassette was first inserted between the *Hind*III and *Xho*I sites in the polylinker of the pIC19R ((Marsh et al. 1984) to give the plasmid pIC19RHL (carried out by J.W. Kronstad). This was then *Xho*I and *Xba*I digested, treated with T4 polymerase to blunt the ends, and recircularized; then *Sal*I digested, treated with T4 polymerase, and recircularized. These two deletion procedures reduced the size of the Hsp70/HygB cassette to 1.9 kb. This cassette was then excised with *Bam*HI and *Sal*I and inserted between the *Bam*HI and *Bgl*II sites of pIC19H (Marsh et al. 1984) to give pIC19HHL1. The unique *Sma*I site was removed by *Xma*I digestion and recircularization to give pIC19HHL2.

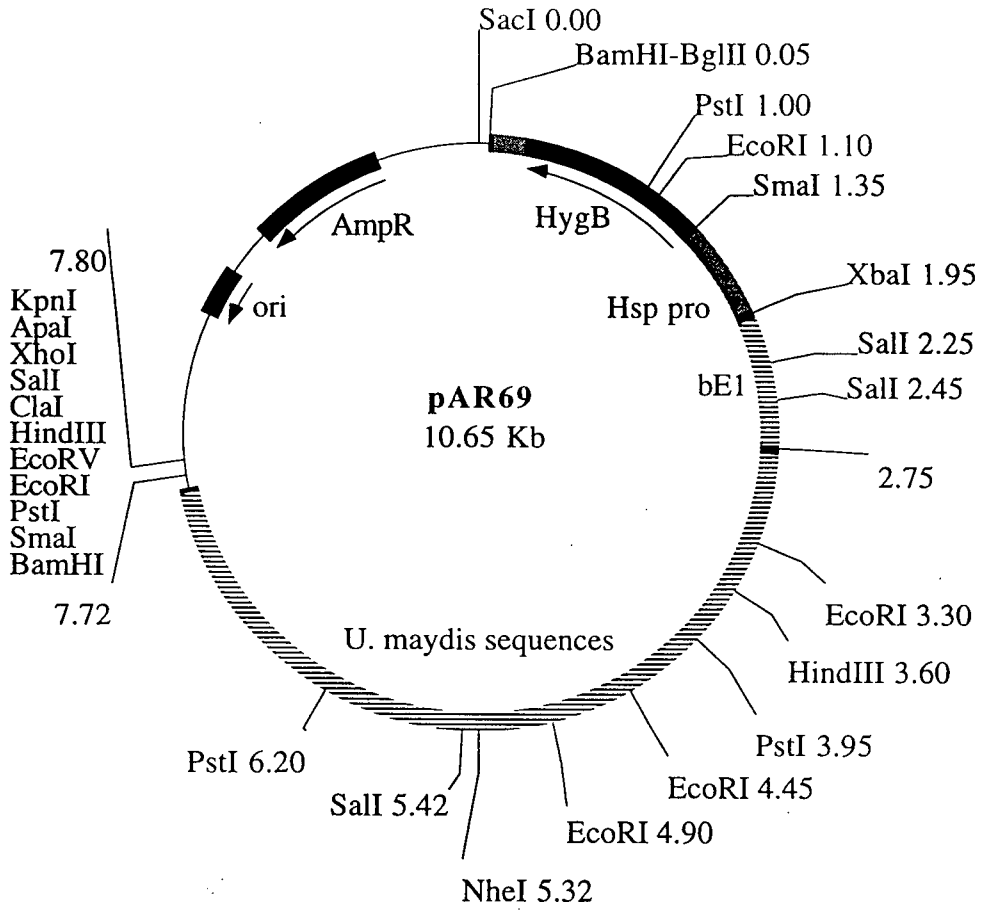

Figure 2-9. Plasmid map of pAR69.
The plasmid pAR69 contains a disrupted *bE1* ORF, the Hygromycin/Hsp70 cassette, and about 50 bp of a remnant *bW1* promoter region. The cloned *bW* chimeric or mutant PCR product was inserted into pAR69 previously linearized with *Hind*III, made blunt with T4 polymerase, and dephosphorylated.

2.2.4 Direct Cell PCR Screen for Homologous Integration in *U. maydis*

A rapid method for detecting homologous integration and one-step gene replacement in *U. maydis* was developed by adaptation of direct cell PCR techniques on *Saccharomyces* (Sathe 1991) and *Aspergillus* (Aufauvre-Brown 1993) cells. A small clump of cells from *U. maydis* transformants was transferred with a wooden toothpick from solid PDA media to 20 ul of sterile distilled water in a 0.5 ml thinwall MicroAmp (Perkin Elmer) reaction tube. Then, the reaction buffer components, primers, and polymerase were added to a final volume of 50 ul. The PCR reaction conditions were 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 40 mM Tris-HCl approximately pH 9, 2 mM $MgSO_4$, 100 uM each dNTP, 100 ug/ml BSA, 0.1% Triton X-100, 1 uM primers, and 2.5 U of Taq polymerase mixed with 0.025 U of Vent polymerase per 50 ul reaction. These conditions were essentially Vent Polymerase reaction conditions modified by increasing the alkaline buffering, adding BSA, and using a Taq:Vent polymerase mix. These modifications, adapted from "long distance PCR" (Barnes 1994; Cheng et al. 1994) allowed consistent amplification of PCR products up to 2 kb in length directly from *Ustilago* cells. Thermal cycling was performed on a Perkin-Elmer 9600, with an initial 4 min at 94°C, then 30 cycles of 30 sec at 94°C, 30 sec at 55°C, and 30 sec at 72°C, and a final 10 min extension at 72°C.

The primers used to detect homologous integration were designed to give a PCR product only if the transforming DNA had integrated at the *b* locus. One of the primers, designated Hph1 (5'-CGA GGA TCC GAG GGC AAA GGA ATA GAG -3') was located at the stop codon of the hygromycin resistance gene (Gritz and Davies 1983) and the other primer was located within the *b* locus but just outside of the transforming *b* sequences. Integration of the transforming DNA in any location other than the *b* locus would put these two primers too far apart to give a PCR product (Frohman and Martin 1990).

2.2.5 Southern Analysis of *Ustilago* Transformants

Southern analysis (Southern 1975) on selected *U. maydis* transformants was carried out to verify single copy homologous integration of transforming DNA. The genomic DNA used for

Southern analysis was prepared as described in Section 2.3.4. This genomic DNA was digested by *Bam*HI (Gibco BRL) according to the manufacturers instructions and approximately 1-2 ug of DNA was electrophoresed on a BRL Model H4 apparatus. Rather than blotting the genomic DNA to a membrane, an in-gel hybridization procedure was used, whereby the agarose gel was dried, rehydrated, and then incubated with ^{32}P labeled probe (Ehtesham and Hasnain 1991). The probe used for Southern analysis was a 400 bp *Hind*III fragment from pUC18*b1* (Figure 2-1) spanning the 5' ends of both *bE1* and *bW1*. It was labeled with ^{32}P using the Pharmacia Biotech Oligolabeling Kit with random hexamer primers. The in-gel hybridization procedure saved a blotting step, but some difficulties with excessive background were experienced. Transformants with single copy homologous integration displayed a single hybridizing band at about 10.5 kb. The wild-type strains displayed a single band at about 8.5 kb.

2.3 *Ustilago* Techniques

2.3.1 Transformation of *U. maydis*

Transformation of *U. maydis* was accomplished by a protoplast-PEG-$CaCl_2$ procedure modified from Wang (1988) and Specht (1988). Protoplast were prepared with cells suspended in Buffer I (1 M sorbitol, 50 mM sodium citrate, pH 5.8) by the addition of Sigma Lysing Enzyme (Sigma L-2265). Protoplasts were stored frozen at -70°C in Buffer II (1 M sorbitol, 25 mM Tris pH 7.5, 50 mM $CaCl_2$) with PEG 3350 (Sigma P-3640), dimethyl sulfoxide (Sigma D-8418), and β-mercaptoethanol (Bio-Rad Cat. No. 161-0710). Addition of DNA to the protoplasts was done in 1.5 ml microfuge tubes. In brief, approximately 1 to 5 ug of linearized DNA was incubated with the protoplasts in the presence of PEG and $CaCl_2$. The protoplasts were washed twice with Buffer II and plated on Double Complete Media containing 1 M Sorbitol, 1% glucose, and 300 ug/ml Hygromycin B (Calbiochem, La Jolla, CA), and then incubated at 30°C for about 4 days. Each transformation tube gave about 50 to 100 stable transformants from 1 to 5 ug of DNA. A detailed description of the transformation protocol is given in Appendix 6.2.

2.3.2 Mating Tests

Routine mating tests were done by the following "drop-on-drop" procedure. A small lump of *U. maydis* cells was inoculated from a PDA plate into a 18 x 150 mm test tube containing 5 ml of PDB. It was important that the culture on the PDA plate was less than 2 weeks old (stored at 4°C) in order in achieve consistent inoculation. This culture was grown at 30°C, 225 rpm for approximately 16 to 20 hours until late log or early stationary phase (OD_{600} of 1.8 to 2.2). Then, about 10 to 30 ul of this culture was dropped on a charcoal mating plate containing 1% glucose and allowed to dry in a laminar flow hood for a few minutes. Then a drop of the tester culture grown in the same manner was placed on top of the first drop and allowed to dry. The plates were then taped with a double layer of Parafilm (American National Can, Greenwich, CT, USA) and then placed in the dark at room temperature for 24 to 48 hours, after which time the colonies were scored for their mycelial phenotype.

It was important not to allow the liquid cultures to incubate too long at stationary phase or to not store the cultures for more than a couple hours before spotting the cultures. Experience revealed that the cells had to be in an actively growing state for successful mating reactions. It was also found that the mycelial reactions were stronger if the first culture was allowed to dry before spotting the second, and if the mating plates were incubated at room temperature rather than 30°C. The use of room temperature for incubating mating reactions is consistent with previously reported methods (Day and Anagnostakis 1971; Puhalla 1968). A five point rating scale was used for scoring the mycelial phenotype (Table 2-3).

Table 2-3. Five point rating scale for mycelial phenotype of *U. maydis*

Rating	Description
-	Nonmycelial - no hairs visible projecting from fungal surface when observed under a dissecting microscope
+/-	Trace mycelial growth - sparse hairs or spikiness visible by dissecting scope only
+	Slightly mycelial - hairs clearly visible by dissecting scope, but white mycelium not clearly evident by naked eye
++	Moderately mycelial - white mycelial surface clearly visible by the naked eye but white color is not intense
+++	Strongly mycelial - intense white mycelial growth

2.3.3 Replica Mating Tests

Single transformants were transferred by toothpick onto master PDA 50 ug/ml hygromycin B (CalBiochem Corporation, San Diego, CA) plates in a 8 by 6 grid pattern that matched one half of a standard 96 microwell plate. The master plates were then put in plastic bags and incubated at 30°C overnight. These transformants were then transferred from the PDA master plates to charcoal mating plates using a prong type replica transferring device. The prong replica transfer device had 48 stainless steel prongs 1.5 mm in diameter and spaced 9 mm apart in a 8 by 6 pattern that matched a 96 microwell plate. The charcoal mating plates were incubated at 30°C overnight to allow colonies from 1-3 mm in size to form. Then a drop of late log phase to early stationary phase liquid PDB culture (OD_{600} of 1.8 to 2.2) was dropped on the colonies and the plates were allowed to dry in a laminar flow hood. The plates were then taped with Parafilm (American National Can, Greenwich, CT, USA) and incubated at room temperature for 24 to 48 hours. The colonies were then scored for their mycelial phenotype.

Selected transformants were streaked for single colonies from the PDA master plates onto new PDA hygromycin (50 ug/ml) plates. Single colony isolates were then inoculated into 18 x 150 mm test tubes containing 5 ml of PDB. These were grown without hygromycin selection for about 20 hours at 30°C and 225 rpm until early stationary phase. Integrative transformants have been previously shown to be mitotically stable (Kronstad et al. 1989; Wang et al. 1988) even without hygromycin selection. This liquid culture was used for retesting the mating type of selected transformants, for small scale preparation of genomic DNA, and for storage of the strains at -70°C.

2.3.4 Small Scale Preparations of *U. maydis* Genomic DNA

Preparations of *U. maydis* genomic DNA were done on a small scale in 1.5 ml microfuge tubes, using a glass bead and phenol-chloroform procedure adaped from *Saccharomyces* (Elder et al. 1993).

2.3.5 Storage of *U. maydis* Strains

U. maydis strains were stored for up to two months on PDA or PDA with 50 ug/ml hygromycin petri plates sealed with Parafilm. The *U. maydis* cultures survive longer on PDA than on rich media such as CM. For long term storage of up to several years, the strains were inoculated into 18 x 150 mm test tubes containing 5 ml of PDB and incubated at 30°C and 225 RPM for 18 to 24 hours until the cultures reached stationary phase (OD_{600} of 2.0 to 2.4). Then 800 ul of the culture was pipetted into a 1.5 ml microfuge tube, 70 ul of dimethyl sulfoxide (Sigma D-8418) was added, and this was vortexed briefly. The tube was placed in a -70°C freezer without any special prefreezing treatment. It was important that the cultures reached stationary phase because log phase cultures did not survive -70°C storage as well as stationary phase cultures.

3. RESULTS

3.1 Construction of Chimeric Alleles

As discussed in the Introduction, the N-terminal variable domains of *bE* and *bW* are responsible for allelic specificity. For *bE*, the variable domain occurs from amino acid positions 1 to 110, with the protein sequences of *bE1* and *bE2* being 66% identical, whereas the remainder of the *bE* open reading frame is 93% identical. Similarly, the *bW* variable domain from amino acid positions 1 to 150 is 54% identical between *bW1* and *bW2*, while the remainder of the open reading frame is 96% identical. A comparison of amino acid sequences between *bE1* and *bE2* was given in Figure 1-4 and between *bW1* and *bW2* in Figure 1-5.

Given this structure, the obvious experiment was to construct a chimeric allele containing the N-terminal variable region from one allele with the remaining C-terminal portion of another allele, and then test whether this altered specificity. An added dimension to this experiment was to construct a series of alleles with recombination points that scanned the variable region, so as to map the regions involved in allelic specificity. The following sections describe the results of these experiments for both *bE* and *bW* genes.

3.1.1 Construction of Chimeric *bE* Alleles and Fine Mapping of the *bE1-bE2* Specificity Classes

The earlier work of Yee and Kronstad (1993) demonstrated the utility of the chimeric allele approach for identifying regions responsible for allelic specificity. This work involved the *in vivo* construction of chimeric *bE1-bE2* alleles by transformation of a *bE2* strain of *U. maydis* with differing portions of the *bE1* variable region. Homologous integration of the transforming DNA produced *bE1-bE2* chimeric alleles that contained recombination points that spanned the *bE* variable region. Recombination points for *bE1-bE2* alleles were arbitrarily defined as the first *bE2* codon seen when scanning from the N to the C-terminus of the variable domain. The chimeric alleles are designated by the position of the recombination point, that is, a chimeric allele with its recombination point at codon 28 was thus designated *bEx28*.

Subsequent work, presented in this thesis, produced 5 additional chimeric alleles, allowing a fine structure map of the *bE1-bE2* specificity classes. The additional alleles *bEx57*, *bEx82*, and *bEx89* were constructed using the previously reported *in vivo* strategy, and the alleles *bEx31* and *bEx45* were constructed using an *in vitro* "Megaprimer" PCR strategy (Sarkar and Sommer 1990, see Materials and Methods). The *in vitro* PCR strategy allowed the construction of alleles with specific recombination points, without having to sequence a large number of transformants for the desired chimeric allele. Chosen alleles were spaced such that they differed by a single amino acid and they were positioned across a region where a change in specificity class occurred (see below). Thus, they allowed the study of single amino acid substitutions that altered specificity.

As reported in Yee and Kronstad (1993), the specificity of *bE1-bE2* chimeric alleles fell into three classes. If the parent *bE2* strain incorporated too little of the *bE1* variable region to alter its specificity, then it remained as a *bE2* allele. These were designated as class I. The opposite occurred when the transformant incorporated a large amount of *bE1* sequence such that the allele switched its specificity from *bE2* to *bE1*. These were designated as class III. The final possibility, which was less obvious at the onset of these experiments, was when a transformant incorporated an intermediate amount of *bE1* sequences such that its specificity was different from *bE1* or *bE2*. These were designated as class II. These three class of transformants are illustrated in Figure 3-1.

The *b* locus DNA sequences were amplified from individual chimeric allele strains by PCR and sequenced to determine the chimera position. A list of the chimeric *bE* strains is given in Table 3-1 and the fine map of the chimeric *bE* alleles, revised from Yee and Kronstad (1993) is given Figure 3-2. All chimeric allele strains reported were verified for single copy homologous integration by Southern analysis (see Materials and Methods and Figure 3-6). A total of 21 chimeric alleles are reported here for *bE*. Within the first 92 codons of *bE*, alleles for 16 of the potential 27 *bE1-bE2* chimera points (59%) were obtained.

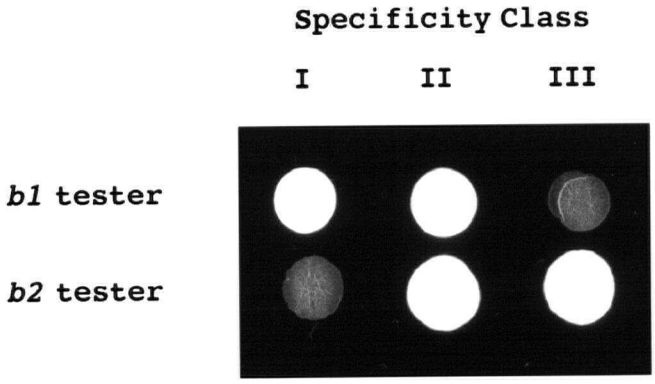

Figure 3-1. Three specificity classes of transformants with chimeric alleles. Mating reactions were carried out by co-spotting two *U. maydis* strains on DCM charcoal media, sealing with Parafilm, and incubating at room temperature for 24 to 48 hours. The white mycelia phenotype indicates a compatible mating reaction. Class I is unchanged from the parent *b2* specificity (compatible with *b1* but not *b2*). Class II is altered to a specificity different from *b1* or *b2* (compatible with *b1* and *b2*). Class III is switched to *b1* specificity (compatible with *b2* but not *b1*).

Table 3-1. List of chimeric *bE* strains.
The sequence for strain t17-3 was not available.

Strain	Allele	Mating 521 (*b1*)	Mating 031 (*b2*)	Alone
t11-1	*bEx152*	-	+++	-
ta-1	*bEx112*	-	+++	-
ta-3	*bEx142*	-	+++	-
ta-23	*bEx107*		+++	-
ta-28	*bEx128*	-	+++	-
tb-5	*bEx79*	+++	+++	-
tb-7	*bEx87*	-	+++	-
tb-11	*bEx48*	+++	+++	-
tb-14	*bEx90*	-	+++	-
tb-21	*bEx92*	-	+++	-
tb-23	*bEx70*	+++	+++	-
tc-1	*bEx49*	+++	+++	-
tc-3	*bEx51*	+++	+++	-
tc-7	*bEx60*	+++	+++	-
tc-9	*bEx49*	++	+++	-
td-4	*bEx39*	+++	+++	-
td-23	*bEx39*	+++	+++	-
td-24	*bEx28*	+++	-	-
td-43	*bEx39*	+++	+++	-
td-53	*bEx3*	+++	-	-
td-54	*bEx28*	++	-	-
t4-609	*bEx70*	+++	+++	-
t4-849	*bEx51*	+++	+++	-
t17-1	*bEx79*	+++	+++	-
t17-2	*bEx90*	-	+++	-
t17-3	*na*	+++	-	-
t17-4	*bEx87*	-	+++	-
t17-5	*bEx70*	+++	+++	-
t17-6	*bEx57*	+++	+++	-
t17-7	*bEx89*	-	+++	-
t17-8	*bEx82*	-	+++	-
t17-9	*bEx82*	-	+++	-
tBEX31-1	*bEx31*	+++	-	-
tBEX31-2	*bEx31*	+++	-	-
tBEX45-1	*bEx45*	+++	+++	-
tBEX45-2	*bEx45*	+++	+++	-

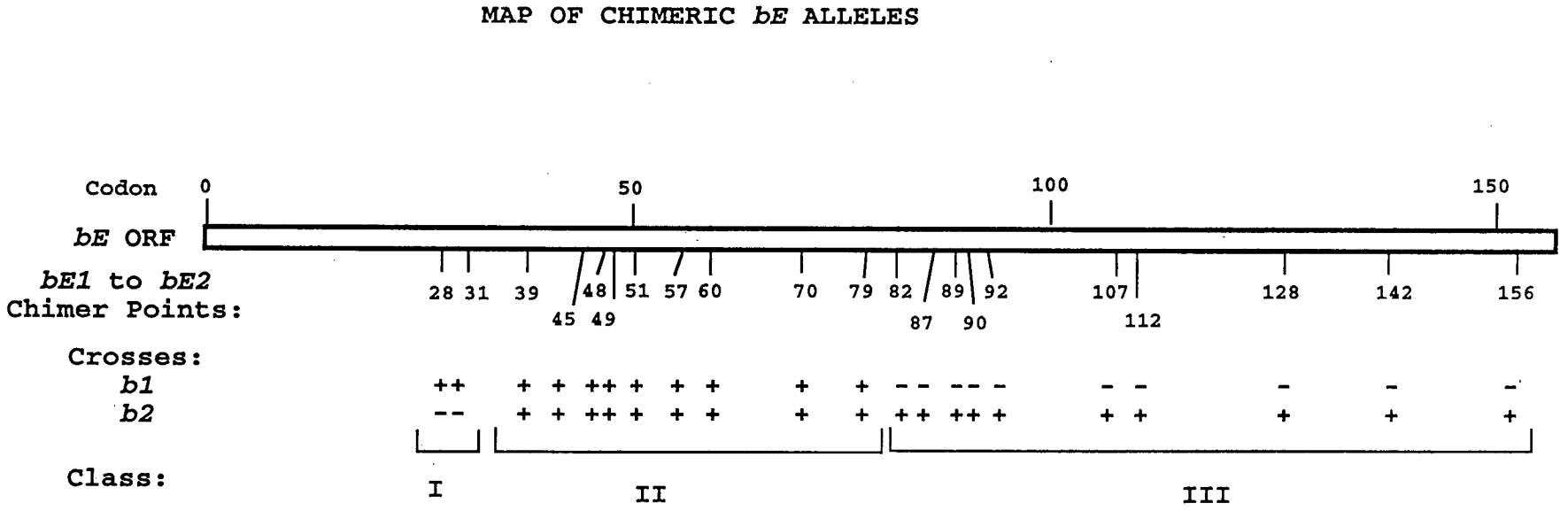

Figure 3-2. Map of chimeric *bE* alleles.
This figure represents a more detailed version of the map published in Yee and Kronstad (1993). The alleles *bEx31*, *bEx45*, *bEx57*, *bEx82*, and *bEx89* are added alleles. The transformant with a chimeric allele at position 39 was originally reported as class I, but subsequent isolation of more transformants with the same chimera point indicated that the original transformant was scored incorrectly.

The map of chimeric *bE* alleles (Figure 3-2) showed that the unaltered class I alleles contained chimera points up to codon 31. This indicates that replacement of *bE2* sequences with *bE1* sequences from codons 1 to 30 did not alter specificity.

The class II alleles possessed chimera points from amino acids positions 39 to 79. These alleles have a specificity different from either parent allele, as indicated by their dual compatiblity with both *b1* and *b2* testers. It is interesting how a chimera of two alleles produces a third allele, with a specificity distinct from the parents. This may point to a mechanism for the development of a multiallelic recognition system from a biallelic system, as discussed in Section 4.4.

The boundary between class I and class II alleles is delineated by *bEx31* and *bEx39*. The *bEx39* allele contains *bE1* sequences up to position 38, and then *bE2* sequences from codon 39 and beyond. It is apparent from these data that the *bE1* sequences in the region from codon 31 to codon 38 are responsible for a change in allelic specificity. It should be noted that the *bEx39* allele was originally reported as a class I allele (Yee and Kronstad 1993), but subsequent retesting and isolation of additional strains having the same chimeric allele indicated that the *bEx39* allele was a class II allele and that the original transformant was misclassified

The class III alleles (allele with *bE1* specificity) had chimera points from codon 82 and beyond. This indicates that incorporation of amino acids 1 to 81 from *bE1* was sufficient to impart *bE1* specificity to the parent *bE2* allele. More importantly, the amino acids from position 79 to 81 were critical to this change in specificity. As discussed in the next section, these boundary regions between different classes of chimeric alleles provided valuable insight into determinants of allelic specificity.

3.1.2 The Borders Between *bE* Specificity Classes Reveal Individual Amino Acids that Affect Specificity

Fine mapping of the specificity classes allowed the identification of single amino acid positions that affect allelic specificity. This can be illustrated by the boundary between specificity classes I and II, delineated by the two chimeric alleles *bEx31* and *bEx39* (Figure 3-3a). These two alleles had different specificities; *bEx31* behaved like an unaltered *bE2* allele in that it was compatible only with a *bE1* tester but not a *bE2* tester. In contrast, *bEx39* had dual compatibility; i.e., it was compatible with *bE1* and *bE2*. These two alleles differed by only one amino acid at position 31. This result indicates that a Tyr (Y) for Arg (R) substitution in *bEx31* changes a incompatible cross with *bE2* to a compatible one. This substitution to a Tyr (Y) results in a loss of a positively charged side chain and a gain in hydrophobicity. It is difficult to say with this limited data whether electrostatic charge, hydrophobicity, or even hydrogen bonding or side chain shape are the critical factors in this change in specificity. Nevertheless, it is significant that a change in specificity can result from mutation of a single amino acid.

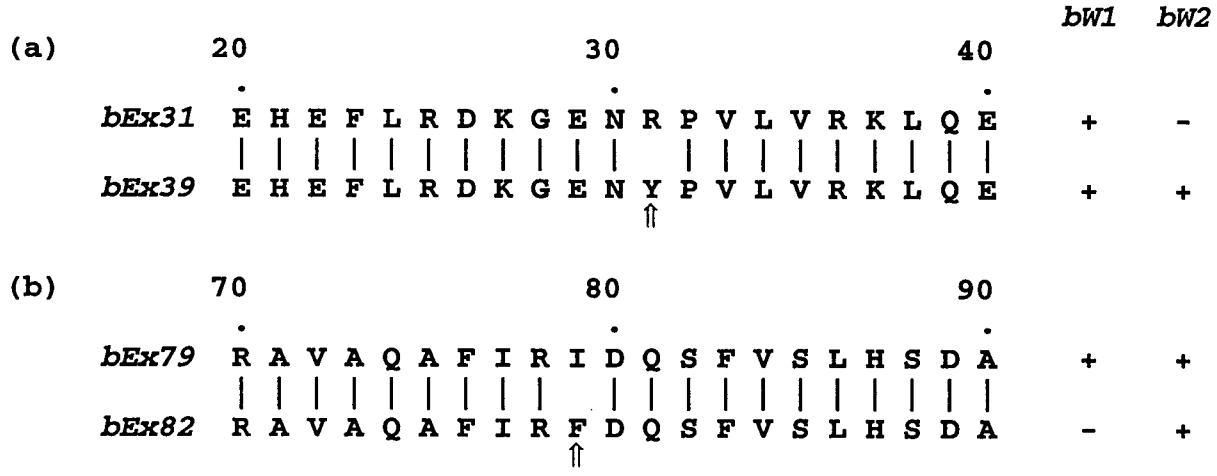

Figure 3-3. Comparison of sequences of *bEx31*, *bEx39*, *bEx79*, and *bEx82*.

The boundary between specificity classes II and III illustrated a similar situation whereby a single amino acid substitution resulted in altered specificity. The chimeric allele *bEx82* differs from *bEx79* (Figure 3-3b) by having a Ile (I) to Phe (F) substitution and this resulted in change from compatiblity to incompatiblity with *bW1*. These two amino acid residues are hydrophobic and their hydrophobicities are comparable (Table 1-1). Neither side chain is charged but they do differ in size and shape. The Ile side chain has a slightly smaller van der Waals volume (124 $Å^3$) than the Phe side chain (135 $Å^3$) and Ile is β-branched while Phe is γ-branched with an aromatic ring. This difference in side chain shape and the presence of the aromatic ring may result in steric hindrance, preventing dimerization of proteins derived from *bEx87* with *bW1*.

3.1.3 Construction of Chimeric *bW* Alleles and Fine Mapping of the *bW1-bW2* Specificity Classes

The similarity in domain structure between bE and bW proteins, namely a variable N-terminal domain, prompted an analogous chimeric allele analysis of bW. A series of chimeric alleles between *bW1* and *bW2* were constructed and the specificity of these alleles was determined by mating tests. As with *bE*, the transformants fell into the three previously described specificity classes (Figure 3-1). Class I alleles retained their original *bW2* specificity; class II alleles had an altered specificity different from the parental *bW1* or *bW2* alleles; and class III transformants had switched their specificity completely from *bW2* to *bW1*.

Sequencing of chimeric alleles from all three classes revealed the location of the chimera points and allowed the construction of a map of the specificity classes (Figure 3-4). A list of *bW* chimeric allele strains is given in Table 3-2. Again, special effort was taken to obtain closely spaced chimeric alleles at the boundaries between specificity classes. Chimeric alleles were obtained for 24 of the 42 potential *bW1-bW2* chimera positions (57%) in the first 109 codons of *bW*. The *bW1-bW2* chimera point was arbitrarily defined as the first codon that was unique to *bW2* and the allele was designated based on this codon position. Thus, an allele which had *bW1* sequence up to codon 5 and *bW2* sequence starting at codon 6 was designated *bWx6*.

Table 3-2. List of chimeric *bW* strains.

Strain	Allele	Mating 032 (*b1*)	Mating 518 (*b2*)	Alone
t51-1	*bWx171*	-	+++	-
t52-1	*bWx171*	-	+++	-
t53-1	*bWx48*	+++	+++	-
t57-3	*bWx48*	+++	+++	-
t57-4	*bWx107*	-	+++	-
t58-1	*bWx107*	-	+++	-
t58-2	*bWx52*	+++	+++	-
t58-3	*bWx36*			-
t60-2	*bWx48*	+++	+++	-
t60-3	*bWx48*	+++	+++	-
t60-4	*bWx36*	+++	+++	-
t60-5	*bWx36*	+++	+++	-
t61-4	*bWx19*	+++	+++	-
t61-5	*bWx48*	+++	+++	-
t62-5	*bWx49*	+++	+++	-
t62-7	*bWx38*	+++	+++	-
t64-1	*bWx48*	+++	+++	-
t66-1	*bWx19*	+++	+++	-
t66-2	*bWx9*	+++	+++	-
t66-3	*bWx12*	+++	+++	-
t66-4	*bWx36*	+++	+++	-
t66-5	*bWx2*	+++	-	-
tBWX4-1	*bWx4*	+++	-	-
tBWX4-2	*bWx4*	+++	-	-
tBWX6-1	*bWx6*	+++	-	-
tBWX6-2	*bWx6*	+++	-	-
tBWX68-1	*bWx68*	+++	+++	-
tBWX68-1	*bWx68*	+++	+++	-
tBWX72-1	*bWx72*	+++	+++	-
tBWX72-2	*bWx72*	+++	+++	-
tBWX73-1	*bWx73*	+++	+++	-
tBWX73-2	*bWx73*	+++	+++	-
tBWX74-1	*bWx74*	+++	+++	-
tBWX74-3	*bWx74*	+++	+++	-
tBWX76-1	*bWx76*	-	+++	-
tBWX76-3	*bWx76*	-	+++	-
tBWX77-1	*bWx77*	-	+++	-
tBWX77-3	*bWx77*	-	+++	-
tBWX79-1	*bWx79*	++	+++	-
tBWX79-2	*bWx79*	++	+++	-
tBWX80-1	*bWx80*	-	+++	-
tBWX80-2	*bWx80*	-	+++	-
tBWX81-1	*bWx81*	-	+++	-
tBWX81-2	*bWx81*	-	+++	-
tBWX81-3	*bWx81*	-	+++	-
tBWX82-1	*bWx82*	++	+++	-
tBWX82-3	*bWx82*	++	+++	-
tBWX83-1	*bWx83*	-	+++	-
tBWX83-2	*bWx83*	-	+++	-
tBWX88-1	*bWx88*	-	+++	-
tBWX88-2	*bWx88*	-	+++	-
tBWX91-1	*bWx91*	-	+++	-
tBWX91-2	*bWx91*	-	+++	-
tBWX140-1	*bWx140*	-	+++	-
tBWX140-2	*bWx140*	-	+++	-

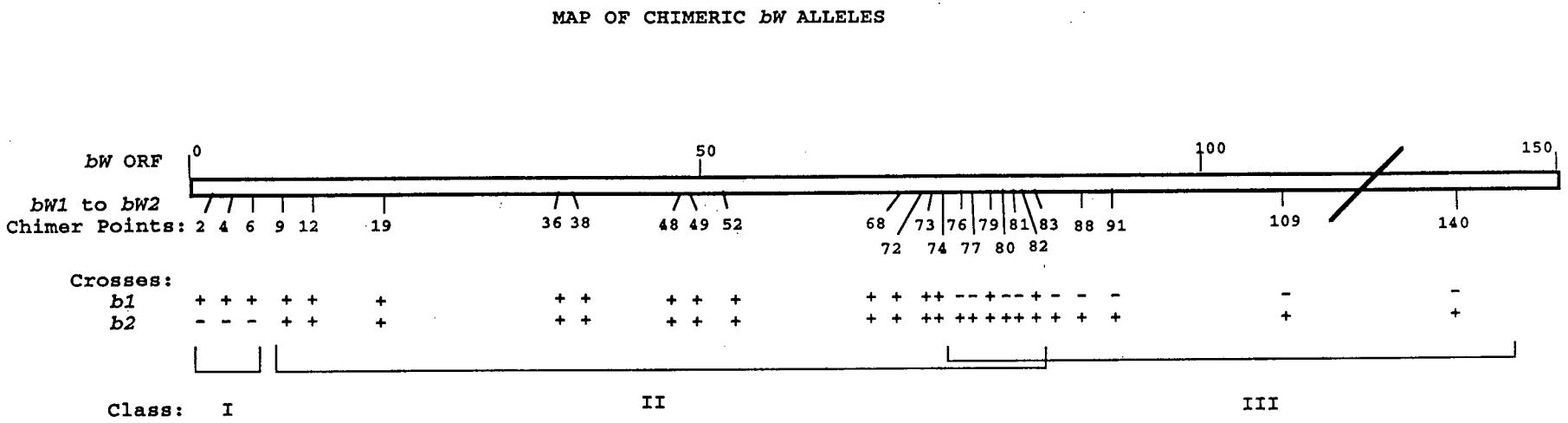

Figure 3-4. Map of chimeric *bW* alleles.

Two different strategies were used to construct these alleles. The first was the *in vivo* strategy, analogous to the one used for construction of most of the chimeric *bE* alleles. The *in vivo* strategy was successful in producing chimeric *bW* alleles, from codons 9 to 52, and within the constant region at codon 171. Unfortunately, this strategy met with very limited success in producing chimeric alleles within the region from codons 1 to 6, and 52 to 150. Even after screening approximately 5000 transformants, the only chimeric alleles obtained in these areas by the *in vivo* strategy were *bWx2* and *bWx109*. An inspection of *bW* sequences revealed a low level of homology between *bW1* and *bW2* in these problem areas (Figure 1-5). At the amino acid level, the region from codons 1-6 had 50% identity, the region from codons 52 to 91 had only 30% identity, and the region from codons 109 to 150 had only 34% identity. These levels of homology are lower than the overall level of 54% identity in the variable region of *bW* (codons 1-150). It is likely that this lack of homology prevented the homologous integration of transforming DNA into these areas. In support of this, the two chimeric alleles produced in the problem areas by the *in vivo* strategy lie in short stretches which have high homology between *bW1* and *bW2*.

The alternative *in vitro* strategy using "megaprimer" PCR (Sarkar and Sommer 1990, see Materials and Methods) was used to construct chimeric *bW* alleles in the region between codons 1 and 6 and between codons 52 and 150. The 16 alleles *bWx4*, *bWx6*, *bWx68*, *bWx72*, *bWx73*, *bWx74*, *bWx76*, *bWx77*, *bWx79*, *bWx80*, *bWx81*, *bWx82*, *bWx83*, *bWx88*, *bWx91*, and *bWx140* were constructed *in vitro* and introduced into a *bW2* strain of *U. maydis* by transformation and homologous integration at the *b* locus. This *in vitro* strategy was time consuming but effective at solving this technical problem, allowing the construction of closely spaced chimeric alleles in the areas of low homology between *bW1* and *bW2* . These alleles proved to be essential for the fine mapping of *bW* specificity classes, as well as providing key data in the experiments involving crosses between chimeric alleles (Section 3.2).

The three specificity classes of *bW* chimeric alleles occupied the following regions: class I had chimera points from codons 2 to 6, class II from 9 to 82, and class III from 76 to 140 (Figure 3-4). Two aspects of this specificity map appear interesting, and are in contrast to the analogous

map of chimeric *bE* alleles. First, the class I alleles covered a small region at the N terminus, resulting in a class I to class II boundary being very close to the N-terminus. Second, there was an area occupied by a mix of class II and class III alleles between codons 74 and 83. This effect is visually quite apparent in Figure 3-5b, which presents plate mating reactions involving a selection of strains carrying chimeric alleles.

Alleles belonging to class I (unaltered *bW2* specificity) were represented by chimeric alleles *bWx2*, *bWx4*, and *bWx6*. A PCR screen for homologous recombination (Frohman and Martin 1990) was used to identify the transformants where integration was at the *b* locus but specificity was not altered. Only one class I transformant, with allele *bWx2*, was obtained using the *in vivo* recombination strategy. The *bWx2* allele was unique because it was found to result from a chimera point upstream of the first codon of the *bW* ORF. This transformant had a *bW1* nucleotide sequence up to nucleotide -20 and a *bW2* nucleotide sequence beyond this. Thus the chimera point was actually in the promoter region upstream of the *bW* ORF. Using the naming convention described above, it was designated *bWx2* because codon 2 is the first codon unique to *bW2* (the Met at position one is common to both *bW1* and *bW2*) but it should be noted that this allele has the same amino acid sequence as *bW2*.

Class II alleles, with specificities different from either *bW1* or *bW2*, were found to have chimera points from codons 9 to 82. These were represented by transformants with alleles *bWx9*, *bWx12*, *bWx19*, *bWx36*, *bWx38*, *bWx48*, *bWx49*, *bWx52*, *bWx68*, *bWx72*, *bWx73*, *bWx74*, *bWx79*, and *bWx82*.

The border between the class I and class II alleles was very close to the N-terminus of the *bW*, delineated by the alleles *bWx6* and *bWx9*. The *bWx9* allele consists of *bW1* sequence up to codon 8 and then *bW2* sequence beyond this. Since the specificity of *bWx9* is altered from the parent *bW2*, this indicates that the amino acid sequence within the first 8 amino acids of *bW* play a critical role in allelic specificity.

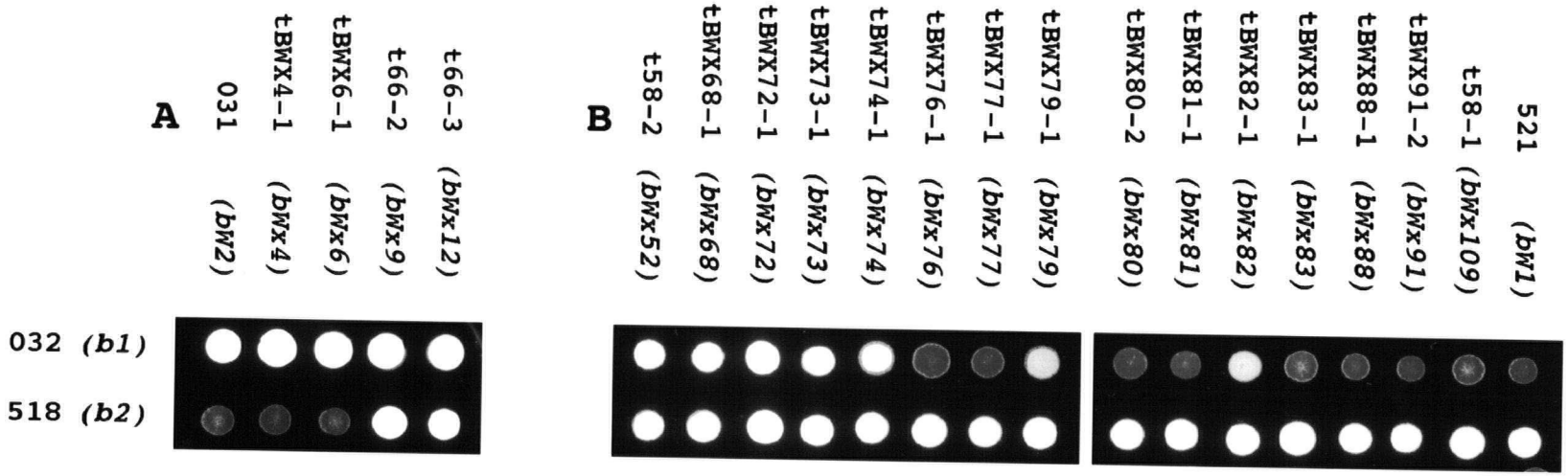

Figure 3-5. Plate mating reactions for *bW* chimeric allele strains at the specificity class borders. PDB liquid cultures of chimeric allele strains in late log to early stationary phase were co-spotted on DCM charcoal mating media with *b1* and *b2* testers. The plates were sealed with Parafilm and incubated at room temperature for 24 to 48 hours. (a) mating reactions involving alleles surrounding the class I to class II boundary; (b) mating reactions involving alleles surrounding the class II to class III boundary.

Class III transformants, which have switched completely to *bW2* specificity, had chimera points from codons 76 to 171. These were represented by the transformants with the alleles *bWx76*, *bWx77*, *bWx80*, *bWx81*, *bWx83*, *bWx88*, *bWx91*, *bWx109*, *bWx140*, and *bWx171*.

The border between class II and class III alleles was novel because it was not a discreet boundary; it was spread over an area between codons 74 and 83, and contained a mix of class II and class III alleles. These border regions are important because they identify areas where amino acid differences between bW1 and bW2 proteins affect allelic specificity. These individual amino acid substitutions will be discussed in Section 3.1.5, but existence of a broad specificity class border region may indicate a concentration of amino acids involved in allelic specificity.

All transformants carrying the chimeric alleles reported in Table 3-2 were verified for single copy homologous integration by Southern analysis (see Materials an Methods). A representative example of this is presented in Figure 3-6. The chimeric allele strains displayed a single band at 10.5 kb, and the wild type strain 032 displayed a single band at 8.5 kb. Note that one transformant strain tBWX91-1 displayed two bands, one at 10.5 kb and one at 9 kb. Even though this transformant displays two integration events (one homologous and one ectopic) its specificity was identical as the strain tBWX91-2 (which shows only a single homologous integration), when tested against wild type and chimeric allele tester strains (data not shown). It appears that in this situation, the ectopic integration had no effect on the *b* locus specificity. Approximately 5% of the chimeric alleles displayed more than one integration event (one homologous and one ectopic) and the majority of these displayed the same hybridization pattern, with two bands, one at 10.5 kb and one at 9 kb. This may indicate the presence in the genome, possibly close to the *b* locus, of nonfunctional *b* like sequences because these sequences seem to be preferentially targeted in ectopic integration events.

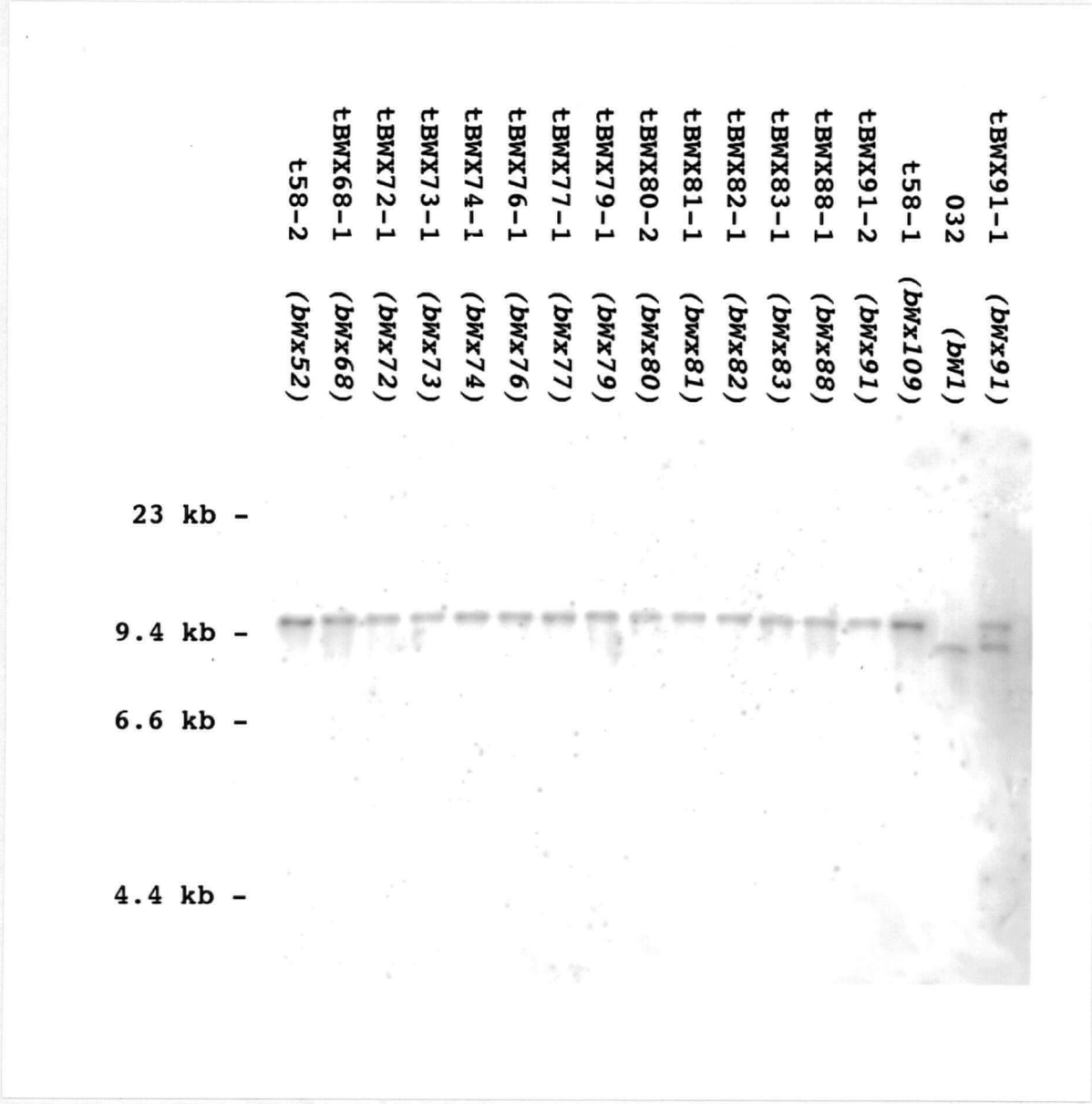

Figure 3-6. Southern analysis of some *bW* chimeric alleles strains.
Chimeric allele strains with single copy homologous integration displayed a single band at 10.5 kb and the wild type strain 032 displayed a single band at 8.5 kb. The strain tBWX91-1 displayed two bands, one at 10.5 kb and one at 9 kb, indicating two integration events, one homologous and one ectopic.

3.1.4 The Recombination Point for Transformation with a Given Fragment of *bW1* Occurs Close to the End.

Approximately 100 transformants from plasmid pAR66 were screened for homologous integration using PCR (Frohman and Martin 1990, see Materials and Methods) and the five chimeric alleles *bWx2*, *bWx9*, *bWx12*, *bWx19*, and *bWx36* were isolated from this screen. The bW sequence for plasmid pAR66 ends at codon 22. The recombination points were positioned 0, 3, 10, 13, and 20 codons away from the end of the plasmid pAR66 (Figure 3-7). Note that homologous integration at the very end of pAR66 at codon 22 resulted in an allele with a designation *bWx36*. This is due to the 100% identity between *bW1* and *bW2* between codons 20 and 36, making codon 36 the first position that is unique to *bW2* when a *bW1-bW2* chimera is created at codon 22. Thus, the convention used for naming chimeric alleles designates this allele as *bWx36*.

These data showed that homologous recombination tends to occur within about 60 basepairs from the end of the transforming DNA and confirm previous reports from *S. cerevisiae* and *U. maydis* that ends of linear DNA are much more recombinogenic than middle portions (Fotheringham and Holloman 1990; Orr-Weaver et al. 1981).

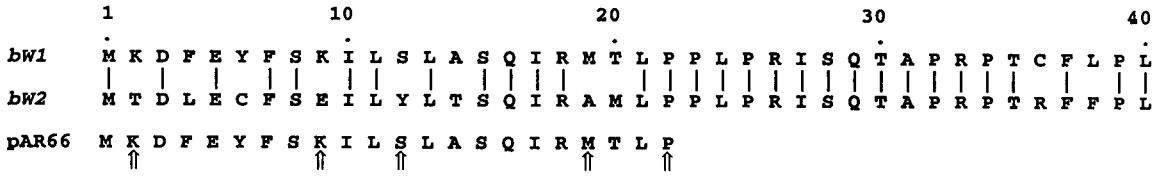

Figure 3-7. Recombination points for alleles resulting from transformation with pAR66. The recombination points for alleles *bWx2*, *bWx9*, *bWx12*, *bWx19*, and *bWx36* are marked by vertical arrows. . Note that homologous integration between the very end of pAR66 at codon 22 resulted in an allele with a designation *bWx36*. This is due to the homology between *bW1* and *bW2* from codons 20 to 36, making codon 36 the first position that is unique to *bW2* when a *bW1-bW2* chimera is created at codon 22.

3.1.5 Borders Between *bW* Specificity Classes Reveal Single Amino Acid Substitutions that Alter Specificity

The border between specificity classes I and II was defined by the chimeric alleles *bWx6* and *bWx9*. A comparison of these two alleles shows that *bWx9* allele differs from *bWx6* by only one amino acid; *bWx9* has a Cys (C) to Tyr (Y) substitution (Figure 3-8) which resulted in a change from an incompatible to a compatible mating with a *bE2* tester. The Tyr side chain differs from Cys in that it has a aromatic ring which is larger in size and strongly hydrophobic. Both Tyr and Cys are capable of hydrogen bonding but at different positions and orientations. The Tyr side chain has both polar and nonpolar groups and it is said to be amphiphilic (Creighton 1993, page 162). Therefore, it is not clear from this limited information whether van der Waals interactions, hydrophobic interactions, or hydrogen bonding are responsible for the difference in specificity between *bWx6* and *bWx9*.

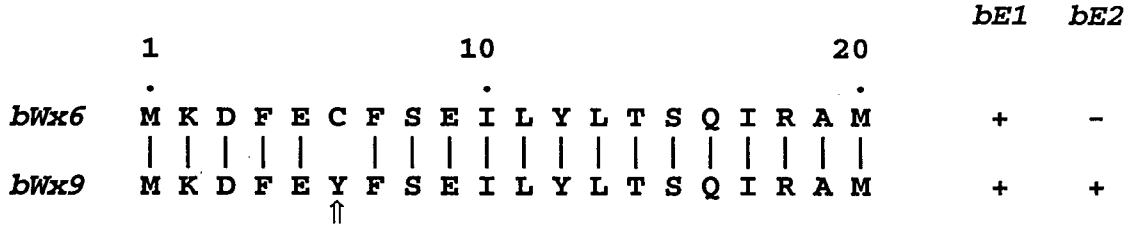

Figure 3-8. Comparison of sequences of *bWx6* and *bWx9*.
Corresponding mating reactions with *bE1* and *bE2* testers are given at right.

The boundary between specificity classes II and III was somewhat complex. The region from positions 74 to 83 contained an interesting series of chimeric alleles that switched back-and-forth from dual to single compatibility with the *bE1* and *bE2* testers (Figure 3-5b).

The two chimeric alleles *bWx74* and *bWx76* had different specificities; that is *bWx74* was compatible with both *bE1* and *bE2* while *bWx76* was only compatible with *bE2* (Figure 3-4). This switch from compatiblity to incompatiblity with *bE1* was associated with a His (H) to Tyr (Y) substitution (Figure 3-9). Although both side chains are ringed structures, the Tyr side chain has a six membered aromatic ring and a van der Waals volume of 141 $Å^3$ while His has a five membered

imidazole ring and a volume of 118 Å^3. The Tyr side chain is more hydrophobic than His so this was not likely the factor preventing compatiblity and thus preventing dimerization. It may be that the significantly larger size of the Tyr side chain at *bW* position 74 is interfering with dimerization of the *bWx76* allele with *bE1*. There are also other factors such as charge and positioning of hydrogen bonds that differ between Tyr and His.

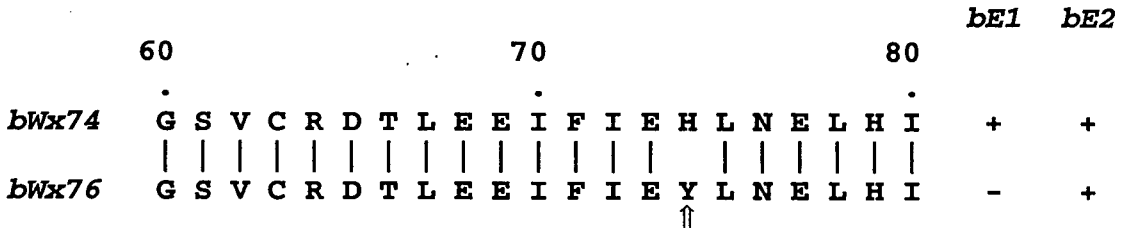

Figure 3-9. Comparison of sequences of *bWx74* and *bWx76*. Corresponding mating reactions with *bE1* and *bE2* testers are given at right.

The chimeric alleles *bWx77*, *bWx79*, *bWx80*, and *bWx81*, *bWx82*, *bWx83* all showed changes in specificity in relation to mating tests with *bE1* (Figure 3-4). A comparison of the amino acid sequences for these alleles is given in Figure 3-10.

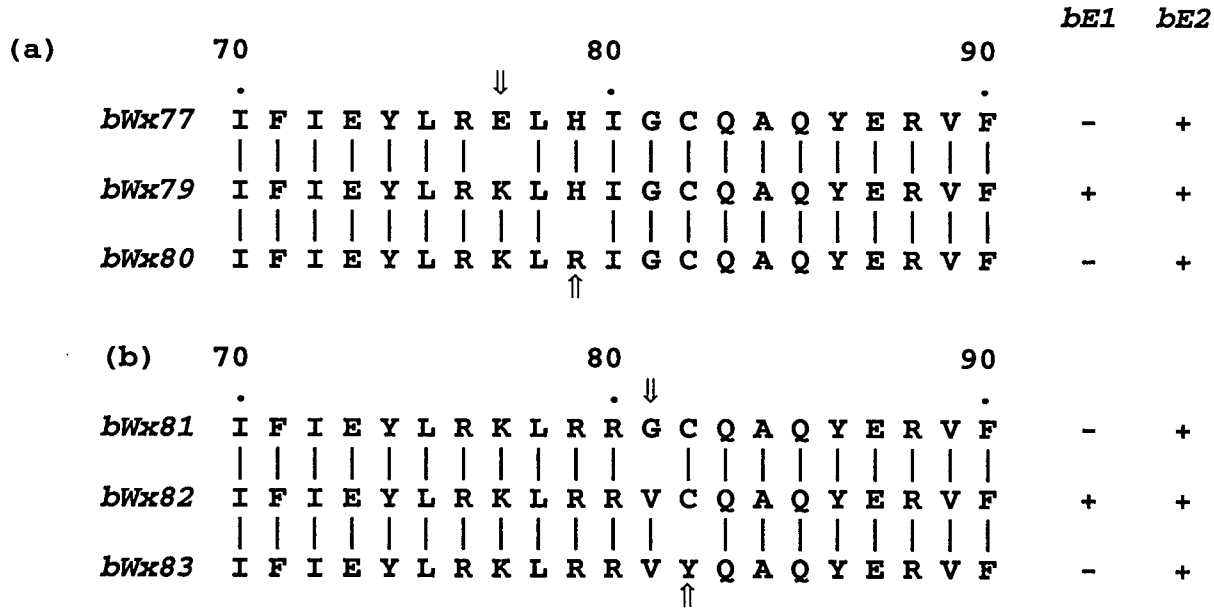

Figure 3-10. Comparison of sequences of (a) *bWx77*, *bWx79*, *bWx80*, and (b) *bWx81*, *bWx82*, *bWx83*.

These six alleles illustrate a variety of interactions that affect *bE-bW* dimerization. A substitution of a positively charged Lys (K) for negatively charged Glu (E) at position 77 in *bWx79* resulted in compatibility with *bE1*. This is a clear charge reversal and likely causes a switch from unfavorable to favorable electrostatic interactions between *bWx79* and *bE1*.

A His (H) to Arg (R) substitution at position 79 made *bWx80* incompatible with *bE1*. The van der Waals volume of Arg is much larger that His (148 $Å^3$ versus 118 $Å^3$) and Arg is much less hydrophobic than His so it could be either unfavorable steric interactions or unfavorable hydrophobic interactions, or both, that was preventing dimerization of *bWx80* and *bE1*.

A Gly (G) to Val (V) substitution at position 81 in *bWx82* versus *bWx81* gives a compatible mating reaction between *bWx82* and *bE1*. The Val side chain is much larger than Gly but it is also much more hydrophobic, so it is likely that this increase in hydrophobicity was favoring dimerization.

Comparison of *bWx82* and *bWx83* shows that a Cys (C) to Tyr (Y) substitution at position 82 resulted in an incompatible cross between strains carrying *bWx83* and *bE1*. The Tyr side chain is more hydrophobic than Cys so it would seem as if this substitution would favor dimerization as it did in *bWx6* versus *bWx9*. The difference here may be that the Tyr (Y) side chain is also much bulkier that Cys (van der Waals volumes of 141 $Å^3$ versus 86 $Å^3$) and that there is a yet unidentified large residue in bE1 that interacts with the Tyr at position 82 of *bWx83* to prevent dimerization. Thus, unfavorable van der Waal interactions may have been the factor in this situation.

3.1.6 Chimeric alleles of *bW1* and *bW3*

All of the chimeric allele construction described above has focused on the *bE* and *bW* genes of *b1* and *b2* specificity. We were also interested whether the specificity class structure differed with other allele combinations. A comparison of the alleles *bW1* and *bW3* showed a surprisingly small number of differences in the amino acid sequence between the two (Gillissen et al. 1992).

There are only nine differences between *bW1* and *bW3* in the first 150 codons (94% identity) and four of those are in the first 10 codons.

Three chimeric alleles were obtained between *bW1* and *bW3* by transforming a *bW3* strain of *U. maydis* ATCC 18604 *(a1b3)* with plasmids pAR53 and pAR66 using the previously described methods for *in vivo* construction of chimeric alleles. These chimeric alleles had recombination points at codons 9, 20, and 34. The specificities of these chimeric alleles were tested by crossing with the strains 032 *(a2b1)* and 023 *(a2b3)*. Surprisingly, all three chimeric alleles had class III specificity; that is they were incompatible with *b1* and compatible with *b3*. This indicated that there were sequences that caused incompatibility with *bE1* but allowed compatibility with *bE3*, contained within the first 8 codons of *bW1*. A comparison of the wild type *bW1*, *bW3*, and the *bW1/3x9* chimeric allele (a *bW1-bW3* chimeric allele with a recombination point at position 9) is given in Figure 3-11. There are three amino acid differences between *bW3* and *bW1/3x9*; Ala at position 2 is changed to Lys, Leu at position 4 is changed to Phe, and Ser at position 6 is changed to Tyr. One of these substitutions is to a positively charged side chain and the other two substitutions are to aromatic amino acids. It is difficult to interpret these results precisely because the changes in specificity cannot be attributed to single amino acid changes, but these data indicate that the amino acid positions that determine specificity are likely to change depending upon which allele pair is interacting.

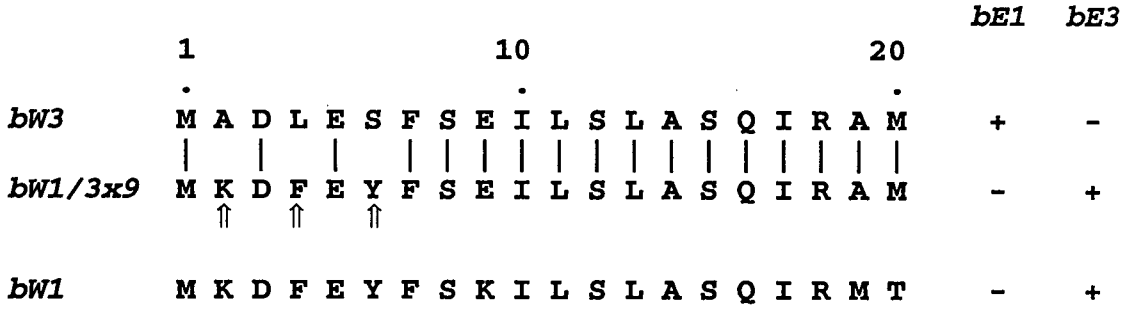

Figure 3-11. Comparison of sequences of *bW3*, *bWx9*, and *bW1*. Compatibility with *bE1* and *bE1* testers is given at the right.

3.1.7 Summary

Overall, these comparisons of chimeric alleles that differ in specificity at single amino acids provide insight into the types of amino acid changes and protein-protein interactions that affect *b* locus specificity.

This analysis has identified several amino acid positions in bE and bW that affect specificity. These were bE positions 31 and 79, and bW positions 6, 74, 77, 79, 81, and 82 (Table 3-3). The nature of the amino acid changes that affected specificity involved van der Waals, hydrophobic, electrostatic, and hydrogen bond interactions. On first inspection, the data seems to lack a consistent pattern with regards to interactions that affect specificity. It should be kept in mind however, that allelic specificity is likely determined by interactions that prevent dimerization between bE and bW from like alleles. If we focus on the key question presented in the Introduction (section 1.5) - "what prevents bE1 from dimerization with bW1?" - then the picture becomes clearer. The substitutions that brought about incompatibility with bE1 or bW1 are summarized in Table 3-3. Out of these eight substitutions, one of them consists of a charge reversal (Lys at bW position 77 to Glu) and three of them involve the introduction of an aromatic side chain (either Tyr or Phe). These aromatic side chains have a large van der Waals volume and a bulky shape. Could it be these bulky side chains that are interfering with dimerization between bE and bW?

Table 3-3. Summary of single amino acid changes in bE or bW that cause incompatibility. Chimeric alleles crossed with wild-type *b1* or *b2* alleles. Superscripts on amino acids denote their position in bE or bW, i.e. Tyr ^{E31} denotes Tyr on bE position 31.

Allele Pair	Substitution	Effect on compatibility
bEx39/31	Tyr^{E31} to Arg	*bEx31* incompatible with *bW2*
bEx79/82	Ile^{E79} to Phe	*bEx79* incompatible with *bW1*
bWx9/6	Tyr^{W6} to Cys	*bWx6* incompatible with *bE2*
bWx74/76	His^{W74} to Tyr	*bWx76* incompatible with *bE1*
bWx79/77	Lys^{W77} to Glu	*bWx77* incompatible with *bE1*
bWx79/80	His^{W79} to Arg	*bWx79* incompatible with *bE1*
bWx81/82	Val^{W81} to Gly	*bWx81* incompatible with *bE1*
bWx82/83	Cys^{W82} to Tyr	*bWx82* incompatible with *bE1*

3.2 Crosses Between Strains with Chimeric *b* Alleles

The previous section has shown how two analogous sets of chimeric alleles have provided some clues to the determinants of specificity at the *b* locus. These clues were obtained by crossing the chimeric *bE* or chimeric *bW* alleles with wild-type *b1* or *b2* partners.

Due to the genetic and biochemical evidence for interaction between *bE* and *bW* proteins, it was surmised that further clues to specificity could be obtained by pairwise crossing, in all combinations, of strains carrying chimeric *bE* alleles with the strains carrying chimeric *bW* alleles. It was envisaged that these crosses would produce a pattern of compatible and incompatible interactions that could identify additional determinants of allelic specificity. Again, one would look for differences in specificity between alleles that differed by only one amino acid.

3.2.1 Crosses Between Strains with Chimeric *bE* and *bW* Alleles

U. maydis strains carrying 19 different chimeric *bE* alleles and 2 wild-type *bE* alleles were crossed in all combinations with strains carrying 26 different chimeric *bW* alleles plus 2 wild type *bW* alleles, resulting in a total of 588 combinations. The *U. maydis* strains carrying *bE* and *bW* chimeric alleles were originally designed to be compatible at the *a* locus, so these crosses could be carried out directly, without any further manipulation of genetic backgrounds. These results are presented in Table 3-4. All crosses were repeated a minimum of two times to assure consistent data.

One of the most salient features of the data in Table 3-4 is the large block of incompatible mating combinations in the central area of the matrix. This block of data represents crosses between the class II chimeric *bE* alleles and the class II chimeric *bW* alleles. The class II chimeric alleles were those with a new specificity, different from the parental *b1* and *b2* alleles. The observation that most of the class II chimeric alleles from both bE and bW had the same specificity suggested that the region at the borders of the specificity classes are critical for allelic specificity.

Table 3-4. Crosses between strains carrying chimeric alleles of *bE* and *bW*.
Compatible matings are denoted by + and incompatible matings are denoted by -.

Strain		518	td-24	tBWX31-1	td-23	tBEX45-1	tb-11	tc-1	tc-3	t17-6	tc-7	tb-23	t17-1	t17-8	tb-7	t17-7	tb-14	tb-21	ta-23	ta-28	t11-1	032
	Allele	*bE2*	*bEx28*	*bEx31*	*bEx39*	*bEx45*	*bEx48*	*bEx49*	*bEx51*	*bEx57*	*bEx60*	*bEx70*	*bEx79*	*bEx82*	*bEx87*	*bEx89*	*bEx90*	*bEx92*	*bEx107*	*bEx128*	*bEx156*	*bE1*
031	*bW2*	-	-	-	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++
tBWX4-2	*bWx4*	-	-	-	-	-	+++	+++	+++	+++	+++	+++	+++	+++	++	+++	++	++	+++	+++	+++	+++
tBWX6-1	*bWx6*	-	-	-	-	-	+++	+++	+++	+++	+++	+++	+++	+++	++	+++	+++	+++	+++	+++	+++	+++
t66-2	*bWx9*	+++	+++	+++	-	-	+++	+++	+++	+++	+++	++	+++	+++	++	+++	++	++	+++	+++	+++	+++
t66-3	*bWx12*	+++	+++	+++	-	-	-	-	-	-	-	-	-	-	-	+++	++	++	+++	+++	+++	+++
t66-1	*bWx19*	+++	+++	+++	-	-	-	-	-	-	-	-	-	-	-	+++	++	++	+++	+++	+++	+++
t66-4	*bWx36*	+++	+++	+++	-	-	-	-	-	-	-	-	-	-	-	+++	++	++	+++	+++	+++	+++
t62-7	*bWx38*	+++	+++	+++	-	-	-	-	-	-	-	-	-	-	-	++	++	+++	+++	+++	+++	+++
t60-2	*bWx48*	+++	+++	+++	-	-	-	-	-	-	-	-	-	-	-	++	++	++	+++	+++	+++	+++
t62-5	*bWx49*	+++	+++	+++	-	-	-	-	-	-	-	-	-	-	-	++	++	++	+++	+++	+++	+++
t58-2	*bWx52*	+++	+++	+++	-	-	-	-	-	-	-	-	-	-	-	+	++	+++	+++	+++	+++	+++
tBWX68-1	*bWx68*	+++	+++	+++	-	-	-	-	-	-	-	-	-	-	-	-	-	+++	+++	+++	+++	+++
tBWX72-1	*bWx72*	+++	+++	+++	-	-	-	-	-	-	-	-	-	-	-	-	-	+++	+++	+++	+++	+++
tBWX73-1	*bWx73*	+++	+++	+++	-	-	-	-	-	-	-	-	-	-	-	-	-	+++	+++	+++	+++	+++
tBWX74-1	*bWx74*	+++	+++	+++	-	-	-	-	-	-	-	-	-	-	-	-	-	++	+++	+++	+++	+++
tBWX76-1	*bWx76*	+++	+++	+++	-	-	-	-	-	-	-	-	-	-	-	+	++	-	++	++	++	-
tBWX77-1	*bWx77*	+++	+++	+++	-	-	-	-	-	-	-	-	-	-	-	-	-	-	++	++	++	-
tBWX79-1	*bWx79*	+++	+++	+++	-	-	-	-	-	-	-	-	-	-	-	+++	+++	-	+++	+++	+++	++
tBWX80-2	*bWx80*	+++	+++	+++	-	-	-	-	-	-	-	-	-	-	-	+++	++	-	+++	+++	+++	-
tBWX81-1	*bWx81*	+++	+++	+++	-	-	-	-	-	-	-	-	-	-	-	-	++	-	-	-	-	-
tBWX82-1	*bWx82*	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	-	-	+++	++	++	+++	+++	+++	++
tBWX83-1	*bWx83*	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	-	-	-	-	-	-	-	-	-	-
tBWX88-1	*bWx88*	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	-	-	-	-	-	-	-	-	-	-
tBWX91-2	*bWx91*	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	-	-	-	-	-	-	-	-	-	-
t58-1	*bWx109*	+++	+++	+++	+++	++	+++	+++	+++	+++	+++	++	-	-	-	-	-	-	-	-	-	-
tBWX140-1	*bWx140*	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	-	-	-	-	-	-	-	-	-
t52-1	*bWx171*	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	-	-	-	-	-	-	-	-	-
521	*bW1*	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	-	-	-	-	-	-	-	-	-

The pattern of incompatible matings in the top left and lower right of Table 3-4 was consistent with the fact that these crosses were between alleles with the same specificity. Crosses between *bE* and *bW* alleles with *b2* specificity occurred in the top left, and crosses between alleles with *b1* specificity occurred in the bottom right. In a similar line of reasoning, the blocks of compatible matings in the top right and lower left of the table were consistent with crosses between alleles having different specificities.

3.2.2 Borders Between Compatible and Incompatible Crosses Reveal Individual Amino Acid Residues that Affect Specificity

Single amino acid changes that affect specificity can be identified by looking at specificity boundaries between compatible and incompatible crosses in Table 3-4. The specificity boundaries valuable for revealing important codon positions are those involving two chimeric alleles that differ from each other by a single amino acid. This can be illustrated by the following examples which reveal the types of amino acid changes that affect dimerization between bE and bW.

The alleles *bW2* and *bWx4* displayed different specificities when crossed against *bEx39* and *bEx45* (upper left of Table 3-4) and their sequences differ by only one amino acid (Figure 3-12). The chimeric allele *bWx4* differs from the wild type allele *bW2* by having a Thr (T) to Lys (K) substitution at position 2. The Lys (K) side chain introduces a positive charge at this position, and it has a large van der Waals volume (135 $Å^3$) compared to Thr (93 $Å^3$). This change in electrostatic interactions or in van der Waals interactions appears to have altered the dimerization of bE and bW and thus altered specificity. It is interesting to note that this difference in specificity was evident only when crossed with the chimeric alleles *bEx39* and *bEx45*, and not when tested with any other chimeric or wild type alleles. The explanation for this may be: (1) the *bEx39* and *bEx45* proteins are close to the dimerization "balance point" with the *bW2* protein, such that small changes in the interaction will tip this balance one way or the other, and (2) the region of bW position 2 is interacting in some way with the region around bE position 39. It is possible that

these positions are adjacent to each other on the dimer interface. Further data would be required to test these possibilities.

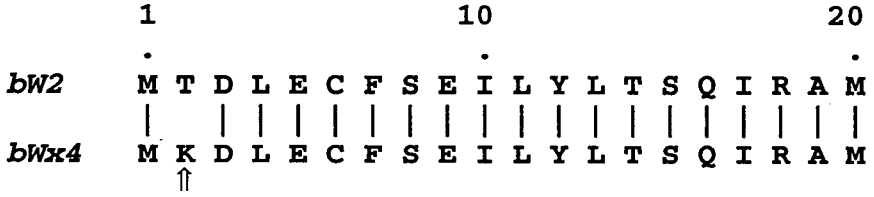

Figure 3-12. Comparison of sequences of *bW2* and *bWx4*.

The second example of two alleles that differ from each other by one amino acid involved *bWx6* and *bWx9* . The single difference between *bWx6* and *bWx9* is a Cys (C) to Tyr (Y) substitution at position 6 (Figure 3-13). From Table 3-5 (extracted from Table 3-4) it was evident that this single substitution resulted in altered specificity when the two alleles *bWx6* and *bWx9* were crossed with *bE2*, *bEx28*, and *bEx31*. This single substitution of Cys to Tyr resulted in a change from an incompatible to a compatible mating reaction. This is possibly due to an increase in hydrophobicity of Tyr or to the ability of Tyr to hydrogen bond, both of which would favor dimerization between *bE* and *bW*. This is the same substitution that was discussed in Section 3.1.5 on the mapping of the *bW* specificity classes.

Note that two chimeric alleles do not have to be designated with consecutive position numbers (i.e. *bWx6* and *bWx7*) in order to differ by one amino acid. This is because there are many positions that share the same sequence between *bW1* and *bW2*. Thus, *bW1* and *bW2* differ at positions 6 and 9 but not at positions 7 and 8.

Table 3-5. Mating reactions involving *bWx6* and *bWx9*.
Extracted from Table 3-4.

Allele	bE2	bEx28	bEx31
bWx6	-	-	-
bWx9	+++	+++	+++

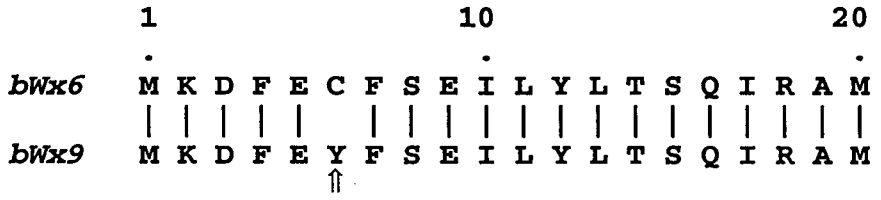

Figure 3-13. Comparison of sequences of *bWx6* and *bWx9*.
Two chimeric alleles that differ by one amino acid at position 6.

The third example of these "specificity boundaries" can be defined for *bWx9* and *bWx12*. This is shown in Table 3-6 (extracted from Table 3-4). The alleles *bWx9* and *bWx12* differed from each other when crossed with the *bE* chimeric alleles *bEx48* to *bEx87*; *bWx9* gave compatible mating reactions and *bWx12* gave incompatible reactions. These two chimeric alleles differ from each other by only one amino acid (Figure 3-14); *bWx9* possesses a Glu (negatively charged) at position 9 and *bWx12* possesses a Lys (positively charged) at this position. This shows that a charge reversal from negative to positive at position 9 of *bWx9* caused incompatibility with many *bE* chimeric alleles. It is possible that the positively charged residue resulted in electrostatic repulsion with an unidentified residue on bE and prevented dimer formation.

Table 3-6. Mating reactions involving *bWx9* and *bWx12* Extracted from Table 3-4.

Allele	*bEx48*	*bEx49*	*bEx51*	*bEx57*	*bEx60*	*bEx70*	*bEx79*	*bEx82*	*bEx87*
bWx9	+++	+++	+++	+++	+++	++	+++	+++	++
bWx12	-	-	-	-	-	-	-	-	-

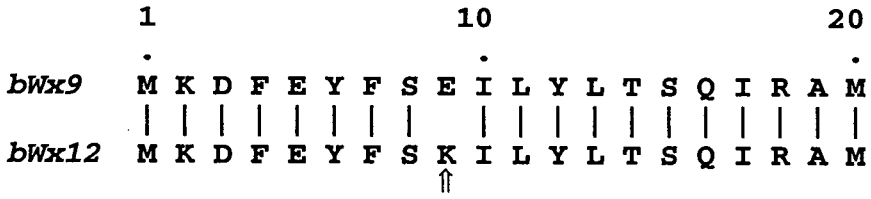

Figure 3-14. Comparison of sequences of *bWx9* and *bWx12*

A fourth boundary between compatible and incompatible crosses was found between *bWx76* and *bWx77*, when crossed with *bEx89* and *bEx90* (Table 3-7). These two alleles differed by only one amino acid, a replacement of Gln (N) at position 76 of *bWx76* with Arg (R) resulted in an incompatible reaction. Both Gln (N) and Arg (R) have polar side chains but Arg (R) is positively charged and its side chain is much larger (van der Waals volume of 148 $Å^3$ versus 96 $Å^3$). The effect of the Arg (R) side chain could be due to the introduction of a positive charge, possibly repelling another positive charge on *bE* at the dimer interface, or due to the large size of Arg (R) resulting in steric hindrance to dimerization between bE and bW.

Table 3-7. Mating reactions involving *bWx76*, *bWx77*, and *bWx79*. Extracted from Table 3-4.

Allele	bEx89	bEx90
bWx76	+	++
bWx77	-	-
bWx79	+++	+++

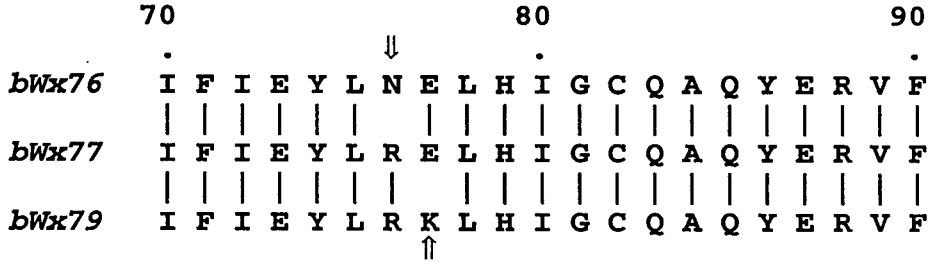

Figure 3-15. Comparison of sequences of *bWx76*, *bWx77*, and *bWx79*.

A fifth specificity boundary also involved *bWx77*; this time with the adjacent allele *bWx79*. The *bWx79* allele contains a Lys (K) at position 77, replacing the Glu (E) of *bWx77* (Figure 3-15). Lys is positively charged and Glu is negatively charged so this substitution is a clear charge reversal. It has changed the incompatible mating reactions with *bEx89* and *bEx90* to compatible ones (Table 3-7). It is possible that this charge reversal at position 77 of *bW* resulted in a switch from unfavorable to favorable electrostatic interactions with an unidentified positively charged residue on *bE*.

The sixth, seventh, eighth specificity boundaries involved *bWx80*, *bWx81*, *bWx82*, and *bWx83*, when crossed with *bEx89*, *bEx90*, *bEx107*, *bEx128*, *bEx156*, and they identified a series of three consecutive amino acid substitutions that altered the specificity of these chimeric *bW* alleles in a "back and forth" manner (Table 3-8). The change from *bWx80* to *bWx81* resulted in a loss of compatibility and is associated with a loss of hydrophobicity (substitution of an Ile (I) for Arg (R) at position 80; Figure 3-17). The second change from *bWx81* to *bWx82* reversed this effect by regaining compatibility with an increase in hydrophobicity (substitution of Val (V) for a Gly (G) at

position 82). The third change from *bWx82* to *bWx83* resulted in a loss of compatibility again, but this time it was associated with the introduction of a bulky aromatic side chain. (a Cys (C) to Tyr (Y) substitution at position 82, this substitution was previously discussed in Section 3.1.5). This series of three changes is illustrative of the effect of increased hydrophobicity on favoring dimerization, and on the effect of a bulky Tyr side chain disfavoring dimerization between bE and bW.

Table 3-8. Mating reactions involving *bWx80*, *bWx81*, *bWx82*, and *bWx83*. Extracted from Table 3-4.

Allele	bEx89	bEx90	bEx92	bEx107	bEx128	bEx156
bWx80	+++	++	-	+++	+++	+++
bWx81	-	++	-	-	-	-
bWx82	+++	++	++	+++	+++	+++
bWx83	-	-	-	-	-	-

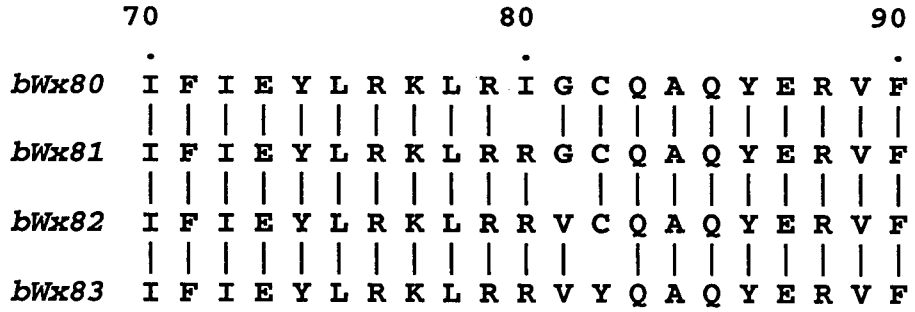

Figure 3-16. Comparison of sequences of *bWx80*, *bWx81*, *bWx82*, and *bWx83*.

A ninth specificity boundary occurred with strains carrying alleles *bEx45* and *bEx48*, when crossed with strains carrying alleles *bWx4*, *bWx6*, and *bWx9* (Table 3-8). The allele *bEx48* has become compatible with the three *bW* chimeric alleles and this was associated with a Thr (T) to Ile

(I) substitution at position 45 of *bW* (Figure 3-17). The Ile is much more hydrophobic than Thr and this is likely to have favored dimerization between bE and bW.

Table 3-9. Mating reactions involving *bEx45* and *bEx48*. Extracted from Table 3-4.

Allele	bEx45	bEx48
bWx4	-	+++
bWx6	-	+++
bWx9	-	+++

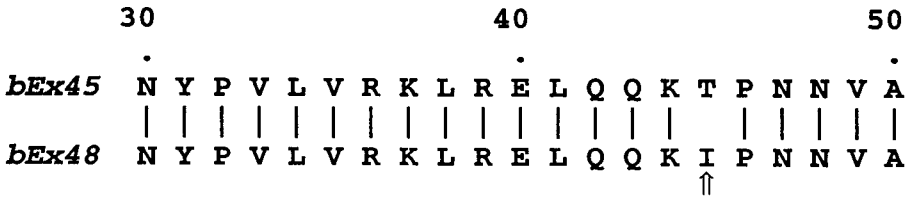

Figure 3-17. Comparison of sequences of *bEx45* and *bEx48*.

A tenth specificity boundary occurred with *bEx79* and *bEx87* and was associated with a loss of compatibility due to a Ile (I) to Phe (F) substitution. This substitution was discussed previously in Section 0 and its effect is likely due to unfavorable steric interactions involving the Phe side chain.

An eleventh boundary between compatible and incompatible mating combinations occurred with alleles *bEx87* and *bEx89*. This is presented in Table 3-10. The comparison of the two sequences is given in Figure 3-18. These data show that a His (H) to Cys (C) substitution changed an incompatible mating to a compatible one. The Cys side chain has about the same hydrophobicity as His but it is much smaller in size, therefore this substitution may favor dimerization of bE and bW by decreased steric hindrance.

Table 3-10. Mating reactions involving *bEx87* and *bEx89*. Extracted from Table 3-4.

Allele	bEx87	bEx89
bWx12	-	+++
bWx19	-	+++
bWx36	-	+++
bWx38	-	++
bWx48	-	++
bWx49	-	++
bWx52	-	+

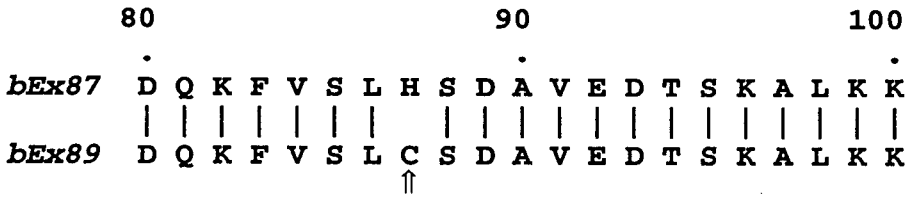

Figure 3-18. Comparison of sequences of *bEx87* and *bEx89*.

3.2.3 Some Boundaries Identify Blocks of Amino Acids that Affect Specificity

There are additional specificity boundaries between compatible and incompatible crosses, but they involve chimeric alleles that differ by more that one amino acid. For these, the data cannot be interpreted in such a way as to link a change in specificity to a single amino acid difference. Nevertheless, these additional specificity boundaries are still useful in identifying blocks of amino acids that affect specificity.

The crosses involving the alleles *bEx70* and *bEx79* are an illustrative example of this. The strains carrying these two alleles had different specificities when crossed with strains carrying *bWx83*, *bWx88*, *bWx91*, and *bWx109* (Table 3-11). A comparison of their sequences (Figure 3-19) shows that they differ at positions 70, 71, 72, and 74. Note that 2 of the 4 changes involve Val for Ala substitutions, resulting in a significant change in side chain size and hydrophobicity. It is not possible with existing data to attribute the altered specificity to a single amino acid difference.

What can be said is that this region in bE from positions 70 to 73 contains amino acids that affect specificity and this region is likely to be on the bE to bW dimer interface.

Table 3-11. Mating reactions involving *bEx70* and *bEx79*. Extracted from Table 3-4.

Allele	bEx70	bEx79
bWx83	+++	-
bWx88	+++	-
bWx91	+++	-
bWx109	++	-

```
          60                  70                  80
          .                   .                   .
bEx70     Q Q I H Q T T H R I K V A A K A F I R I D
          | | | | | | | | | |     |   | | | | | |
bEx79     Q Q I H Q T T H R I R A V A Q A F I R I D
```

Figure 3-19. Comparison of sequences of *bEx70* and *bEx79*.

A second example involves the alleles *bEx92* and *bEx107*. The strains carrying these two alleles behaved differently when crossed with strains carrying *bWx76*, *bWx77*, *bWx79*, and *bWx80* (Table 3-12) but a comparison of their sequences (Figure 3-20) shows that they differ at ten positions. It can be concluded that this region of *bE* from positions 92 to 104 also contains amino acids that affect specificity.

Table 3-12. Mating reactions involving *bEx92* and *bEx107* alleles. Extracted from Table 3-4.

Allele	bEx92	bEx107
bWx76	-	++
bWx77	-	++
bWx79	-	+++
bWx80	-	+++

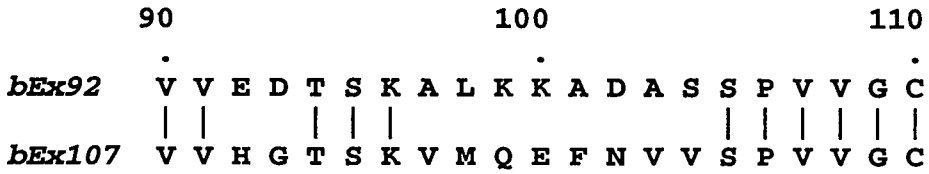

Figure 3-20. Comparison of sequences of *bEx92* and *bEx107*.

A third example of a specificity border across multiple amino acid changes occurred between the alleles *bEx156* and *bE1*. The *bEx156* allele contains amino acids 1 to 156 of *bE1* (all of the variable region plus part of the homeodomain) fused to amino acids 157 to 473 of the *bE2* gene. A comparison of the amino acid sequence of *bEx156* and *bE1* reveals five differences in the homeodomain and three differences in the constant region. It is possible that these differences in the homeodomain have an effect on specificity, either through some effect on dimerization or through an effect on interactions between the *b* proteins and the target DNA.

A fourth example involved the strains with alleles *bWx52* and *bWx68*, which had different specificities when crossed with *bEx89* and *bEx90* (Table 3-13). These alleles differ at 12 positions (Figure 3-21). Nevertheless, this analysis indicates that one or more of these differences with the region from position 52 to 68 of *bW* had an affect on specificity and this region is likely to be on the dimer interface.

Table 3-13. Mating reactions involving *bWx52* and *bWx68*. Extracted from Table 3-4.

Allele	*bEx89*	*bEx90*
bWx52	+	++
bWx68	-	-

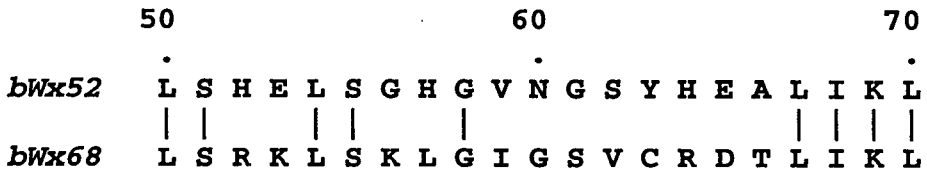

Figure 3-21. Comparison of sequences of *bWx52* and *bWx68*.

The last example of a specificity boundary between two alleles that differ at multiple positions occurred between *bWx109* and *bWx140*. These two alleles differ at 19 positions between codons 109 and 140, which is still within in the variable region of *bW*. This result indicates that this region may lie in the bE to bW dimer interface.

3.2.4 Genetic Evidence for Individual Amino Acid Interactions Between bE and bW Proteins

Some valuable insight into specific amino acid interactions between bE and bW can be derived from a fortuitous combination of crosses resulting in a "checkerboard" pattern of mating reactions involving the four alleles, *bEx90, bEx92, bWx74,* and *bWx76* (Table 3-14). A close examination of this checkerboard pattern reveals a possible direct interaction between amino acids at bE position 90 and bW position 74.

Table 3-14. Mating reactions involving *bEx90, bEx92, bWx74*, and *bWx76* Extracted from Table 3-4.

Allele		*bEx90*	*bEx92*
	a.a.	Ala90	Val90
bWx74	His74	-	++
bWx76	Tyr74	++	-

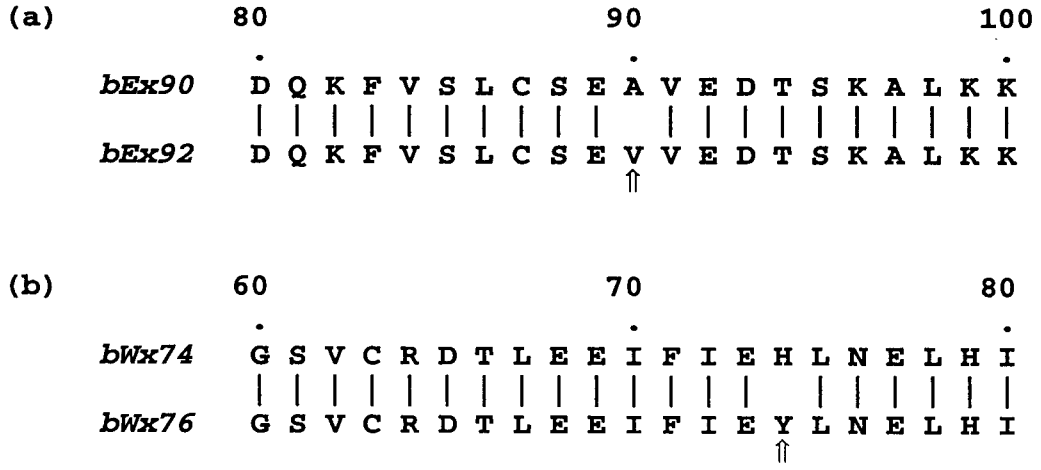

Figure 3-22. Comparison of sequences of *bEx90, bEx92, bWx74, and bWx76.*

Starting from the upper left of Table 3-14, the incompatible cross involving *bEx90* and *bWx74*, is changed to a compatible cross between *bEx92* and *bWx74* strains at the upper right. The two alleles *bEx90* and *bEx92* differ by only one amino acid (Figure 3-22a); *bEx90* has an Ala at position 90 and *bEx92* has a Val. This means that the Val for Ala substitution changed an incompatible cross to a compatible one. It is likely that the Val substitution increased the hydrophobic interaction between the bE and bW proteins, thus favoring dimerization.

If we go back to incompatible cross at the upper left of Table 3-14, and move down, we observe that the cross between *bEx90* and *bWx76* is compatible. Again, the alleles *bWx74* and *bWx76* differ by only one amino acid; *bWx74* has a His (H) at position 74 and *bWx76* has a Tyr (Y) (Figure 3-22b) The substitution of Tyr for His changed an incompatible cross to a compatible one. Tyr is larger than His (van der Waals volume of 141 $Å^3$ versus 118 $Å^3$), so the size

difference would not favor dimerization. Although Tyr is not strictly considered a hydrophobic amino acid, it does have hydrophobic characteristics due to its aromatic ring, and it is more hydrophobic than His. Thus, it is plausible that the Tyr substitution increased the hydrophobic interaction between the bE and bW proteins, favoring dimerization.

The two substitutions described above, one in *bEx90* and the other in *bWx74*, favored compatibility when carried out independently. Surprisingly, their effects appeared to cancel each other out when combined. This was the case with the incompatible cross of *bEx92* and *bWx76*, at the lower right of Table 3-14. A possible explanation of this canceling effect is that this cross may have positioned a Val at position 90 of bE opposite a Tyr at position 74 of bW on the dimer interface. Val has a β-branched aliphatic side chain and Tyr has a bulky aromatic ring. These two side chains, due to their shape and position on the dimer interface, could interfere with the dimerization of bE and bW, through unfavorable "knob-to-knob" interactions. Note that the wild type sequence of bE1 contains a Val at position 90 and bW1 contains a Tyr at position 74. This is consistent with the premise that the Val 90 and Tyr 74 result in unfavorable steric interactions on the dimer interface and prevents dimerization between bE1 and bW1.

The compensating effect of two independent amino acid changes does not necessarily prove that these two amino acids interact directly. One can imagine a situation where the changes interact indirectly; i.e. one amino acid might cause a change in conformation, which is then canceled out by a conformational change induced by the other amino acid. However, the nature of the compensating changes suggest that a direct amino acid interaction is plausible. Further evidence to support this hypothesis could be obtained by manipulating the residues at these two positions through site-directed mutagenesis. These experiments are described in Section 3.3.

The apparent sensitivity of these four alleles to amino acid changes is likely because the proteins are close to a dimerization "balance point", so that relatively small changes in interactions favoring or disfavoring dimer formation will have a detectable effect on compatibility.

The second example of a interaction between positions on bE and bW involved strains carrying alleles *bEx31* and *bEx39* when they were crossed with strains carrying alleles *bW2* and *bWx9* (Table 3-15). A comparison of *bEx31* and *bEx39* showed that *bEx39* has a Arg (R) to Tyr (Y) substitution at position 31 (Figure 3-23), and this single amino acid change had two opposite effects depending upon what was at positions 2 to 6 of bW. The Tyr gave a compatible reaction when the strain carrying *bEx39* is tested against *bW2*, but it gave an incompatible reaction when tested against *bWx9*. There are three amino acid changes between alleles *bW2* and *bWx9* (Figure 3-24). The amino acids Lys (K), Phe (F), and Tyr (Y) are substituted at positions 2, 4, and 6 respectively in allele *bWx9*. These data indicate an interaction between bE position 31 and bW positions 2 to 6. The interaction could involve only one of the amino acids at positions 2 to 6 of bW or more than one. It is interesting to note that the incompatible combination of alleles bEx39 and bWx9 would potentially bring together a Tyr from bE with a Lys, Phe, and Tyr from bW; 3 out or 4 of these are aromatic amino acids.

Table 3-15. Mating reactions involving *bEx31* and *bEx39*. Extracted from Table 3-4.

Allele	bEx31	bEx39
	Arg31	Tyr31
bW2	-	+++
bWx4	-	-
bWx6	-	-
bWx9	+++	-

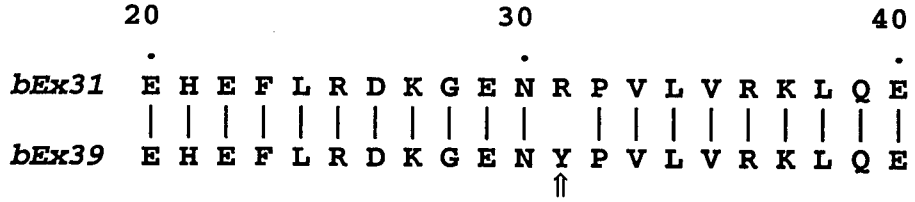

Figure 3-23. Comparison of sequences of *bEx31* and *bEx39*.

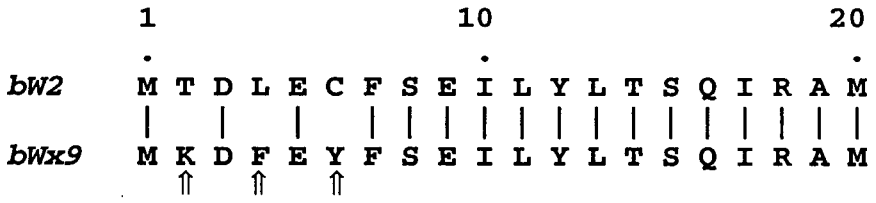

Figure 3-24. Comparison of sequences of *bW2* and *bWx9*.

3.2.5 Summary

A matrix of mating reactions involving two series of chimeric alleles, one from *bE* and the other from *bW*, has revealed additional amino acid positions that have affected bE to bW dimerization. For bE, in addition to the positions 31 and 79 previously revealed in Section 3.1, the positions 45, 87, and 90 were shown to be relevant to specificity. For bW, in addition to the positions 6, 74, 77, 79, 81, and 82 presented in Section 3.1, the positions 2, 9, 76, and 80 have been identified. The nature of the amino acid changes that alter specificity continue to range from hydrophobicity changes, to electrostatic charge reversals, to size and shape changes. However, some key information was obtained, identifying amino acid positions that may be interacting with each other on the dimer interface. These data indicate there is an intermolecular interaction between position 31 of bE and positions 2 to 6 of bW. Also there is an interaction between position 90 of bE and position 74 of bW, and this interaction may interfere with dimerization between bE1 and bW1 through steric hindrance between side chains.

3.3 Site-Directed Mutagenesis

There are often two sides to the problem of specificity in protein-protein interactions. One involves attractive forces that favor dimerization, and the other involves forces that prevent dimerization - I shall call these the anti-dimerization forces. An excellent example of this is the specific protein-protein interactions between the leucine zipper domains of Fos and Jun, or Myc and Max. The heterodimers of Fos to Jun or Myc to Max form due to attractive hydrophobic and electrostatic interactions, but the homodimers are prevented from forming due to repulsive electrostatic interactions (Amati et al. 1993; O'Shea et al. 1992).

The results presented in the previous chapter, obtained from crosses between strains with bE and bW chimeric alleles, indicate a possible interaction between amino acid residues at bE position 90 and bW position 74. The interactions appeared to be steric in nature, with the larger side chains such as Tyr and Val preventing compatibility, presumably by interfering with the dimerization of bE and bW.

The experiment presented in this section was designed to provide clearer evidence that these two positions do indeed interact with each other, and to clarify the nature of that interaction. The evidence was obtained by site-directed mutagenesis to test the effect of several different amino acids at both positions.

3.3.1 Site-Directed Mutants of bE Position 90 and bW Position 74

The residues at position 90 of bEx90 and position 74 of bWx74 were each mutated to the amino acids Ala (A), Val (V), Tyr (Y), Phe (F), Met (M), or Leu (L). These residues were chosen on the basis of side chain size and shape. The premise was that various side chains would affect dimerization through favorable or unfavorable interactions on the dimer interface. Charged amino acid were not tested because these did not naturally occur at these two positions, therefore it was concluded that electrostatic charge was not involved in allelic specificity at these two positions. The Ala side chain is the smallest, containing only one hydrocarbon (van der Waals volume 67 $Å^3$), and would be expected to give the least steric hindrance on the dimer interface. The Val

side chain is considerably larger, with three hydrocarbons in a "V" arrangement (105 Å^3). Its β-carbon branched structure is likely to give different packing characteristics compared with a side chain that is unbranched. The Tyr (141 Å^3) and Phe (135 Å^3) side chains both have bulky aromatic rings attached to the β-carbon, with Tyr differing only in having a hydroxy group attached to the aromatic ring. The Met side chain is somewhat bulky but unbranched, containing a four atom backbone with sulfur as the δ-atom (124 Å^3). The Leu side chain is also somewhat bulky, consisting of four hydrocarbons with a branch at the γ-carbon (124 Å^3). The His side chain has a smaller five member ring structure attached to the β-carbon (118 Å^3), so it is not as bulky as Tyr or Phe. These side chains differ in their hydrophobicities, as this is related to the size of the hydrocarbon side chain. The side chains are all uncharged, with the exception of His which potentially can have a positive charge. At pH 7.0, the His side chain is largely uncharged, with a ratio of neutral to charged species of 10:1 (Rawn 1989).

These site-directed mutants were made in *bEx90* and *bWx74* chimeric alleles backgrounds. These two alleles were part of the "checkerboard pattern" that illustrated the interaction between bE position 90 and bW position 74 (Section 3.2.4). The proteins from these two alleles are likely close to a dimerization balance point.

The site-directed mutagenesis was carried out using the megaprimer PCR technique as outlined in Materials and Methods. Tables 3-16 and 3-17 list the mutant *U. maydis* strains, the designations for the mutant alleles, and their phenotypes when crossed with the standard tester strains. All the *U. maydis* mutant strains used for these mating reactions were checked for the correct mutant sequence and they all displayed single copy homologous integration except for the strains tBWX74-H74V-1 and tBWX74-H74V-2. These two strains displayed double integration, one at the *b* locus giving a 10.5 kb band, and one ectopic giving a 9 kb band. This specific pattern of ectopic integration in other *U. maydis* transformants did not affect the *b* phenotype (See Section 3.1.3). Strains carrying *bE* alleles have inactivated *bW* alleles and visa versa.

Table 3-16. Site-directed mutants of *bEx90* and their phenotype when crossed with 521 *(a1b1)* and 031 *(a1b2)*, and when grown alone.

		Mating Reaction with Tester Strain		
U. Maydis Strain	Allele Designation	521 *(a1b1)*	031 *(a1b2)*	Alone
t17-2	*bEx90*	-	+++	-
tb-21	*bEx92*	-	+++	-
tBEX90-A90Y-1	*bEx90-A90Y*	-	+++	-
tBEX90-A90Y-2	*bEx90-A90Y*	-	+++	-
tBEX90-A90F-1	*bEx90-A90F*	-	+++	-
tBEX90-A90F-2	*bEx90-A90F*	-	+++	-
tBEX90-A90M-1	*bEx90-A90M*	+++	+++	-
tBEX90-A90M-2	*bEx90-A90M*	+++	+++	-
tBEX90-A90L-1	*bEx90-A90L*	+++	+++	-
tBEX90-A90L-2	*bEx90.A90L*	+++	+++	-

Table 3-17. Site-directed mutants of *bWx74* and their phenotype when crossed with 032 (a2*b1*) and 518 (a2*b2*), and when grown alone.

		Mating Reaction with Tester Strain		
U. Maydis Strain	Allele Designation	032 *(a2b1)*	518 *(a2b2)*	Alone
tBWX74-1	*bWx74*	+++	+++	-
tBWX74-3	*bWx74*	+++	+++	-
tBWX76-1	*bWx74-H74Y*	-	+++	-
tBWX76-3	*bWx74-H74Y*	-	+++	-
tBWX74-H74A-1	*bWx74-H74A*	+++	+++	-
tBWX74-H74A-2	*bWx74-H74A*	+++	+++	-
tBWX74-H74V-1	*bWx74-H74V*	-	+++	-
tBWX74-H74V-2	*bWx74-H74V*	-	+++	-
tBWX74-H74F-1	*bWx74-H74F*	-	+++	-
tBWX74-H74F-2	*bWx74-H74F*	-	+++	-
tBWX74-H74M-1	*bWx74-H74M*	-	+++	-
tBWX74-H74M-2	*bWx74-H74M*	-	+++	-
tBWX74-H74L-1	*bWx74-H74L*	-	+++	-
tBWX74-H74L-2	*bWx74.H74L*	-	+++	-

3.3.2 Opposing Val and Tyr Residues at bE Position 90 and bW Position 74 Prevent Compatibility

The site-directed mutants of *bEx90* were crossed with those of bWx74, in all combinations. The expectation was that large side chains, when present in both bE and bW would lead to an incompatible reaction. Results of these crosses are given in Table 3-18. These data are quite revealing in terms of possible residue to residue interactions at these two positions. With a small residue Ala at both positions, the cross gave a strong compatible reaction. The substitution of a larger Val or Tyr at bW position 74 resulted in a weaker compatible reaction, presumably because the size of the side chain was interfering slightly with the dimerization of bE and bW. Even further revealing is the substitution of either Val or Tyr at both positions. This gave incompatible matings reactions for all combinations of Val and Tyr. The combination of two larger side chains was preventing dimerization between bE and bW.

It is interesting to note that in bE1 and bW1 wild type sequences, the side chains are Val 90 and Tyr 74 respectively. This combination, as shown above, potentially results in steric hindrance between these two positions. This is consistent with the incompatible combination of bE1 and bW1. These side chains are altered in bE2 and bW2 to Ala 90 and His 74 respectively. The compatible combinations of bE1 and bW2, or bE2 and bW1, would result in the pairing of Val^{E90} and His^{W74}, or of Ala^{E90} and Tyr^{W74} respectively. Again, this is consistent with the site-directed mutagenesis data which showed that the combinations of Val^{E90} and His^{W74}, or of Ala^{E90} and Tyr^{W74} allowed compatiblity.

Superscripts on amino acids denote their position in bE or bW, i.e. Val ^{E90} denotes Val on bE position 90.

Table 3-18. Crosses between *U. maydis* strains carrying alleles with site-directed mutations at position 90 of bEx90 and position 74 of bWx74.

Amino Acid Residue at Position 74 of bWx74	Amino Acid Residue at Position 90 of bEx90					
	Ala	Val	Tyr	Phe	Met	Leu
Ala	+++	+++	+++	+++	+++	+++
Val	++	-	-	-	+++	+++
Tyr	++	-	-	+++	+++	+++
Phe	-	-	-	-	+++	+++
Met	-	-	-	-	+++	+++
Leu	-	-	-	-	++	+++
His	-	++	**na**	**na**	**na**	**na**

na - data not available

The site-directed mutations with Phe, Met, and Leu gave results that were less clear and there was a divergence as to how these substitutions behaved at bEx90 position 90 versus bWx74 position 74. At bE position 90, the presence of a Phe gave compatible crosses when combined with Ala or Tyr, but not with Val, Phe, Met, or Leu. Compatibility when Ala is at position 90 but not when the larger side chains are present can be explained by steric hindrance, but the compatibility with Tyr is not consistent with this. The Tyr side chain is exactly the same as the Phe side chain, except for an additional hydroxy group on the aromatic ring of Tyr which is capable of hydrogen bonding. Possibly the hydroxy group is hydrogen bonding with some unknown group on bEx90, and this may be tilting the bE-bW dimerization balance enough to give a compatible reaction.

The presence of Met at bE position 90 gave compatible reactions regardless of the residue at bW position 74. The Met side chain is not as bulky as the aromatic side chains and it is

unbranched. It is also significantly more hydrophobic than Ala. It is possible that the unbranched Met side chain at bE position 90 packs better than the other bulky side chains, and the hydrophobicity of Met favors dimerization regardless of the amino acid at bW position 74. The effect of Met at bW position 74 does not follow the same pattern. It is compatible only with Met and Leu at bE position 90. This lack of reciprocity may be due to differences in the positioning and orientation of the side chains on bE position 90 versus bW position 74. For example, if the side chain at bE position 80 was oriented at an angle to the dimer interface, then a long side chain would not offer as much steric interference compared to a side chain oriented perpendicular to the dimer interface. This explanation is supported by the secondary structure predictions of the protein in these two regions. The structure surrounding bE position 90 is predicted to be nonhelical, whereas the protein surrounding bW position 74 is predicted to be helical (see Section 4.2).

Substitution of Leu at bE position 90 gave a pattern similar to that of Met, such that it give a compatible reaction with all residues at bW position 74. Again, Leu is not as bulky as the aromatic side chains (van der Waal volume 124 $Å^3$), it is strongly hydrophobic, and although it is branched, it is γ-carbon branched, as opposed the β-carbon branched Val. This γ-carbon branching, which may be imagined as a knob at the end of flexible peg, could pack in a "zipper tooth" like fashion with other side chains, whereas the β-carbon branched Val is imagined more as a knob on short stiff peg, and would not be able to pack in the same manner. Differences in side chain packing of Leu versus Val have been reported in leucine zipper structures (Hu et al. 1990). Most importantly, the hydrophobicity of Leu would favor dimerization as long as packing of the Leu side chain is favorable. The lack of reciprocity with Leu at bW position 74 follows the same pattern as for Met at that position, and is likely due to the same reasons of position and orientation of the side chains.

3.3.3 Summary

The results of site-directed mutagenesis for Ala, Val, and Tyr showed a clear effect of side chain size and shape. The simultaneous presence of bulky Val or Tyr residue at bE position 90 and bW position 74 had a negative effect on dimer formation. Dimerization occured when one or both of these positions were replaced by Ala. These results support the hypothesis that dimerization of bE1 and bW1 is prevented by unfavorable "knob-to-knob" interactions between bE1 Val 90 and bW1 Tyr 74.

The effect of Phe, Met, and Leu side chains were less clear, probably complicated by factors such as hydrophobicity and orientation of the side chain. The effects of these side chains are less relevant to the question of *b* locus specificity because these amino acids do not occur at these positions in the wild type *b* alleles. The empirical effect of a specific side chain is a result not only of its size and shape, but also of its hydrophobicity and its environment in the dimer interface. It is not possible to understand the net effect of all these factors from genetic data alone.

4. DISCUSSION

4.1 Chimeric Allele Analysis Identifies Multiple Positions in bE and bW which Affect Specificity

Dual sets of chimeric alleles have been constructed in both *bE* and *bW* genes. These alleles can be thought of as a series of site-directed mutants, many differing from each other by single amino acid substitutions. The effect of these amino acid substitutions on specificity was determined by crossing the chimeric alleles against wild type "tester" strains, as well as by crossing the strains carrying *bE* chimeric alleles with strains carrying *bW* chimeric alleles. This allowed identification of the position and nature of amino acid substitutions affecting specificity in both bE and bW.

The power of this chimeric allele approach to studying *b* locus specificity lies in three aspects of the experimental design. First, it focuses on amino acid differences that exist naturally between wild type alleles, and thus provides more relevant information to the question of *b* locus specificity. This is in contrast to the random PCR mutagenesis approach of Kamper (1995) where only 1 out of a total of 5 single mutations to *bE2* represented an amino acid naturally present at these positions. All of the mutations isolated were to hydrophobic residues (Val, Ile, or Leu) and they all had the effect of promoting dimerization between bE2 and bW2. Mutations to hydrophobic residues are likely to increase the strength of hydrophobic interactions between bE and bW and would be expected to promote dimerization, since hydrophobic interactions are a significant, if not dominant, driving force behind protein dimerization (Jones and Thornton 1996; Young et al. 1994). Single hydrophobic residue replacements have been reported to promote binding between proteins (Lee et al. 1995).

Secondly, the construction of a series of closely spaced chimeric alleles allowed the identification of individual amino acid positions (rather than blocks of amino acids) that were important for specificity. This provided information on specific amino acid substitutions that affected specificity and even identified possible intermolecular interactions between individual

residues of bE and bW. This fine mapping strategy provided far greater insight into the determinants of specificity at the *b* locus.

A third aspect of the experimental design, that of crosses between strains with chimeric *bE* and *bW* alleles, unveiled many important amino acid substitutions that were hidden from detection when the chimeric alleles were crossed with wild type *b1* and *b2* testers. When wild type *b1* and *b2* alleles were used, these crosses revealed bE positions 31 and 79, and bW positions 6, 74, 77, 79, 81, and 82. When chimeric alleles were used as testers, three additional positions in bE and four in bW were shown to affect specificity. In bE these were positions 45, 87, and 90, and in bW these were positions 2, 9, 76, and 80. This resulted in a combined total of 15 positions in bE and bW that have been identified by chimeric allele analysis. All these single amino acid changes have been summarized in Table 4-1. More importantly, valuable information was revealed on possible direct interactions between specific bE and bW residues. The mating tests involving bE position 90 and bW position 74 suggest a "knob-to-knob" interaction exists between the Val at position 90 of bE1 and the Tyr at position 74 of bW1. This interaction, which exists in the wild type alleles, is likely to hinder dimerization between bE1 and bW1 proteins. A similar "knob-to-knob" interaction may exist between Tyr 31 of bE1 and Tyr 6 of bW1.

Examination of the data in Table 4-1 shows that the type of substitutions that affect specificity, and thus dimerization of bE and bW, span the range of interactions from hydrophobic, electrostatic, to van der Waals (steric). The single amino acid substitutions are presented in Table 4-1 from the perspective of substitutions that cause incompatiblity. This can be interpreted as substitutions that have a negative effect on dimerization. The reason for this originates from the premise that allelic specificity is determined by interactions that prevented dimerization between proteins from same allele combinations, and is related to the critical question asked in Section 1.5; "why does dimerization not occur between bE and bW proteins from the same allele?".

To answer the above question, it is important to examine the data with the intent of distinguishing between the "signal" and the "noise" - the "signal" being the interactions that actively prevent dimerization between bE and bW from the same allele, and the "noise" being all

the other interactions that affect dimerization, but do not necessarily determine recognition between self alleles. These "noise" interactions are likely to affect the general cohesiveness of the dimer interface, and are more relevant to the nonspecific ability of an allele to be compatible with other nonself alleles. To separate the "signal" from the "noise", one would look for amino acid changes that could actively prevent dimerization. Table 4-1 summarizes the 18 individual amino acid changes at 15 positions in bE and bW that cause incompatibility. The bE positions 31 and 90, and bW position 74 contributed two amino acid changes (substitutions in either direction) because they were positions that showed an interaction between bE and bW.

The most apparent type of substitution that could prevent dimerization are amino acid changes that could produce unfavorable steric interactions. Possible unfavorable steric substitutions made up 9 of the total 18 substitutions. What is striking about these positions is that 4 of these involve the introduction of an aromatic side chain, usually Tyr. Tyr and Phe have a large van der Waals volume and their aromatic structure results in a "knob" like shape. The presence of two bulky residues opposite each other on the dimer interface could interfere with dimerization through unfavorable steric interactions. An example of this was reported by Ridgeway et al. (1996) as applied to the dimerization of antibody heavy chains. Here they showed that an engineered "knob-into-holes" arrangement of Tyr opposed to Thr on the interface between two heavy chain antibody molecules favored dimerization while the "knob-to-knob" arrangement of opposing tyrosines had the opposite effect. A similar explanation was presented for the negative effect on antigen binding of a Trp substitution at site 33 of the monoclonal antibody HyHEL10 (Tsumoto et al. 1995). Trp has a very large side chain (163 $Å^3$) and could potentially disrupt the shape complementarity of the antibody-antigen interface.

Tyr has a combination of characteristics that make it special. Tyr has been shown to have an unusually high occurrence in the antigen binding sites of antibodies (Davies and Cohen 1996; Padlan 1990). It has a unique ability to play multiple roles in antibody-antigen interactions (Tsumoto et al. 1995), such as forming H-bonds (Bhat et al. 1994; Padlan et al. 1989), forming favorable aromatic-aromatic and aromatic-Arg interactions (Burley and Petsko 1988; Schreiber and

Fersht 1995; Serrano et al. 1991), contributing to hydrophobic interactions (Tsumoto et al. 1995), and forming favorable van der Waals contacts (Rini et al. 1992). Thus, its presence on a dimer interface would generally favor dimerization, except when it is positioned opposite another bulky amino acid.

A second type of change that could actively prevent dimerization is electrostatic in nature. Possible unfavorable electrostatic changes were present at 2 positions (Glu^{W9} to Lys and Lys^{W77} to Glu), and they involved clear charge reversals. These residues could be positioned on the dimer interface close to charged residues belonging the partner protein. Oppositely charged residues would favor dimerization and like charged residues would disfavor dimerization. This is similar to the mechanism of specificity for Fos and Jun leucine zipper motifs (O'Shea et al. 1992).

The third type of change that could actively prevent dimerization is a nonpolar to polar substitution. These are apparent at 2 positions (Ile^{E45} to Thr and Ile^{W80} to Arg). Conceivably, the introduction of a polar residue at a location in the dimer interface dominated by nonpolar residues would be disruptive on the hydrophobic interactions between bE and bW, and have a negative effect on dimerization.

The remaining 5 positions involve mostly decreases in hydrophobicity of the side chains. These decreases in hydrophobicity would disfavor dimerization by reducing the strength of the hydrophobic interaction. This is a passive process, as opposed to actively preventing dimerization. As such, these changes are not likely related to the question of recognition between same alleles. They constitute the "noise" component of the data.

Noted that chimeric allele analysis is not capable of identifying all of the positions affecting specificity between two alleles. It is only capable of identifying positions having amino acids that differ between the two alleles; the role of conserved positions cannot be determined using this strategy. Thus, there are likely to be many more positions on the dimer interface that are capable of affecting specificity than those identified in this study. Also, without three dimensional protein structural data on bE and bW, it is not possible to rule out intramolecular conformational effects as an explanation for specifcity changes resulting from the amino acid substitution discussed above.

Table 4-1. Summary of all individual amino acid changes that cause incompatibility. Superscripts on amino acids denote their position in bE or bW, i.e. Tyr ^{E31} denotes Tyr on bE position 31.

Allele Pair	Substitution	Possible Interaction Change	Incompatibility Effect
bEx31/39	Arg^{E31} to Tyr	steric; charge	*bEx31* incomp. with *bWx9*
bEx39/31	Tyr^{E31} to Arg	hydrophobicity loss; charge	*bEx31* incomp. with *bW2*
bEx48/45	Ile^{E45} to Thr	hydrophobicity loss	*bEx48* incomp. with *bWx4/12*
bEx79/82	Ile^{E79} to Phe	steric	*bEx79* incomp. with *bWx82/bW1*
bEx89/87	Cys^{E87} to His	steric	*bEx87* incomp. with *bWx12 /82*
bEx92/90	Val^{E90} to Ala	hydrophobicity loss	*bEx90* incomp. with *bWx74*
bEx90/92	Ala^{E90} to Val	steric	*bEx92* incomp. with *bWx76*
bW2/bWx4	Thr^{W2} to Lys	steric; charge	*bW2* incomp. with *bEx39*
bWx9/6	Tyr^{W6} to Cys	hydrophobicity loss	*bWx6* incomp. with *bEx31/bE2*
bWx9/12	Glu^{W9} to Lys	charge reversal	*bWx9* incomp. with *bEx48*
bWx74/76	His^{W74} to Tyr	steric	*bWx76* incomp. with *bEx92/bE1*
bWx76/74	Tyr^{W74} to His	hydrophobicity loss	*bWx74* incomp. with *bEx90*
bWx76/77	Gln^{W76} to Arg	steric; charge	*bWx76* incomp. with *bEx90*
bWx79/77	Lys^{W77} to Gln	charge reversal	*bWx77* incomp. with *bEx89/bE1*
bWx79/80	His^{W79} to Arg	steric; charge	*bWx79* incomp. with *bE1*
bWx80/81	Ile^{W80} to Arg	hydrophobicity loss, charge	*bWx81* incomp. with *bEx89*
bWx82/81	Val^{W81} to Gly	hydrophobicity loss	*bWx81* incomp. with *bEx89/bE1*
bWx82/83	Cys^{W82} to Tyr	steric	*bWx82* incomp. with *bEx89/bE1*

incomp. - incompatible

The chimeric allele analysis was also able to identify blocks of amino acids in bE and bW that affected specificity (Table 4-2), but because these regions were not "fine mapped", it was not possible to attribute specificity changes to individual amino acids. This information is still useful for locating addition regions that may be part of the dimer interface or affect dimerization in some indirect manner. One of these additional regions was in the bE homeodomain from codons 156 to 169. This result suggests that differences between alleles in this region may affect dimerization between bE and bW proteins, either through an indirect effect on the conformation of the specificity region or possibly through an effect on binding of the dimer to the DNA target.

Table 4-2. Summary of multiple amino acid changes that affect specificity.

Allele Pair	Positions	Changes	Effect
bEx70/79	bE pos. 70 to 74	4	*bEx79* incomp. with *bWx83* to *109*
bEx92/107	bE pos. 92 to 104	10	*bEx92* incomp. with *bWx76* to *80*
bEx156/bE1	bE pos. 156 to 473	8	*bE1* incomp. with *bWx76* to *80*
bWx2/9	bW pos. 2 to 6	3	*bWx9* incomp. with *bWx39*
bWx52/68	bW pos. 52 to 66	12	*bWx68* incomp. with *bEx89* and *90*
bWx109/140	bW pos. 109 to 140	19	*bWx109* incomp. with *bEx79*

pos. - positions

4.2 Identification of Interactions Between bE and bW

Chimeric allele analysis has also resulted in evidence for interaction between individual amino acids in bE and bW. This evidence was revealed by the rare "checkerboard" patterns produced from crossing strains with chimeric *bE* and *bW* alleles. The best example of this involved the interaction between bE position 90 and bW position 74 (previously presented in Section 3.2.4 and re-illustrated in Table 4-3). The independent substitutions of a Val at bE position 90 or a Tyr at bW position 74 caused a change from incompatibility to compatiblity, but surprisingly, these effects are reversed when carried out simultaneously. It was hypothesized that there was a direct interaction between amino acids at these two positions and that the presence of both Val and Tyr side chains were interfering with dimerization between bE and bW through "knob-to-knob" interactions.

Table 4-3. Mating reactions involving *bEx90, bEx92, bWx74, and bWx76.* The pair of chimeric alleles *bEx90* and *bEx92* differ by one amino acid, as do the pair *bWx74* and *bWx76.*

Allele		bEx90	bEx92
	a.a.	Ala90	Val90
bWx74	His74	-	++
bWx76	Tyr74	++	-

The second example of an interaction between individual amino acids in bE and bW occurred near the N-terminus of both proteins. This interaction was revealed by a checkerboard pattern involving the alleles *bEx31*, *bEx39*, *bW2*, and *bWx9* (previously presented in Section 3.2.4 and re-illustrated in Table 4-4 below). These mating tests revealed an interaction between position 31 of bE and positions 2 to 6 of bW. In this case, the data does not allow an individual amino acid to be identified for bW. Nevertheless, the ability to localize the interaction to five residues is valuable. The hypothesized interacting positions could be extremely valuable for determining the orientation of the bE and bW proteins once their individual structures are obtained.

The task of determining the correct "docking" arrangement of two proteins is not a trivial one, even when the individual structures are known (Weng et al. 1996; Young et al. 1994).

Table 4-4. Mating reactions involving *bEx31, bEx39, bW2, and bWx9.* This representation has been simplified from Table 3-4 to highlight the checkerboard pattern. The alleles bEx31 and bEx39 differ by one amino acid, and the alleles bW2 and bWx9 differ by 3 amino acids.

Allele	bEx31	bEx39
	Arg31	Tyr31
bW2	-	+++
bWx9	+++	-

4.3 Four Specificity Subdomains Defined by Chimeric Allele Analysis

Four locations, two on bE and two on bW, are involved in intermolecular interactions, and could be positioned opposite each other on the dimer interface. The secondary structure in the vicinity of these four locations can be predicted with reasonable accuracy, due to recent improvements in secondary structure prediction algorithms (Rost and Sander 1996). Figure 4-1 and Figure 4-2 present secondary structure predictions for bE and bW using the PHD (Rost and Sander 1993; Rost et al. 1993), SOPMA (Geourjon and Deleage 1994; Geourjon and Deleage 1995), and DSC (King and Sternberg 1996) prediction programs. These programs have been shown to have high accuracy rates in the range of 70% for predicting secondary structure of proteins (Geourjon and Deleage 1995; King and Sternberg 1996; Rost and Sander 1995). The PHD and DSC programs are considered to be the two most reliable secondary structure prediction tools currently accessible through the Internet (Dunbrack et al. 1997). All three use an input of multiple sequence alignments of proteins closely related to the "guide" sequence (the submitted sequence). This significantly improves the accuracy of secondary structure predictions because the profile of amino acid exchanges in naturally evolved protein families are highly specific to their common structure (Rost and Sander 1995).

The predicted secondary structures obtained by all three programs correlate well with each other, even though they use very different algorithms for prediction. The known helices of the homeodomain have been identified from alignments of the bE and bW protein sequence with the homeodomains of *S. cerevisiae* MATα2 and *Drosophila* engrailed (Kues and Casselton 1992; Wolberger et al. 1991). These helices were predicted accurately by DSC and to a lesser extent by PHD and SOPMA. This indicates that the secondary structure predictions have a reasonable amount of accuracy.

The placement of the four interacting positions (bE positions 31 and 90, bW positions 2 to 6 and 74) into the predicted secondary structures allows the defining of specificity subdomains that are likely involved in protein-protein interactions between bE and bW. The first of these centers around bE position 31 (designated subdomain EN for bE N-terminal) and consists of the predicted loop structure from residues 26 to 32. The other subdomains are defined as EC, occupying the loop region in bE between positions 86 and 95, subdomain WN occupying the helix in bW between positions 3 and 19, and subdomain WC, occupying the helix in bW from residues 62 to 90. These four subdomains range in size from 7 to 28 amino acids. They all lie on or close to the N and C-terminal borders of the specificity class II regions for both bE and bW, confirming the importance of these specificity border regions.

There is an interesting difference between bE and bW when comparing predicted structures of the bE and bW subdomains. The bE subdomains are both predicted to be loop structures. This situation is reminiscent of immunoglobulin molecules where the antigen interface consists of hypervarible loops known as complementarity determining regions (CRD) (Creighton 1993; Davies et al. 1990). In contrast, the two bW subdomains are both predicted to be helical.

The four specificity subdomains defined by chimeric allele analysis are likely to reside on the dimer interface between bE and bW. Thus, this secondary structure prediction analysis presents a possible picture of the dimer interface, that of at least four contacting subdomains, with the two loop subdomains of bE contacting the two helical subdomains of bW. Additionally, there is a fixed orientation of the interacting subdomains of bE and bW, EN interacts with WN and EC

interacts with WC. This orientation is inferred from the evidence on interacting amino acids between bE and bW (previous Section 4.2) and on the proposed model for interactions between the specificity subdomains (following in Section 4.4). The proteins are likely to have other contacting interfaces, some identified by chimeric allele analysis (summarized in Tables 4-1 to 4-3 and labeled in Figure 4-1 and Figure 4-2) such as the helical region in bE from positions 57 to 79. These other contact interfaces may play a greater role in general cohesiveness between bE and bW as opposed to determining specificity, or they may play a role in determining specificity for allele combinations other *b1* and *b2*.

The importance of the proposed specificity subdomains for bE is confirmed by independent random mutagenesis data published by Kamper et al. (1995). They reported that mutations to hydrophobic residues in bE2 positions 31, 70, 87, 90, and 94 resulted in its dimerization with bW2. Four out of five of these positions are within the specificity subdomains defined for bE and three out of five of these positions are in common with the positions identified by chimeric allele analysis (bE positions 31, 87, and 90).

It must be emphasized that the arrangement of subdomains discussed here is hypothetical, based on genetic data and not structural data. Nevertheless, these proposed subdomains can be used to model the observations obtained from chimeric allele analysis, allowing for the formation of hypotheses which can be further tested.

```
                                                    50
bE1         MSSDPNFSLISFLECLNEIEHEFLRDKGENYPVLVRKLRELQQKIPNDIA
                                                 ⇑       •
DSC         .........HHHHHHHHHHHHHHHHH.....HHHHHHHHHHHH.......
SOPMA       ......HHHE.HHHHHHHHHHHHHHHH......HHHHHHHHHHH.......
PHD         .........HHHHHHHHHHHHHHHH.......HHHHHHHHHHHH..HHHH

                                                            100
bE1         NLPRDPETIQQIHQTTHRIRAVAQAFIRFDQKFVSLCSEVVHGTSKVMQE
                          ****     •       •  ⇑ **    ****
DSC         ......HHHHHHHHHHHHHHHHHHHHHHH.............HHHHHH
SOPMA       ......HHHHH..........HHHHHHEEE....EE........HHHHEE
PHD         ......HHHHHHHHHHHHHHHHHHHHHEEE..EEEEEE.....HHHHHHH

                                                            150
bE1         FNVVSPDVGCRNLSEDLPAYHMRKHFLLTLDNPTPRQEEKETLVRLTNES
            ****                  ----helix I----     --helix II-
DSC         ............HHH...HHHHHHHHH.........HHHHHHHHHH...
SOPMA       EEEE.........HHHHHHHHHHHHHEEH...........HHHHH....
PHD         H..............HHHHHHHHHHHH......HHHHHHEEE....

                                                            200
bE1         TARVGQSSVNRPPLEVHHVTLWFINARRRSGWSHILKKFAREDRSRMKHL
                 * **       ----helix III----
DSC         ............HHHHHHHHHH......HHHHHHHHHHHHHHHHHH
SOPMA       ..EEE..EE.....HHHEEEEE..........HHHHHHH..HHHHHH
PHD         .EE...........EE..HHHEHHHHHH..HHHHHHH..HHHHHHHH
```

Figure 4-1. Secondary structure prediction for bE variable domain and homeodomain. Amino acid positions hypothesized to interact between *bE* and *bW* labeled by ⇑. Other individual amino acids that affect specificity are marked by •. Groups of multiple amino acid that affect specificity are marked by *. Helix - H, Beta sheet - E. Other - . Regions marked as helix I, II, and III refer to homeodomain structures determined from alignments with MATα2 and *Drosophila* engrailed.

```
                    .         .         .         .         .         .  60
bW1         MKDFEYFSKILSLASQIRMTLPPLPRISQTAPRPTCFLPLSLEGPNQQALSRKLSKLGIG
             ⇑ ⇑ ⇑  •                                              **  ** *
DSC         ...HHHHHHHHHHHHHHH.................EEE.......HHHHHHHHHH....
SOPMA       HH..HHHHHHHH..HH..EEE............EEEEEEE.......EEHHHHHH...EE
PHD         ...HHHHHHHHHHHHHHH.................EEEE........HHHHHHHHHH....

                    .         .         .         .         .         .
bW1         SVCRDTLEEIFIEYLRKLRRVYEAQYENAFVTWQQENLYEEAYDQAFRKLLNRLFAMHSQ 120
            ******       ⇑ ••  •••                          *  *** *** *
DSC         .HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH..HHHHHHHHHHHHHHHHHH..HH
SOPMA       EEEHHHHHHEEEEEHHHHHH.EEHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH.....
PHD         ...HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH....

                    .         .         .         .         .         . 180
bW1         ETWHMVLDEVSKVFRTDSSLTVTQRDNASYEGAPLKTGRGHDSEAVRILEQAFKHSPNIT
            ** * * * *  *****                          ---helix I---
DSC         HHHHHHHHHHHHH.................. ..........HHHHHHHHHH.......
SOPMA       .HHHHHHHH..........EEEHH..H.............HEEEEEHHHE.......
PHD         ...EHHHHHHHHH.....................HHHHHHHHHH.......

                    .         .         .         .         .         . 240
bW1         PAEKFRLSEVTGLKPKQVTIWFQNRRNRKGKKNLNVEPTESTQPDLSPSRHESPPPSSPS
            --helix II-    -----helix III----
DSC         .HHHHHHHHH........EEEEHHHHHH.............................
SOPMA       .HHHH.EEEE......EEEEEEHHHH.......E.........EEEEEEE.........
PHD         HHHHHH..EE.......EEEEEE.....................................
```

Figure 4-2. Secondary structure prediction for bW variable domain and homeodomain. Amino acid positions hypothesized to interact between *bE* and *bW* labeled by ⇑. Other individual amino acids that affect specificity are marked by •. Groups of multiple amino acid that affect specificity are marked by *. Helix - H, Beta sheet - E. Other - . Regions marked as helix I, II, and III refer to homeodomain structures determined from alignments with MATα2 and *Drosophila* engrailed.

4.4 Implications of Specificity Subdomains

The four defined specificity subdomains can be used as a framework for interpreting the results of the chimeric allele experiments. First, the observation that class II chimeric alleles were compatible with both parental alleles indicates that compatibility requires nonself sequences interacting at only one of the subdomains on each protein. This is illustrated by Figure 4-3 where the specificity subdomains have been designated according to their alleles (e.g. EN1 and EC1 for the *bE1* subdomains). A natural compatible combination of bE1 and bW2 brings together nonself combinations of both subdomains on each partner (Figure 4-3a) while a natural incompatible combination of bE1 and bW1 results in self combinations of the subdomains (Figure 4-1b). The proteins are labeled at their N-termini and vertical lines indicate dimerization. A class II chimeric protein, bEx57, is capable of dimerizing with either parental bW1 or bW2 (Figure 4-1c and d). This suggests that dimerization is not prevented when sequences of self specificity are present at just one of the subdomains of each partner. That is to say, self recognition requires sequences with the same specificity at both subdomains simultaneously. Similarly, recognition of nonself (i.e. dimerization) at only one subdomain is sufficient for compatibility. This reasoning can be extended to combinations of class II chimeric proteins from both bE and bW (i.e. bEx57 and bWx52). These are incompatible because self sequences are brought together at both subdomains on each partner (Figure 4-3e). Finally, incompatibility between class II chimeric proteins from bE and bW, also confirm the hypothesized orientation of the bE and bW proteins; namely that the N-terminal subdomains interact with each other as do the C-terminal subdomains (an N to N and C to C orientation). Figure 4-3f represents the two proteins bEx57 and bWx52 in the opposite N to C orientation. This would bring together nonself combinations of subdomains, predicting compatibility between these combinations of class II chimeric alleles. This was clearly not observed, providing support for the proposed N to N orientation of the subdomains.

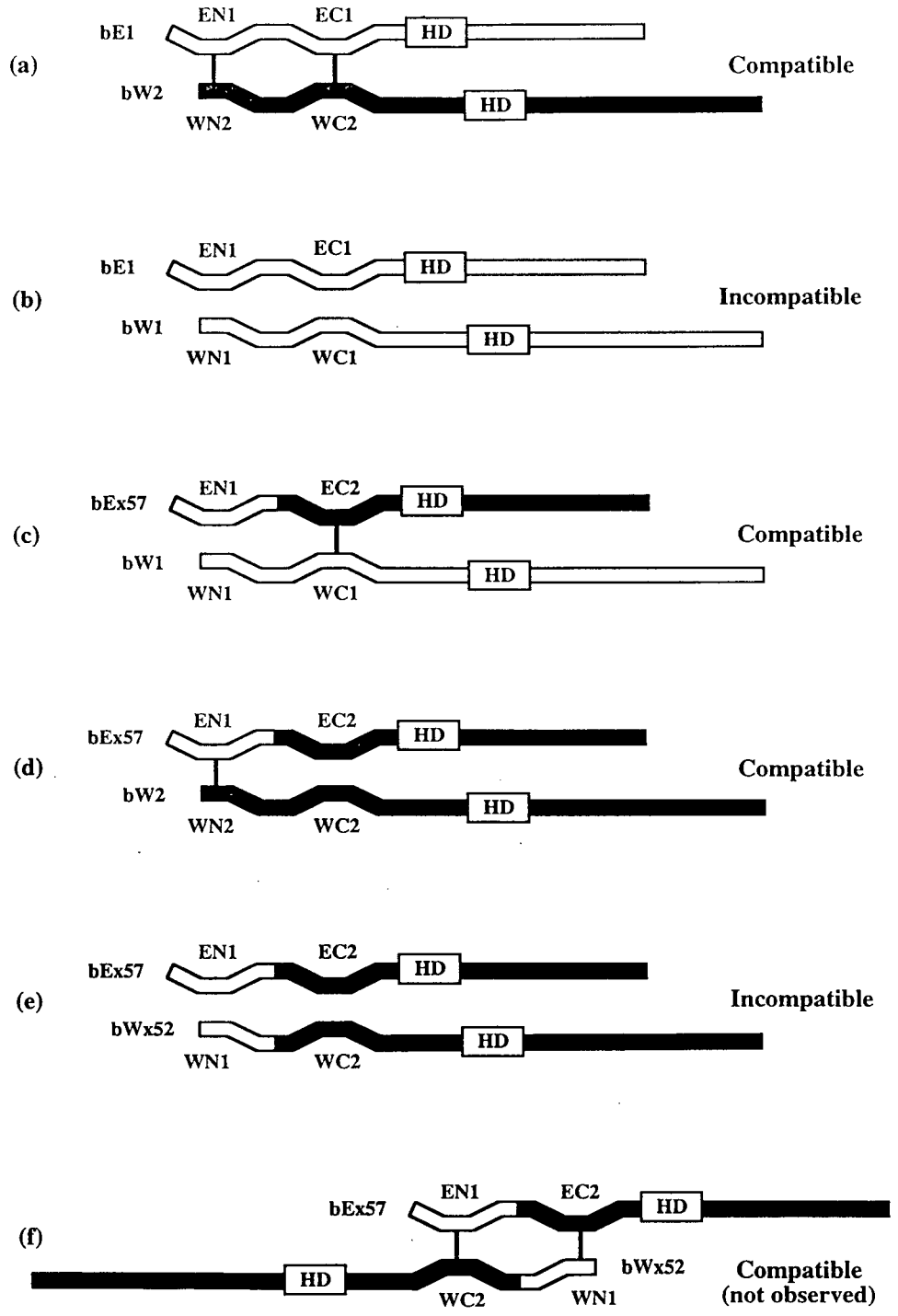

Figure 4-3. Model for interactions between the specificity subdomains. The four subdomains are designated EN, EC, WN, and WC, followed by a number (1 or 2) to indicate its parental allele. The proteins are labeled at their N-termini and vertical lines indicate nonself dimerization.

The above discussed requirements for compatible and incompatible allele combinations lead to two separate mechanisms for the evolution of the multiallelic *b* mating type system from a progenitor biallelic system. One of these mechanisms involve point mutations, the other involves meiotic recombination.

The first mechanism requires an initial set of two point mutations within one of the specificity subdomains to remove an interfering residue by substituting it with a less bulky residue, and to place a bulkier residue at a different location on the dimer interface. This might allow dimerization between one mutant b protein and its normally incompatible partner. Dimerization at only one of the subdomains in each protein is sufficient for compatiblity. These initial mutations produce an unnatural self compatible pair of *bE-bW* alleles. A compensating second set of mutations must occur to remove and reintroduce an interfering residue in the corresponding subdomain of the partner protein, such that the mutant *bE* and *bW* proteins are no longer self compatible. Thus, a new allele bE-bW combination is created that is able to dimerize with the original parent allele but not with itself. These point mutations are likely to involve substitutions of less bulky residues for bulky ones or visa versa; substitutions such as Tyr to Cys, Tyr to His, or Val to Ala. The codons for these amino acids allow for these substitutions with only single base mutations. It is noteworthy that a comparison of multiple *b* alleles reveals a significant number of these substitutions between alleles (data not shown).

The second mechanism for the creation of new alleles requires an initial meiotic crossover between two different parental alleles (i.e. *b1* and *b2*) occurring in the region between the two specificity subdomains of one partner protein (e.g. *bE*). This essentially results in a class II chimeric allele that is compatible with both parental alleles. Again, this produces an unnatural self compatible pair of *bE-bW* alleles. Then, the other partner allele(e.g. *bW*) must also undergo a meiotic crossover in the region between the two specificity subdomains, producing a second class II chimeric alleles. This would create a new pair of partner *bE-bW* genes that are self incompatible because they are both class II chimeric alleles, but are compatible with both parental alleles.

The first mechanism described, called the mutation mechanism, would produce new *b* alleles that have high levels of identity with the original allele. This situation is found between *bW1* and *bW3* alleles, which have only nine amino acid differences in the first 150 codons. The second mechanism, called the recombination mechanism, would produce a new *b* allele that has a chimeric sequence. An inspection of several *bE* sequences presented in Kronstad and Leong (1990) suggests that *bE6* may have been created from a chimera of *bE3* positions 1-50 and an unidentified allele. The two alleles *bE6* and *bE3* are identical up to position 49 and then divergent from positions 50 to 110, with 30 amino acid differences between the two alleles over this latter portion of the variable region. The driving force behind the evolution of a multiallelic mating type system may be that a greater proportion of the population is likely to be compatible with any one individual. At the same time, genetic diversity is maintained by a self incompatibility.

Finally, some thought should be given as to the number of specificity subdomains on each of bE or bW proteins required to give approximately 25 multiple alleles. Why are there two or possibly more specificity subdomains rather than just one? It is likely that since a single specificity subdomain of approximately 10 to 20 amino acids has a limited number of contact residues, then two or more interacting subdomains are necessary to achieve the 25 or so combinations of interacting residues that mediate self recognition.

4.5 Steric Hindrance Between bE Position 90 and bW Position 74 May Interfere with Dimerization

The site-directed mutagenesis at bE position 90 and bW position 74 in a chimeric allele background clearly showed the negative effect on compatibility of opposing Val or Tyr side chains. The presence at both positions of Val or Tyr produced incompatibility; the replacement of one of the Val or Tyr with Ala (half the size of Tyr) allow a moderate compatible reaction, and the replacement of Ala at both positions gave a strong compatible reaction. A crucial point to make in the interpretation of these data is that the wild type sequences contain a Val at position 90 of bE1 and a Tyr at position 74 of bW1. These two side chains may interfere with the dimerization of

these self alleles through unfavorable steric interactions, thus providing at least a partial answer to the question "why does bE1 not dimerize with bW1?".

It is likely that more than one location of steric interference is required to prevent bE1 from dimerizing with bW1. Also, the removal of just one of the interference points is not sufficient to allow dimerization. A experiment mating two strains with mutant alleles, bE1(Val^{E90} to Ala) and bW1(Tyr^{W74} to Ala) did not produce a compatible mating reaction (data not shown). The other interfering side chain combinations may include bE1 Tyr^{E31} and bW1 Phe^{W4} or Tyr^{W6} (subdomains EN interacting with WN) and possibly bE1 Val^{E72} and bW1 Tyr^{W82} (inferred from Table 3-4).

The presence of several opposing bulky residues would disrupt the shape complementarity that is so critical for dimer formation. Without shape complementarity, many of the dimer stabilizing interactions are weakened considerably. Van der Waals interactions are effective only at very close range, H-bonds have a longer range but are very dependent upon orientation (linearity) of the H-bond, electrostatic interactions and hydrophobic interactions are disrupted by the entry of water molecules to the dimer interface that would result from disruption of shape complementarity. The importance of shape complementarity is well illustrated by the incredibly close conformation of many antibody-antigen interfaces (Davies et al. 1990). Even single amino acid changes resulting in unfavorable steric interactions, such as a Gly to Ser mutation in the 48G7 antibody (Wedemayer et al. 1997), and mutations to large volume side chains at position 101 of the antibody HEL-10 (Kam-Morgan et al. 1993) have significant negative effects on antibody binding.

4.6 The Inhibitory Determinant Model for Specificity at the *b* Locus

There are often two sides to the problem of specificity in protein-protein interactions. One side involves positive determinants of specificity such as shape complementarity and attractive electrostatic interactions, which actively bring two proteins together in a specific manner. But sometimes the specificity problem is solved by actively preventing unwanted combinations through "inhibitory determinants". This is often the case when the selection of protein partners are closely related by homology. Specificity is mediated by a process of elimination. A simple example of

this occurs with the Fos and Jun heterodimer, where negatively charged residues adjacent to the leucine zipper backbone of Fos destabilize the Fos homodimer, thus favoring the Fos-Jun heterodimer (O'Shea et al. 1992).

An inhibitory determinant model was proposed for hormone-receptor specificity, prompted by chimeric protein experiments with a family of homologous human reproductive hormones. The chimeras of human chorionic gonadotropin (hCG) and human follitropin (hFSH) were shown to exhibit activity unique to a third family member, human thyrotropin (hTSH) (Campbell et al. 1997). This result was explained by a model stating that specificity between ligand and receptor was mediated by "inhibitory determinants" that restricted binding to only the desired combinations (Campbell et al. 1997; Moyle et al. 1994). The formation of chimeric hormones was thought to disrupt the inhibitory determinants and unmask activities characteristic of other members of the protein family.

In an independent study of hormone-receptor specificity, Tian et. al. (1996) proposed a role for inhibitory domains in determining specificity of receptors for neurokinin peptides. Again, a set of chimeric proteins were constructed between two homologous receptor genes, NK_1 and NK_3. It was found that the affinites for several natural peptide ligands were higher in these chimeric receptors, than in the parental receptors. It was hypothesized that selectivity may have evolved in receptors through the acquiring of steric hindrance domains that would allow only the preferred ligands to bind. Presumably, the formation of chimeric receptors disrupted the inhibitory domains, resulting in higher affinities to certain ligands compared to the parental receptors.

The problem of specificity between bE and bW proteins is not unlike the hormone-receptor situation. Allelic chimeras (class II) of both bE and bW displayed a specificity different from each of the parental alleles. The nature of the "partner choice", where each b protein is able to dimerize with all 25 or more partners, except for one, strongly suggests that specificity is determined by interactions that actively prevent dimerization between the unwanted pair.

The genetic data in this study support a inhibitory determinant model for specificity between bE and bW proteins, whereby specificity is mediated by residues that prevent interactions

between self combinations. The inhibitory determinant model assumes that bE and bW partner proteins have an intrinsic ability to dimerize and that the dimerization interface is specific to the partner protein, but is largely nonspecific for the different alleles of the partner. It is likely that the interface between bE and bW has a high degree of shape complementarity and that the protein backbone of bE or bW is largely unchanged between alleles. Shape complementarity and backbone rigidity are general principles governing specificity in protein-protein interactions (Creighton 1993) and they are well illustrated in classic examples such as antibody-antigen binding (Davies et al. 1990; Lescar et al. 1995) . Any surface plasticity required to adapt to different allelic forms of bE or bW is likely accommodated by side chain mobility (Malby et al. 1994). Surface adaptability through side chain movement has been proposed as one of the mechanisms by which different cross reacting antibodies recognize the same epitope (Arevalo et al. 1993; Lescar et al. 1995). Also, the interfaces must be sufficiently hydrophobic, to favor dimer formation. There is likely to be polar groups involved in hydrogen bonding, and charged groups forming "salt bridges" on the dimer interface, most likely on the periphery.

The specificity occurs with bE and bW partners from the same allele. Here, the bE and bW proteins do not dimerize, likely because of steric (or electrostatic) interactions that prevent dimerization. Unfavorable steric interactions could arise from the positioning of bulky side chains (such as Val, Tyr, or Phe) in an opposed pattern on the dimer interface. The locations of these sterically interfering residues change for each allelic combination of bE and bW, such that nonself combinations of proteins experience a staggered pattern of interfering residues. The staggered arrangement allows dimerization. It is possible that inhibitory interactions also occur from electrostatic repulsion and unfavorable polar to nonpolar interactions on the dimer interface but these are less suited for a situation where specific interactions are required for only one partner out of a total of 25 or more. Electrostatic and polar to nonpolar interactions require specific parings of residues in order to favor, as well as inhibit dimerization. In contrast to this, bulky Val, Tyr, or Phe side chains can prevent dimerization in a specific manner when they are opposing one another

and favor dimerization in a nonspecific manner when they are staggered, simply due to their moderate hydrophobicity.

The data presented in this study has experimentally identified two amino acid residues, one on each of bE and bW, that function as inhibitory determinants of specificity. Many other positions have been identified that affect specificity, and some of these may prove in future studies, to function as inhibitory determinants also. The ability to manipulate *U. maydis* genetically, and the relative simplicity of the *b* locus, has facilitated these chimeric allele experiments, leading to valuable insight into the mechanism of specificity between the two proteins bE and bW. The lessons learned here, can undoubtedly be applied to the more complex but related mating type systems of *S. commune* and *C. cinereus*. Chimeric alleles of the *Y* mating type locus of *S. commune* have been shown to possess a novel specificity, different from either parental allele (Yue et al. 1997), suggesting that an inhibitory determinant model is likely to also apply to these recognition systems.

4.7 Future Directions

The genetic evidence presented here indicates that Val^{E90} and Tyr^{W74} interact as inhibitory determinants of specificity, preventing bE1 and bW1 proteins from forming dimers. There is likely more than one pair of interfering residues for bE1 and bW1. The evidence for this lies in the observation that two single codon mutations in wild type bE1 and bW1 backgrounds consisting of Val^{E90} to Ala and Tyr^{W74} to Ala, did not produce strains that were compatible with each other (data not shown). Other interfering pairs of residues between bE1 and bW1 are likely Tyr^{E31}-Tyr^{W6}, and possibly Val^{E72}-Tyr^{W82}. The former pair of possible inhibitory determinants were deduced from the interaction involving strains carrying chimeric alleles *bEx31, bEx39, bW2*, and *bWx9* (Table 3-15), and the latter pair were deduced from crosses between strains carrying chimeric alleles *bEx70, bEx79, bWx81,* and *bWx82*. (Table 3.4) Conceivably, site-directed mutagenesis of bE1 and bW1, replacing the two or three pairs of interfering residues with Ala, could allow these two

proteins to dimerize. This would further strengthen the inhibitory determinant model for specificity at the *b* locus.

The evidence reported in this study has employed a biological mating assay in *U. maydis* as a measure of interaction between bE and bW. The mating assay has been shown to correlate well with protein-protein interactions between the N-terminal domains of bE and bW by both *in vitro* and *in vivo* methods (Kamper et al. 1995; Romeis et al. 1997). It would be valuable to confirm the hypothesized dimerization effects for the data in this study through the use of additional assays for protein-protein interactions. Two-hybrid analysis (Fields and Song 1989) could be used to show the effect of specific amino acid substitutions on the strength of interaction between bE and bW.

As powerful as genetics can be, the solution to the specificity problem will not be complete without the three dimensional structures of bE and bW, both as monomers and as dimers. Structural data would verify the interactions between Val^{E90} and Tyr^{W74}, and the role of the numerous positions in bE and bW revealed by chimeric allele analysis. Complementary to this, the genetic data would provide valuable functional information to corroborate the structural data. Research to determine the structures of the *b* proteins is presently being carried out. Its success will provide a complete structure-function model for multiallelic recognition at the *b* locus of *Ustilago maydis* and also likely provide insight to mechanisms of specificity in other biological systems.

5. REFERENCES

Agrios, G.N. 1988. Plant pathology. Third Edition, Academic Press.

Alexopoulos, C.J., C.W. Mims, and M. Blackwell. 1996. Introductory mycology. Fourth Edition, John Wiley & Sons, Inc., Toronto, Canada.

Amati, B., M.W. Brooks, N. Levy, T.D. Littlewood, G.I. Evan, and H. Land. 1993. Oncogenic activity of the c-Myc protein requires dimerization with Max. *Cell* **72**: 233-245.

Anderson, D.E., W.J. Becktel, and F.W. Dahlquist. 1990. pH-induced denaturation of proteins: A single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry* **29**: 2403-2408.

Arevalo, J.H., M.J. Taussig, and I.A. Wilson. 1993. Molecular basis of crossreactivity and the limits of antibody-antigen complementarity. *Nature* **365**: 859-863.

Asante-Owusu, R.N., A.H. Banham, H.U. Bohnert, E.J.C. Mellor, and L.A. Casselton. 1996. Heterodimerization between two classes of homeodomain proteins in the mushroom *Coprinus cinereus* brings together potential DNA-binding and activation domains. *Gene* **172**: 25-31.

Aufauvre-Brown, A. 1993. Detection of gene disruption events in Aspergillus transformants by PCR direct from conidiospores. *Curr. Genet.* **24**.

Ausubel, F.M. 1995. Current protocols in Molecular Biology, John Wiley & Sons, Inc., New York.

Banham, A.H., R.N. Asante-Owusu, B. Gottgens, S.A.J. Thompson, C.S. Kingsnorth, J.C. Mellor, and L.A. Casselton. 1995. An N-terminal dimerization domain permits homeodomain proteins to choose compatible partners and initiate sexual development in the mushroom *Coprinus cinereus*. *Plant Cell* **7**: 773-783.

Banuett, F. and I. Herskowitz. 1988. Ustilago maydis, Smut of Maize. In Genetics of Phytopathological Fungi (ed. G.S. Sidhu), Advances in Plant Pathology, vol. 6, pp. 427-455. Academic Press, London.

Banuett, F. and I. Herskowitz. 1989. Different *a* alleles of *Ustilago maydis* are necessary for maintenance of filamentous growth but not for meiosis. *Proc. Natl. Acad. Sci. USA* **86**: 5878-5882.

Banuett, F. and I. Herskowitz. 1994. Identification of Fuz7, a *Ustilago maydis* MEK/MAPKK homolog required for *a*-locus-dependent and -independent steps in the fungal ife cycle. *Genes Dev.* **8**: 1367-1378.

Barnes, W.M. 1994. PCR amplification of up to 35-kb DNA with high fidelity and high yield from λ bacteriophage templates. *Proc. Natl. Acad. Sci. USA* **91**: 2216-2220.

Bhat, T.N., G.A. Bentley, G. Boulot, M.I. Greene, T. D., W. Dall'Acqua, H. Souchon, F.P. Schwarz, R.A. Mariuzza, and R.J. Poljak. 1994. Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl. Acad. Sci. USA* **91**: 1089-1093.

Bolker, M., M. Urban, and R. Kahmann. 1992. The *a* mating type locus of *U. maydis* specifies cell signaling components. *Cell* **68**: 441-450.

Burley, S.K. and G.A. Petsko. 1988. Weakly polar interactions in proteins. *Adnances in Protein Chemistry* **39**: 125-189.

Campbell, R.K., E.R. Bergert, Y. Wang, J.C. Morris, and W.R. Moyle. 1997. Chimeric proteins can exceed the sum of their parts: implications for evolution and protein design. *Nature Biotech.* **15**: 439-443.

Cease, K.B., C.A. Potcova, C.J. Lohff, and M.E. Zeigler. 1994. Optimized PCR using Vent polymerase. *PCR Methods and Applications* **3**: 298-300.

Cheng, S., C. Fockler, W.M. Barnes, and R. Higuchi. 1994. Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl. Acad. Sci. USA* **91**: 5695-5699.

Christensen, J.J. 1963. Corn smut caused by Ustilago maydis. American Phytopathological Society Monograph Number 2, 41 pp, American Phytopathological Society, St. Paul, MN.

Cohen, G.B., R. Ren, and D. Baltimore. 1995. Modular binding domains in signal transduction proteins. *Cell* **80**: 237-248.

Covitz, P.A., I. Herskowitz, and A.P. Mitchell. 1991. The yeast *RME1* gene encodes a putative zinc finger protein that is directly repressed by a1-α2. *Genes Dev.* **5**: 1982-1989.

Creighton, T.E. 1993. Proteins : structures and molecular properties. 2nd Edition, 507 pp, W. H. Freeman and Co., New York.

Dahl, M., M. Bolker, B. Gillissen, F. Schauwecker, B. Schroeer, and R. Kahmann. 1991. The *b* locus of *Ustilago maydis* : Molecular analysis of allele specificity. In Advances in Molecular Genetics of Plant-Microbe Interactions (ed. H. Hennecke and D.P.S. Verma), vol. 1, pp. 264-271. Kluwer Academic Publishers, Dordrecht, Netherlands.

Davies, D.R. and G.H. Cohen. 1996. Interactions of protein antigens with antibodies. *Proc. Natl. Acad. Sci. USA* **93**: 7-12.

Davies, D.R., E.A. Padlan, and S. Sheriff. 1990. Antibody-antigen complexes. *Annu. Rev. Biochem.* **59**: 439-473.

Day, P.R. and S.L. Anagnostakis. 1971. Corn smut dikaryon in culture. *Nat. New Biol.* **231**: 19-20.

Day, P.R., S.L. Anagnostakis, and J.E. Puhalla. 1971. Pathogenicity resulting from mutation at the *b* locus of *Ustilago maydis*. *Proc. Natl. Acad. Sci. USA* **68**: 533-535.

Dolan, J.W. and S. Fields. 1991. Cell-type-specific transcription in yeast. *Biochim. Biophys. Acta* **1088**: 155-169.

Dranginis, A.M. 1990. Binding of yeast a1 and α2 as a heterodimer to the operator DNA of a haploid-specific gene. *Nature* **347**: 682-685.

Drebot, M.A. and S.H.S. Lee. 1994. Direct sequencing of double-stranded PCR products gel purified by centrifugation through blotting paper. *Biotechniques* **17**: 248-250.

Dunbrack, R.L., Jr., D.L. Gerloff, M. Bower, X. Chen, O. Lichtarge, and F.E. Cohen. 1997. Meeting review: the second meeting on the critical assessment of techniques for protein structure prediction (CASP2), Asilomar, California, December 13-16, 1996. *Folding & Design* **2**: R27-R42.

Dyson, H.J. and P.E. Wright. 1993. Peptide conformation and protein folding. *Curr. Opin. Struct. Biol.* **3**: 60-65.

Ehtesham, N.Z. and S.E. Hasnain. 1991. Direct in-gel hybridization, without blotting, using nick-translated cloned DNA probe. *Biotechniques* **11**: 718-722.

Elder, R.T., E.Y. Loh, and R.W. Davis. 1993. RNA from the yeast transposible element Ty1 has both ends in the direct repeats, a structure similar to retrovirus RNA. *Proc. Natl. Acad. Sci. USA* **80**: 2432-2436.

Faux, M.C. and J.D. Scott. 1996. Molecular glue: kinase anchoring and scaffold proteins. *Cell* **85**: 9-12.

Fersht, A.R. and L. Serrano. 1993. Principles of protein stability derived from protein engineering experiments. *Curr. Opin. Struct. Biol.* **3**: 75-83.

Fields, S. and O.K. Song. 1989. A novel genetic system to detect protein-protein interactions. *Nature* **340**: 245-246.

Finer-Moore, J.S., E.B. Fauman, R.J. Morse, D.V. Santi, and R.M. Stroud. 1996. Contribution of a salt bridge to binding affinity and dUMP orientation to catalytic rate: mutation of a substrate-binding arginine in thymidylate synthase. *Protein Eng.* **9**: 69-75.

Fotheringham, S. and W.K. Holloman. 1990. Pathways of transformation in Ustilago maydis determined by DNA conformation. *Genetics* **124**: 833-843.

Froeliger, E.H. and S.A. Leong. 1991. The *a* mating-type alleles of *Ustilago maydis* are idiomorphs. *Gene* **100**: 113-122.

Frohman, M.A. and G.R. Martin. 1990. Detection of homologous recombinants. In PCR protocols: a guide to methods and applications (ed. M.A. Innis, D.H. Gelfand, J.J. Sninsky and T.J. White), pp. 228-236. Academic Press, Inc., San Diego, CA.

Geourjon, C. and G. Deleage. 1994. SOPM: a self-optimized method for protein secondary structure prediction. *Protein Eng.* **7**: 157-164.

Geourjon, C. and G. Deleage. 1995. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biol. Sci.* **11**: 681-684.

Gillissen, B., J. Bergemann, C. Sandmann, B. Schroeer, M. Bolker, and R. Kahmann. 1992. A two-component regulatory system for self/nonself recognition in Ustilago maydis. *Cell* **68**: 647-657.

Gritz, L. and J. Davies. 1983. Plasmid-encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in *Escherichia coli* and *Saccharomyces cerevisiae*. *Gene* **25**: 179-188.

Hanna, W.F. 1929. Studies in the physiology and cytology of *Ustilago zeae* and *Sorosporium relianum*. *Phytopathology* **19**: 415-443.

Herskowitz, I. 1995. MAP kinase pathways in yeast: For mating and more. *Cell* **80**: 187-197.

Ho, C.Y., J.G. Adamson, R.S. Hodges, and M. Smith. 1994. Heterodimerization of the yeast MATa1 and MATα2 proteins is mediated by two leucine zipper-like coiled-coil motifs. *EMBO J.* **13**: 1403-1413.

Holliday, R. 1961. Induced mitotic crossing-over in *Ustilago maydis*. *Genet. Res.* **2**: 231-248.

Holliday, R. 1962. Mutation and replication in *Ustilago maydis*. *Genet. Res.* **3**: 472-486.

Holliday, R. 1964. A mechanism for gene conversion in fungi. *Genet. Res.* **5**: 282-304.

Holliday, R. 1974. The Genetics of *Ustilago maydis*. In Handbook of Genetics (ed. R.C. King), vol. 1, pp. 575-595. Plenum Press, New York.

Hu, J.C., E.K. O'Shea, P.S. Kim, and R.T. Sauer. 1990. Sequence requirements for coiled-coils: analysis with λ repressor-GCN4 leucine zipper fusions. *Science* **250**: 1400-1403.

Janin, J. and C. Chothia. 1990. The structure of protein-protein recognition sites. *J. Biol. Chem.* **265**: 16027-16030.

Jones, S. and J.M. Thornton. 1996. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA* **93**: 13-20.

Kahmann, R. and M. Bolker. 1996. Self/nonself recognition in fungi: old mysteries and simple solutions. *Cell* **85**: 145-148.

Kalvakolanu, D.V.R. and W.H. Livingston. 1991. Rapid and inexpensive protocol for generating >95% recombinants in subcloning experiments. *Biotechniques* **10**: 176-177.

Kam-Morgan, L.N.W., S.J. Smith-Gill, M.G. Taylor, L. Zhang, A.C. Wilson, and J.F. Kirsch. 1993. High-resolution mapping of the HyHEL-10 epitope of chicken lysozyme by site-directed mutagenesis. *Proc. Natl. Acad. Sci. USA* **90**: 3958-3962.

Kamper, J., M. Reichmann, T. Romeis, M. Bolker, and R. Kahmann. 1995. Multiallelic recognition: nonself-dependent dimerization of the *bE* and *bW* homeodomain proteins in *Ustilago maydis*. *Cell* **81**: 73-83.

King, R.D. and M.J.E. Sternberg. 1996. Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Protein Sci.* **5**: 2298-2310.

Korn, A.P. and R.M. Burnett. 1991. Distribution and complementarity of hydropathy in multisubunit proteins. *Proteins Struct. Funct. Genet.* **9**: 37-55.

Kronstad, J.W. and S.A. Leong. 1989. Isolation of two alleles of the b locus of Ustilago maydis. *Proc. Natl. Acad. Sci. USA* **86**: 978-982.

Kronstad, J.W. and S.A. Leong. 1990. The *b* mating-type locus of *Ustilago maydis* contains variable and constant regions. *Genes Dev.* **4**: 1384-1395.

Kronstad, J.W., J. Wang, S.F. Covert, D.W. Holden, G.L. McKnight, and S.A. Leong. 1989. Isolation of metabolic genes and demonstration of gene disruption in the phytopathogenic fungus Ustilago maydis. *Gene* **79**: 97-106.

Kues, U. and L.A. Casselton. 1992. Homeodomains and regulation of sexual development in basidiomycetes. *Trends Genet.* **8**: 154-155.

Kues, U., B. Gottgens, R. Stratmann, W.V.J. Richardson, S.F. O'Shea, and L.A. Casselton. 1994a. A chimeric homeodomain protein causes self-compatibility and constitutive sexual development in the mushroom *Coprinus cinereus*. *EMBO J.* **13**: 4054-4059.

Kues, U., W.V.J. Richardson, A.M. Tymon, E.S. Mutasa, B. Gottgens, S. Gaubatz, A. Gregoriades, and L.A. Casselton. 1992. The combination of dissimilar alleles of the $A\alpha$ and $A\beta$ gene complexes, whose proteins contain homeo domain motifs, determines sexual development in the mushroom *Coprinus cinereus*. *Genes Dev.* **6**: 568-577.

Kues, U., A.M. Tymon, W.V.J. Richardson, G. May, P.T. Gieser, and L.A. Casselton. 1994b. *A* mating-type factors of *Coprinus cinereus* have variable numbers of specificity genes encoding two classes of homeodomain proteins. *Mol. Gen. Genet.* **245**: 45-52.

Kuhn, J. and Y. Parag. 1972. Protein-subunit aggregation model for self-incompatibility in higher fungi: *Schizophyllum commune*. *Theor. Biol.* **35**: 77-91.

Laughon, A. and M.P. Scott. 1984. Sequence of a Drosophila segmentation gene: protein structure homology with DNA-binding proteins. *Nature* **310**: 25-31.

Lee, C., B. Leung, M.A. Lemmon, J. Zheng, D. Cowburn, J. Kuriyan, and K. Saksela. 1995. A single amino acid in the SH3 domain of HcK determines its high affinity and specificity in binding to HIV-1 Nef protein. *EMBO J.* **14**: 5006-5015.

Lescar, J., M. Pellegrini, H. Souchon, D. Tello, R.J. Poljak, N. Peterson, M. Greene, and P.M. Alzari. 1995. Crystal structure of a cross-reaction complex between Fab F9.13.7 and guinea fowl lysozyme. *J. Biol. Chem.* **270**: 18067-18076.

Luo, Y., R.C. Ullrich, and C.P. Novotny. 1994. Only one of the paired *Schizophyllum commune* $A\alpha$ mating-type, putative homeobox genes encodes a homeodomain essential for $A\alpha$-regulated development. *Mol. Gen. Genet.* **244**: 318-324.

Magae, Y., C. Novotny, and R. Ullrich. 1995. Interaction of the $A\alpha$ *Y* and *Z* mating-type homeodomain proteins of *Schizophyllum commune* detected by the two-hybrid system. *Biochem. Biophys. Res. Commun.* **211**: 1071-1076.

Malby, R.L., W.R. Tulip, V.R. Harley, J.L. McKimm-Breschkin, W.G. Laver, R.G. Webster, and P.M. Colman. 1994. The structure of a complex between the NC10 antibody and influenza virus neuraminidase and comparison with the overlapping binding site of the NC41 antibody. *Structure* **2**: 733-746.

Marsh, J.L., M. Erfle, and E.J. Wykes. 1984. The pIC plasmid and phage vectors with versatile cloning sites for recombinant selection by insertional inactivation. *Gene* **32**: 481-485.

Marsh, L. 1991. Signal transduction during pheromone response in yeast. *Annu. Rev. Cell Biol.* **7**: 699-728.

Metzenberg, R.L. and N.L. Glass. 1990. Mating-type and mating-type strategies in *Neurospora*. *BioEssays* **12**: 53-59.

Moyle, W.R., R.K. Campbell, R.V. Myers, M.P. Bernard, Y. Han, and X. Wang. 1994. Co-evolution of ligand-receptor pairs. *Nature* **368**: 251-255.

Munson, M., S. Balasubramanian, K.G. Fleming, A.D. Nagi, R. O'Brien, J.M. Sturtevant, and L. Regan. 1996. What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties. *Protein Sci.* **5**: 1584-1593.

O'Shea, E.K., R. Rutkowski, and P.S. Kim. 1992. Mechanism of specificity in the Fos-Jun oncoprotein Heterodimer. *Cell* **68**: 699-708.

Orr-Weaver, T.L., J.W. Szostak, and R.J. Rothstein. 1981. Yeast transformation: a model system for the study of recombination. *Proc. Natl. Acad. Sci. USA* **78**: 6354-6358.

Padlan, E.A. 1990. On the nature of antibody combining sites: unusual structural features that may confer on these sites an enhanced capacity for binding ligands. *Proteins Struct. Funct. Genet.* **7**: 112-124.

Padlan, E.A., E.W. Silverton, S. Sheriff, G.H. Cohen, S.J. Smith-Gill, and D.R. Davies. 1989. Structure of an antibody-antigen complex: crystal structure of the HyHEL-10 Fab-lysozyme complex. *Proc. Natl. Acad. Sci. USA* **86**: 5938-5942.

Puhalla, J.E. 1968. Compatibility reactions on solid medium and interstrain inhibition in *Ustilago maydis*. *Genetics* **60**: 461-474.

Puhalla, J.E. 1969. The formation of diploids of *Ustilago maydis* on agar medium. *Phytopathology* **59**: 1771-1772.

Puhalla, J.E. 1970. Genetic studies of the *b* incompatability locus of *Ustilago maydis*. *Genet. Res.* **60**: 229-232.

Qian, L. and M. Wilkinson. 1991. DNA fragment purification: removal of agarose 10 minutes after electrophoresis. *Biotechniques* **10**: 736-738.

Rawn, J.D. 1989. Proteins, energy, and metabolism. Revised Edition, Neil Patterson Publishers, Burlington, NC.

Regenfelder, E., T. Spellig, A. Hartmann, S. Lauenstein, M. Bolker, and R. Kahmann. 1997. G proteins in *Ustilago maydis*: transmission of multiple signals? *EMBO J.* **16**: 1934-1942.

Ridgway, J.B.B., L.G. Presta, and P. Carter. 1996. 'Knobs-into-holes' engineering of antibody C_H3 domains for heavy chain heterodimerization. *Protein Eng.* **9**: 617-621.

Rini, J.M., U. Schulze-Gahmen, and I.A. Wilson. 1992. Structural evidence for induced fit as a mechanism for antibody-antigen recognition. *Science* **255**: 959-965.

Romeis, T., J. Kamper, and R. Kahmann. 1997. Single-chain fusions of two unrelated homeodomain proteins trigger pathogenicity in *Ustilago maydis*. *Proc. Natl. Acad. Sci. USA* **94**: 1230-1234.

Rost, B. and C. Sander. 1993. Improved prediction of protein secondary structure by use of sequence profiles and neural networks. *Proc. Natl. Acad. Sci. USA* **90**: 7558-7562.

Rost, B. and C. Sander. 1995. Progress of 1D protein structure prediction at last. *Proteins Struct. Funct. Genet.* **23**: 295-300.

Rost, B. and C. Sander. 1996. Bridging the protein sequence-structure gap by structure predictions. *Annu. Rev. Biophys. Biomol. Struct.* **25**: 113-136.

Rost, B., R. Schneider, and C. Sander. 1993. Progress in protein structure prediction? *Trends Biochem. Sci.* **18**: 120-123.

Rowell, J.B. 1955. Functional role of compatibility factors and an *In Vitro* test for sexual compatibility with haploid lines of *Ustilago zeae*. *Phytopathology* **45**: 370-374.

Rowell, J.B. and J.E. DeVay. 1954. Genetics of *Ustilago zeae* in relation to basic problems of pathogenicity. *Phytopathology* **44**: 356-362.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: a labortory manual, Cold Spring Harbor Laboratory Press, New York.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**: 5463-5476.

Sarkar, G. and S.S. Sommer. 1990. The megaprimer method of site-directed mutagenesis. *Biotechniques* **8**: 404-407.

Sathe, G.M. 1991. Use of PCR for rapid detection of gene insertions in whole yeast cells. *Nucleic Acids Res.* **19**: 4775.

Sauer, R.T., D.L. Smith, and A.D. Johnson. 1988. Flexibility of the yeast $\alpha 2$ repressor enables it to occupy the ends of its operator, leaving the centre free. *Genes Dev.* **2**: 807-816.

Schreiber, G. and A.R. Fersht. 1995. Energetics of protein-protein interactions: analysis of the barnase-barstar interface by single mutations and double mutant cycles. *J. Mol. Biol.* **248**: 478-486.

Schulz, B., F. Banuett, M. Dahl, R. Schlesinger, W. Schafer, T. Martin, I. Herskowitz, and R. Kahmann. 1990. The *b* alleles of *U. maydis*, whose combinations program pathogenic development, code for polypeptides containing a homeodomain-related motif. *Cell* **60**: 295-306.

Serrano, L., M. Bycroft, and A.R. Fersht. 1991. Aromatic-aromatic interactions and protein stability: investigation by double-mutant cycles. *J. Mol. Biol.* **218**: 465-475.

Shen, G.P., D.C. Park, R.C. Ullrich, and C.P. Novotny. 1996. Cloning and characterization of a *Schizophyllum* gene with *Aβ6* mating-type activity. *Curr. Genet.* **29**: 136-142.

Shepherd, J.C.W., W. McGinnis, A.E. Carrasco, E.M. De Roberts, and W. Gehring. 1984. Fly and from homoeo domains show homologies with yeast mating type regulatory proteins. *Nature* **310**: 70-71.

Silva, J. 1972. Alleles at the *b* incompatibility locus in Polish and North American populations of *Ustilago maydis* (DC) Corda. *Physiol. Plant Pathol.* **2**: 333-337.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**: 503-517.

Spassov, V.Z., A.D. Karshikoff, and R. Ladenstein. 1995. The optimization of protein-solvent interactions: Thermostability and the role of hydrophobic and electrostatic interactions. *Protein Sci.* **4**: 1516-1527.

Specht, C.A., A. Munoz-Rivas, C.P. Novotny, and R.C. Ullrich. 1988. Transformation of *Schizophyllum commune:* an analysis of parameters for improving transformation freqencies. *Exp. Mycol.* **12**: 357-366.

Specht, C.A., M.M. Stankis, L. Giasson, C.P. Novotny, and R.C. Ullrich. 1992. Functional analysis of the homeodomain-related proteins of the *Aα* locus of *Schizophyllum commune*. *Proc. Natl. Acad. Sci. USA* **89**: 7174-7178.

Stakman, E.C. and J.J. Christensen. 1927. Heterothallism in *Ustilago zeae*. *Phytopathology* **17**: 827-834.

Stankis, M.M., C.A. Specht, and L. Giasson. 1990. Sexual incompatibility in *Schizophyllum commune*: from classical genetics to a molecular view. In Seminars in Developmental Biology (ed. C.A. Raper and D.I. Johnson), vol. 1, pp. 195-206. Saunders Scientific Publications, London.

Stankis, M.M., C.A. Specht, H. Yang, L. Giasson, R.C. Ullrich, and C.P. Novotny. 1992. The *Aα* mating locus of *Schizophyllum commune* encodes two dissimilar multiallelic homeodomain proteins. *Proc. Natl. Acad. Sci. USA* **89**: 7169-7173.

Tian, Y., L. Wu, D.L. Oxender, and F. Chung. 1996. The unpredicted high affinities of a large number of naturally occurring tachykinins for chimeric NK_1/NK_3 receptors suggest a role for an inhibitory domain in determining receptor specificity. *J. Biol. Chem.* **271**: 20250-20257.

Tsukuda, T., S. Carleton, S. Fotheringham, and W.K. Holloman. 1988. Isolation and characterization of an autonomously replicating sequence from Ustilago maydis. *Mol. Cell. Biol.* **8**: 3703-3709.

Tsumoto, K., K. Ogasahara, Y. Ueda, K. Watanabe, K. Yutani, and I. Kumagai. 1995. Role of Tyr residues in the contact region of anti-lysozyme monoclonal antibody HyHEL10 for antigen binding. *J. Biol. Chem.* **270**: 18551-18557.

Waldburger, C.D., J.F. Schildbach, and R.T. Sauer. 1995. Are buried salt bridges important for protein stability and conformational specificity? *Struct. Biol.* **2**: 122-128.

Wang, J., D.W. Holden, and S.A. Leong. 1988. Gene transfer system for the phytopathogenic fungus Ustilago maydis. *Proc. Natl. Acad. Sci. USA* **85**: 865-869.

Wedemayer, G.J., P.A. Patten, L.H. Wang, P.G. Schultz, and R.C. Stevens. 1997. Structural insights into the evolution of an antibody combining site. *Science* **276**: 1665-1669.

Wells, J.A. 1996. Binding in the growth hormone receptor complex. *Proc. Natl. Acad. Sci. USA* **93**: 1-6.

Wendland, J., L.J. Vaillancourt, J. Hegner, K.B. Lengeler, K.J. Laddison, C.A. Specht, C.A. Raper, and E. Kothe. 1995. The mating-type locus *Bα1* of Schizolphyllum *commune* contains a pheromone receptor gene and putative pheromone genes. *EMBO J.* **14**: 5271-5278.

Weng, Z., S. Vajda, and C. Delisi. 1996. Prediction of protein complexes using empirical free energy functions. *Protein Sci.* **5**: 614-626.

Wilson, I.A. and R.L. Stanfield. 1993. Antibody-antigen interactions. *Curr. Opin. Struct. Biol.* **3**: 113-118.

Wolberger, C., A.K. Vershon, B. Liu, A.D. Johnson, and C.O. Pabo. 1991. Crystal structure of a MATα2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell* **67**: 517-528.

Wu, J., R.C. Ullrich, and C.P. Novotny. 1996. Regions in the *Z5* mating gene of *Schizophyllum commune* involved in Y-Z binding and recognition. *Mol. Gen. Genet.* **252**: 739-745.

Yee, A.R. and J.W. Kronstad. 1993. Construction of chimeric alleles with altered specificity at the *b* incompatibility locus of *Ustilago maydis*. *Proc. Natl. Acad. Sci. USA* **90**: 664-668.

Young, L., R.L. Jernigan, and D.G. Covell. 1994. A role for surface hydrophobicity in protein-protein recognition. *Protein Sci.* **3**: 717-729.

Yue, C., M. Osier, C.P. Novotny, and R.C. Ullrich. 1997. The specificity determinant of the Y mating-type proteins of *Schizophyllum commune* is also essential for Y-Z protein binding. *Genetics* **145**: 253-260.

Zhou, C., Y. Yang, and J. Jong. 1990. Mini-prep in 10 minutes. *Biotechniques* **8**: 172-173.

6. APPENDICES

6.1 Media for *Ustilago maydis*

PDB (Potato Dextrose Broth)

	For 1 liter
distilled water	1000 ml
potato dextrose broth (Difco)	24 g

PDA (Potato Dextrose Agar) - Solid media for fungi

	For 1 liter
distilled water	1000 ml
potato dextrose broth (Difco)	24 g
agar	15 g

CM (Complete Media)

	For 1 liter
distilled water	800 ml
casamino acids (Difco)	2.5 g
ammonium nitrate	1.25g
yeast extract (Difco)	10
Ustilago maydis salt solution	62.5 ml

Dissolve ingredients and make volume to 500 ml; adjust pH to 7.0 with dropwise 10 N NaOH; autoclave; add 10 ml of sterile 50% glucose after autoclaving (glucose added before autoclaving will tend to caramelize). For solid media, add agar at 15 g per liter of liquid medium before autoclaving.

Charcoal Mating Medium

	For 1 liter
Casamino acids	10 grams
Ammonium nitrate	3 grams
Yeast extract	20 grams
Salt solution(see below)	125 ml

Dissolve in 800 ml H_2O; adjust pH to 7.0 with dropwise 10 N NaOH and bring volume to 1 liter.
Distribute 500 ml into each of two 1 liter flasks and add 5 grams of activated charcoal and 10 grams of agar to each flask. Mix by swirling; cover with foil and autoclave. Then add sterile glucose to 1% after autoclaving.

Mix well before and during pouring of plates by swirling. Pour the charcoal media until it covers about 75% of the plate, then swirl the plate to level the media. This produces a thinner layer of agar. Also, while the plates are solidifying, set them out individually, rather in stacks. These measures allow the charcoal agar to solidify quickly, without any settling of the charcoal, producing more consistent mating reactions.

Ustilago maydis salt solution

	For 1 liter
KH_2PO_4	16 g
Na_2SO_4	4 g
KCl	8 g
$MgSO_4 \cdot 7H_2O$	2 g
$CaCl_2 \cdot 2H_2O$	1 g
Trace Element Solution (see below)	8 ml

Add to 900 ml distilled water and make to 1 liter.

Ustilago maydis Trace Element Solution

	For 500 ml
H_3BO_3	30 mg
$MnCl_2 \cdot 4H_2O$	70 mg
$ZnCl_2$	200 mg
$Na_2Mo_4 \cdot 2H_2O$	20 mg
$FeCl_3 \cdot 6H_2O$	50 mg
$CuSO_4 \cdot 5H_2O$	200 mg

6.2 Transformation of *Ustilago maydis* - Modified Wang Protocol

A. Preparation of Competent Protoplasts

Streak a *U. maydis* strain on PDA and grow for 2-3 days. Better transformation results are obtained if the strain has not been in 4°C storage for extended periods (> 2 months).

1. Inoculate 5 ml of PDB with a small amount of cells (about 2 mm diameter) from PDA plate. Do this late in the day so that the culture is not too thick the next day. Incubate in an orbital shaker at 30°C and 225 rpm for about 20 hr until an OD_{600} of about 2.

2. Inoculate 100 ml of CM with 1% glucose in a 1L flask with 30 ul (for 001) to 60 ul (for 031) of U. maydis culture (OD_{600} about 2). This should be done late in the day. Grow overnight at 30°C and 225 rpm for 18 to 20 hr until OD_{600} is 0.9 to 1.0 (about to $5x10^7$ cells/ml for 001; $3x10^7$ cells/ml for 031). Under these conditions, doubling time is about 2 hours. The cells must be in early log phase and actively budding in order to achieve high competency.

3. Pour the 100 ml culture into two 50 ml Falcon tubes. Centrifuge for 5 min at about 3000 rpm (1500 x g; about setting 75 on IEC Centra 4B table top centrifuge); decant.

4. Resuspend the pellets in each tube in 20 ml of fresh 5 mM EDTA and 25 mM β-mercaptoethanol by vortexing on high.

   For 40 ml of this: 40 ml of sdH_2O
   400 ul of 0.5 M EDTA pH 8
   70 ul β-mercaptoethanol

5. Incubate at room temperature with gentle shaking at about 60 rpm for 20 min..

6. Centrifuge the cells again at 3000 rpm (1500 x g) for 5 min and resuspended by vortexing in 10 ml of Buffer I (1 M Sorbitol, 50 mM sodium citrate, pH 5.8). Then add 100 ul of filter sterilized Lysing Enzyme (100 mg/ml in distill water; final concentration of 1 mg/ml Sigma L-2265) to the suspension and mix by swirling. Start a timer now. Then take a 10 ul sample and prepare a slide to monitor protoplasting.

7. Incubate cells at room temperature with shaking (60 rpm) while protoplasting is monitored with the microscope. Within 5-10 minutes, the ends of the rod shaped cells of *U. maydis* should begin to balloon out. When this happens place cells in centrifuge and centrifuge at 600 rpm (60 x g; about setting 35, on IEC Centra 4B table top centrifuge)) for 10 minutes. If pelleting is done at 700 rpm or higher the cells are difficult to resuspend. For 001, centrifuge the cells when about 50% of them show balloon ends (transformation frequencies drop a bit if cells are left too long). Protoplasting continues during the centrifugation so that >90% of the cells become protoplasted by the end of the spin.

8. Decant the supernatant; resuspend the pellet in the residual supernatant by shaking tube back and forth. This should break up pellet and form a thick suspension. Then add 10 ml of Buffer I. The suspension may be clumpy. Centrifuge cells at 600 rpm for 10 min.

9. Decant and resuspend in residual supernatant as before. Then add 10 ml Buffer II (1 M Sorbitol, 25 mM Tris-HCl pH 7.5, 50 mM $CaCl_2$). If suspension is really clumpy, then resuspension can be done by gentle pipeting up and down with a disposable 10 ml pipet but do not overtreat the cells (reusable glass pipets may carry some residual soap or contaminants). Recentrifuge at 600 rpm for 10 min.

10. Decant and resuspend in residual supernatant as before. Then add 2 ml Buffer II. If suspension is clumpy, then resuspend by gentle pipeting with a P1000. Perform a cell count with a hemacytometer on a 1:10 dilution in Buffer II. Then, dilute the cells to $2x10^8$ cells/ml with Buffer II. Typical yield is 3 to 5 ml of $2x10^8$ cells/ml for each 50 ml of starting culture.

11. For each ml of protoplasts add in the following order:
    250 ul 50% PEG 3350 (Sigma P-3640) with 25 mM Tris 7.5 and 50 mM $CaCl_2$
    60 ul DMSO (Sigma D-8418)
    10 ul β-mercaptoethanol (Bio-Rad Cat. No. 161-0710)

    This protoplast mixture can be stored at 0°C for up to 24 hr without loss of transformation frequency or stored at -70°C with minimal loss of transformation frequency. The protoplasts remain competent for up to 12 months at -70°C. Aliquot the final protoplast mixture into 1 ml portions.

B. DNA Addition (double scale transformation volumes are given in brackets)

1. Thaw the protoplast mixture on ice. In a microfuge tube add sdH_2O, 30 (60) ug of heparin (use 2 (4) ul of 15 ug/ul heparin in sdH_2O), and 1-5 ug (2-10 ug) of transforming DNA so that the final volume of mixture is about 20 (40) ul. Place tube on ice for 5 min.

2. Add 125 (250) ul of above protoplast mixture to the DNA mixture, mix by flicking tube, incubate on ice for 10 min.

3. Add 100 (200) ul of room temperature 50% PEG + Tris + $CaCl_2$. Mix by inversion. Incubate at room temperature for 20 min.

4. Add 1 ml of room temperature Buffer II. Centrifuge 3500 rpm on Eppendorf Centrifuge 5415 (1000 x g) for 5 to 10 min. Gently aspirate supernatant.

5. Add 1 ml Buffer II again. No need to resuspend. Centrifuge 2500 to 3000 rpm in an Eppendorf 5415 (500 to 700 x g) for 5 min. Aspirate.

6. Add 200 (400) ul DCM with 1% glucose and 1 M sorbitol. A regeneration incubation is unnecessary.

7. Resuspend by gently pipeting up and down with a P1000. Plate entire 200 ul on one transformation plate. The transformation plates are 20 ml DCM agar with 1% glucose, 1 M sorbitol and 250 ug/ml of Hygromycin B (Calbiochem, La Jolla, CA), . Two layered plates are not necessary for standard strains.

    Transformants are visible within 3-4 days. Integrative transformation produces large and small colonies; some of the small colonies are not stable transformants. Pick the transformants onto PDA containing 50 ug/ml Hygromycin B to test their stability They will store on PDA at 4°C for 1-2 months.

Buffers and Media for Modified Wang Protocol

Buffer I (50 mM sodium citrate and 1 M sorbitol)

To make 200 ml:
to 150 ml of distilled water:
sodium citrate (trisodium salt of citric acid) 2.9 g
Sorbitol (Sigma S-1876) 36.4 g

adjust pH to 5.8 with conc. HCl, about 250 ul
adjust to 200 ml; autoclave.

Buffer II (25 mM Tris pH 7.5 + 50 mM $CaCl_2$ + 1 M sorbitol)

To make 200 ml:
Add to 150 ml of distilled water,
Sorbitol (Sigma S-1876) 36.4 g

make to 185 ml; autoclave; cool; then add sterily
1 M Tris pH 7.5 5 ml
1 M $CaCl_2$ 10 ml

50% PEG Solution

The 50% PEG + 25 mM Tris + 50 mM $CaCl_2$ solution should be made fresh ;
For 10 ml, add to a 25 ml graduated cylinder;
5 g PEG 3350 (Sigma P-3640)
5 ml dH_2O
Cover with foil, autoclave, cool to room temperature; then add sterile
250 ul 1 M Tris pH 7.5
500 ul 1 M $CaCl_2$
Make to 10 ml with about 1 ml of sdH_2O, mix by pipeting.

Transformation Plates -DCM with 1M Sorbitol

	500 ml	1000 ml
sdH_2O	350 ml	700 ml
Casamino acids (Difco)	5 g	10 g
Ammonium nitrate	1.5 g	3 g
Yeast extract (Difco)	10 g	20 g
U. maydis salt solution	62 ml	125 ml
sorbitol	91 g	182 g
pH to 7 with 10 N NaOH	800 ul	1600 ul

Make to volume, distribute to 500 ml portions, add agar (7.5 g/500 ml), autoclave.
Add glucose sterily after autoclaving to 1%.

Comments

All chemicals are analytical grade or better.

The expected transformation frequency for integrative DNA is 50 to 200 stable transformants from 1-5 ug of linear DNA using this protocol. The percentage of homologous integrants can vary from 1 to 50% (typically 20%) depending upon (1) the degree of homology between the target locus and the transforming DNA, (2) the age of the protoplasts, (3) the media used for transformation plates. The highest frequency for homologous integration is obtained when there is at least 500 bp of homologous DNA at each end of the transforming DNA, protoplasts are freshly made or less than 2 weeks old, and rich transformation media is used, such as DCM with 2% glucose.

The centrifuge speeds given above have been worked out for haploid strain 001. For diploid like d410, the centrifuge speeds can be slightly slower because the cells are larger.

The PEG+Tris+$CaCl_2$ should be made fresh and it is critical that it be autoclaved without the Tris and $CaCl_2$. For best results, Buffers I and II should be less than 2 weeks old.

The appearance of the protoplasts under the microscope will give an indication of how well the protoplasting has been carried out. If several unprotoplasted cells are visible and the protoplasts look bumpy and rough, then protoplasting was inadequate. The cells are clumpier and harder to resuspend in this case. Transformation frequencies will probably be lower. If the protoplasts look round and smooth, then protoplasting went well.

The protoplasts may undergo freezing and -70°C storage better with PEG in the freezing buffer (but this has not been thoroughly tested).

Re: DNA Addition: A small amount of osmotic shock to the protoplasts improves the transformation frequencies. A DNA-Heparin volume of 12 to 55 ul works well when using 125 ul of protoplast mix (10 to 50 % of protoplast volume). Transformation frequencies drop if DNA-heparin is added in only 5 ul for 125 ul of protoplasts. The DNA should not be diluted with an osmotic buffer; use sdH_2O or TE. Heparin does improve transformation.

The transforming DNA must be relatively clean for successful transformation for integrative vectors; miniprep plasmid from *E. coli* will give poor results unless it is purified by Phenol-Chloroform-Ethanol precipitation, Qiagen columns or equivalent. DNA that has been restriction digested must be Phenol-Chloroform-Ethanol precipitated or Gene Cleaned before transformation. For replicative transformation (ARS vector) miniprep plasmid will work and give 10-100 transformants/ug DNA; boiling lysis minipreps work better than alkaline-SDS lysis because SDS is damaging to the protoplasts.

The Buffer II has been modified from the original protocol by increasing the $CaCl_2$ concentration from 25 to 50 mM. Other fungal transformation protocols have shown that this is close to optimal for protoplast-PEG-$CaCl_2$ transformation a (Specht et al. 1988: Judelson and Michelmore 1991).

References

This protocol has been adapted from Wang et al. (1988) and Specht et al. (1988).

Specht, C. A. et al. 1988. Transformation of *Schizophyllum commune*: An analysis of parameters for improving transformation frequencies. Exp. Mycol. 12:357-366.

Wang, J. et al. 1988. Gene transfer system for phytopathogenic fungus *U. maydis*. Proc. Natl. Acad. Sci. USA 85:865-869.

6.3 Abbreviations and Conventions

Gene designations are in italics, (e.g. *bE*) and corresponding proteins are designated by nonitalics.

Amino acid residues are denoted by their three letter or one letter code, with their position in a protein given in superscript, e.g. Val^{E90} denotes valine at bE position 90; Tyr^{W74} denotes tyrosine at bW position 74.

Amino Acid	3 Letter Code	One Letter Code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Asn or Asp	Asx	B
Cysteine	Cys	C
Glutamine	Glu	Q
Glutamic acid	Glu	E
Gln or Glu	Glx	Z
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lys	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Trytophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

ARS - autonomously replicating sequence
bp - basepair
kb - kilobasepair
DMSO - dimethyl sulfoxide
EDTA - ethylene-diamine-tetra-acidtic acid
OD - optical density
ORF - open reading frame
PCR - polymerase chain reaction
PEG - polyethylene glycol
TE - 10 mM Tris pH 8.0 and 1 mM EDTA pH 8.0
U - units
UV - ultraviolet