Crystal Structure of Paired Domain - DNA Complex

by

Wenqing Xu

B.S., Biology
University of Science and Technology of China, 1985
M.S., Biochemistry and Crystallography
Chinese Academy of Science, 1988

Submitted to the Department of Biology
in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

at the
Massachusetts Institute of Technology

July 1995

Signature of Author.....
Department of Biology, July 18, 1995

Certified by.....
Carl O. Pabo, Professor of Biophysics and Structural Biology
Thesis Supervisor

Accepted by.....
Frank Solomon, Professor of Biology
Chairman, Biology Graduate Committee

CRYSTAL STRUCTURE OF PAIRED DOMAIN - DNA COMPLEX

by
Wenqing Xu

submitted to the Department of Biology in partial fulfillment of the requirements for the degree of Doctor of Philosophy

ABSTRACT

This thesis describes the determination of a paired domain-DNA complex crystal structure (involving the paired domain of the Drosophila Prd protein), and discusses the structural basis of DNA binding specificity of the paired domain and the structural basis of *Pax* developmental mutations. It also describes the co-crystallization of the human PAX6 paired domain-DNA complex.

Chapter 1 provides an introduction to paired domains and the Pax family. *Pax* genes play very important roles for vertebrate development. Mutations in several *Pax* genes have been associated with mouse and human congenital disorders. The paired domain, a highly conserved DNA-binding domain, is critical for Pax protein function.

Chapter 2 describes the purification of Drosophila Prd paired domain, the crystallization of the Prd paired domain-DNA complex, and the determination of the crystal structure of this complex.

Chapter 3 describes the structure of the Prd paired domain - DNA complex. The crystal structure shows that the paired domain folds as two independent sub-domains, each containing a helical structure that is very similar to the homeodomain. The N-terminal domain makes extensive DNA contacts. It has a novel β-turn motif that fits in the minor groove and a HTH unit that contacts the major groove. The β-turn makes base specific contacts in the minor groove, and is critical for both DNA binding and for Pax *in vivo* function. The HTH unit folds like a homeodomain but docks on DNA like λ repressor. The C-terminal domain of the Prd paired domain does not contact the optimized DNA binding site, and other experiments have shown that it is not required for DNA recognition.

Most *Pax* developmental mutations are found at the protein-DNA interface. This chapter was published as "Crystal Structure of a Paired Domain-DNA Complex at 2.5 Å Resolution Reveals Structural Basis for Pax Developmental Mutations" (Xu, W., Rould, M. A., Jun, S., Desplan, C. and Pabo, C. O. (1995). Cell 80, 639-650).

Chapter 4 further discusses the structural basis of paired domain DNA-binding specificity and *Pax* developmental mutations.

Chapter 5 describes the purification of PAX6 paired domain and the cocrystallization trials of PAX6 paired domain-DNA complex. Several promising cocrystal forms have been obtained.

Thesis supervisor: Professor Carl O. Pabo

To my parents and my wife

ACKNOWLEDGMENT

The work presented in this thesis relied on help and support from many people. I would like to thank my thesis supervisor, Dr. Carl O. Pabo, for his support, advice, patience and generosity. During the six years I have spent in his lab, Carl has been an excellent teacher and a mentor for my development as a scientist. His example as a scientist and his leadership made his laboratory a very exciting place to work. I have appreciated the opportunity to work here.

I wish to thank members of my thesis committee, Dr. Alexander Rich, Dr. Robert Sauer, Dr. Richard Maas and Dr. Stephen Bell, who offered me advice and encouragement while I worked on this project.

My collaborators Dr. Claude Desplan (on the Prd structure, Rockefeller University) and Dr. Richard Maas (on the PAX6 project, Harvard Medical School) were my constant sources of advice and scientific insight on the biology of the paired domain and Pax family.

I am especially grateful to Dr. Mark Rould. We worked together closely on solving the Prd structure. He patiently taught me the techniques for solving the structure, and was a constant source of help for interpreting the structure and learning crystallography.

Susie Jun in Dr. Desplan's lab was my collaborator in solving the Prd structure. Her unpublished results on the roles of paired C-terminal domain were very helpful for interpreting our Prd structure. Jonathan Epstein, my collaborator in Dr. Maas' group, was a constant source of help and comments. Guojun Sheng in Dr. Desplan's group deserves my special thanks. It was the discussion with him that led me to the Pax field.

Within the Pabo lab, I have benefited from the help and expertise of many members. I owe a lot to my baymate Ernest

Fraenkel. He gave invaluable comments for many of my English writings. Lena Nekludova has been very helpful in analysing the DNA structure and making graphics images. Cindy Limb purified many DNA oligomers that I used for PAX6 cocrystallization. Monicia Elrod-Erikson took care of my troublesome vacuum pumps many times. I must offer thanks to Juli Klemm, Kristen Chambers, Eric Xu, Beishan Liu, Harvey Greisman, Edward Rebar, Lisa Tucker-Kellogg, Philip Ma, Jin-Soo Kim, Cynthia Wolberger, Neil Clarke and Chuck Kissinger, for putting up with me, for many stimulating discussions and many other matters. My former baymate, Nikola Pavletich, was a source of inspiration for me, especially in his scientific intensity. I appreciated the many kinds of help from Amy Dunn, Kristine Kelly and Kathleen Kolish.

Finally, I thank my wife, Hongkui Zeng, for her love, support and understanding.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Note: Legends of figures and tables are at the end of each chapter.

Chapter 5

Figure 1: Flow chart for PAX6-PD purification.

Figure 2: Photographic image of the PAX6-PD and DNA co-crystals.

Chapter 1

Paired Domain and Pax Family

DNA-binding Protein Families

DNA-binding proteins are critical for many biological processes, such as transcriptional regulation, DNA recombination, genome replication, repairing damaged DNA, and responding to environment signals. Transcription factors that regulate gene expression comprise one of the largest and most diverse classes of DNA-binding proteins. Among other fields, transcription factors play central roles in the field of development biology --- regulating cell development, differentiation, and cell growth, by binding to specific DNA sites and thereafter activating or inhibiting gene expression.

One of the most important observations of the DNA-binding protein studies is that most DNA-binding proteins can be grouped into classes that use structurally related DNA-binding domains or motifs. Some families, such as the helix-turn-helix family, were recognized by structural similarities. More families were first identified by sequence comparisons and later characterized by structural studies. Some of the largest families include helix-turn-helix proteins, zinc-finger proteins, homeodomain-containing proteins, helix-loop-helix proteins, and leucine-zipper proteins. Structural and recognition aspects of transcription factor families were review by Pabo and Sauer (1992), and Harrison (1991) - more references can be found therein. Structural studies with one family member can usually provide basic information for the whole family.

Cloning and Characterization of *Pax* Genes

The 384 bp long paired box was first identified in three Drosophila segmentation genes *paired (prd), gooseberry (gsb)* and *gooseberry neuro (gsb-n)* (Bopp et al., 1986; Baumgartner et al., 1987), and subsequently in two tissue-specific genes, *Pox meso* and *Pox neuro* (Bopp et al., 1989). Paired boxes have been detected in such divergent organisms as mouse, human, nematode, zebra fish,

frog, turtle, and chicken, and very recently in C. elegans (Deutsch et al., 1988; Dressler et al., 1988; Burri et al., 1989; Walther et al. 1991; Martin et al. 1992; Krauss et al., 1991; Stapleton et al., 1993; Wallin et al., 1993; Chisholm and Horvitz, submitted). So far at least 30-40 paired-box genes have been cloned based on the sequence homology in the paired-box, including 9 murine *Pax* genes (*Pax-1* to *Pax-9*) and 9 human *PAX* genes (*PAX1* to *PAX9*), where *Pax* refers to paired-box-containing genes.

Unlike the developmental regulatory homeobox (*Hox*) genes, which were found clustered on particular chromosomes, each of nine human PAX genes is located on an entirely different chromosome. The most important clue leading to our current understanding of *Pax* biology was the association between Pax genes and several previously known mouse and human developmental phenotypes. For example, mutations in human *PAX3* and *PAX6* genes were found to be responsible for Waardenburg syndrome type 1 and type 3 (Tassabehji et al., 1992; Baldwin et al., 1992; Farrer et al., 1994) and aniridia (Ton et al., 1991; Glaser et al., 1992, 1994), respectively. Mutations in the mouse *Pax-1*, *Pax-3* and *Pax-6* genes are associated with undulated, Splotch, and Small eye phenotypes, respectively (Balling et al., 1988; Epstein et al., 1992; Hill et al., 1991).

The 128 amino acid paired domain encoded by the paired-box is the only region common to all PAX proteins. The DNA binding activity of paired domain was first demonstrated between the Drosophila paired protein (Prd) and the e5 DNA sequence in the even-skipped promoter (Treisman et al., 1991; Chalepakis et al., 1991). All Pax protein showed specific binding to this e5 sequence, and thus it has been used to study Pax protein-DNA interactions. The inference that Pax proteins act as transcription factors is based on their being localized in the nucleus (Dressler and Douglass, 1992; Glaser et al., 1995) and the presence of DNA-binding domain. This has been verified for Pax-5, Pax-6 and Pax-8, which have been shown to regulate cell type-specific gene transcription (Pax-5: Barberis et al., 1989; Kozmik et al., 1992; Waters et al., 1989;

Rothman et al., 1991; Williams and Maziels, 1991; Liao et al., 1992; Pax-8: Zannini, 1992; Pax-6: Cvekl et al., 1994, 1995; Chalepakis et al., 1994b; Richardson et al., 1995; Plaza et al., 1995; see figure 4 of Chapter4). In addition, the Pax-1, Pax-2, Pax-5, Pax-6 and Pax-8 proteins have been shown to activate reporter gene expression upon binding to modified e5 sites in transfection experiments (Czerny et al., 1993; Fickenscher et al., 1993; Kozmik et al., 1993; Zannini et al., 1992). Recently, it has also been shown that Pax-3 contains domains for both transcriptional activation and transcriptional inhibition (Chalepakis et al., 1994; Czerny and Busslinger, 1995).

Pax Gene Structure and Classification

Although *Pax* genes are operationally defined by the presence of a paired domain, they also share overall structural features. *Pax* genes were grouped into at least four subfamilies (Figure 1), initially based on the degree of homology in the paired domain, in conjunction with subfamily-specific amino acids at certain positions of the paired domain (Walther et al., 1991; Figure 1a of Chapter 3). This grouping is consistent with a classification based on the presence or absence of three structural features: 1) a characteristic octapeptide sequence (OP in Figure 1); 2) an intact paired-type homeodomain (HD); or 3) a partial paired-type homeodomain containing only the N-terminal arm and first helix (Hill and Hanson, 1992). The first *Pax* subfamily, which includes *Pax-1* and *Pax-9*, encodes the paired domain and a conserved octapeptide sequence but lacks a homeodomain. The second subfamily consists of *Pax-3* and *Pax-7* and, in addition to the paired domain and octapeptide, also encodes a full-length paired-type homeodomain. Drosophila *paired* and *gooseberry* genes also belong to this subfamily. The third class, represented by *Pax-2*, *Pax-5* and *Pax-8*, encodes paired domain, octapeptide and a partial homeodomain . *Pax-4* and *Pax-6* represent the fourth subfamily, which encodes the paired domain and homeodomain but lacks the octapeptide. The subfamilies and their structural features are summarized in figure 1.

Additional support for this subgrouping can also be found in the genomic organization of *Pax* genes. For example, genes within a given subfamily share specific intron/exon boundaries (Stapleton et al.,1993). Moreover, some Pax proteins in the same subfamily have been shown to have very similar DNA-binding activities (Czerny et al., 1993, 1995; Epstein et al., 1994a).

Pax Gene Expression Pattern

Mouse *Pax* genes are expressed with a distinct spatiotemporal pattern beginning between day 8 and day 9.5 of embryogenesis. Although several *Pax* genes are also expressed in adult tissues, the primary expression of all known functional *Pax* genes is in the embryo. All *Pax* genes (except *Pax-1* and *Pax-9* which are expressed in the developing vertebral column) are expressed in the developing neural tube and brain, and contribute to early nervous system development (Chalepakis et al.,1993; Noll, 1993; Stoykova and Gruss, 1994). Unlike *Hox* genes, which are characterized by region-specific expression along the anterior-posterior axis, *Pax* genes can show expression along the full length of this axis, but often with a progressive reduction as development proceeds.

Individual *Pax* genes are also expressed at high levels in tissues outside the central nervous system, such as *Pax-2* and *Pax-8* expression in the developing kidney (Dressler et al., 1990; Plachov et al., 1990), *Pax-5* expression in B-lymphocytes (Adams et al., 1993), *Pax-3* expression in paraspinal mesoderm (Goulding et al., 1991), and *Pax-8* expression in the thyroid gland (Plachov et al., 1990).

Pax Gene Developmental Mutations

At least seven phenotypes are known to be associated with loss-of-function mutations in three human *PAX* genes. Mutations in human *PAX3* gene cause Waardenburg syndrome (WS) type 1, type 3 and Craniofacial-deafness-hand syndrome (summarized in Farrer et

al., 1995). Mutations in *PAX6* are associated with familial and sporadic aniridia, Peters' anomaly and cataracts (Ton et al., 1991; Glaser et al., 1992,1994,1995). More recently, mutations in *PAX2* gene have been associated with human kidney and retinal defects (Sanyanusin et al., 1995). In addition, mutations in three mouse *Pax* genes, *Pax-1, Pax-3* and *Pax-6* are known to produce the undulated, Splotch and Small-eye mutant phenotypes, respectively (Balling et al., 1988; Epstein et al., 1992; Hill et al., 1991).

Waardenburg syndrome and Aniridia are the best studied of the above syndromes. Waardenburg syndrome type 1 (Waardenburg, 1951) is a heritable autosomal dominant trait occurring with a frequency of approximately 1 in 100,000 of the population (Tassabehji et al.,1993) and is characterized by white forehead, premature graying of the hair, different colored eyes , and an outward displacement of the inner canthii of the eye (da-Silva, 1991). Of the patients with Waardenburg syndrome, approximately one third are deaf, representing 2% of all adult cases of congenital deafness (Hoth et al., 1993). Klein-Waardenburg syndrome or WS type 3 has been described as combination of WS type 1 and limb abnormalities (Goodman et al., 1982). Splotch (mouse *Pax-3* mutation) and WS 1 (human *PAX3* mutation) have similar neural crest deficiency-associated phenotypes (Tassabehji et al., 1994).

The human congenital eye disease aniridia is characterized by hypoplasia of the iris and affects the iris, lens, cornea, filtration apparatus, and retina, leading to cataracts, corneal opacification, and glaucoma that worsen with age (Glaser et al.,1995). It is an important cause of blindness and a paradigm among human geneticists as a Mendelian autosomal dominant disorder. It occurs because of a decreased dosage of *PAX6*, a gene which controls early events in the morphogenesis of the brain and eye (Glaser et al., 1994). *PAX6* mutations have been detected in both sporadic and familial aniridia. *PAX6* mutations have also been described in Peters' anomaly, a congenital defect of the anterior chamber of the eye, that is usually a central corneal opacity overlying a defect in

the posterior layers of the cornea (Hanson et al., 1994). A broad spectrum of *PAX6* mutations have been found in Aniridia / Peters' anomaly. Large deletions may extend to neighboring genes, including the *WT1* Wilms' tumor gene, causing the WAGR contiguous gene syndrome (Wilms tumor, aniridia, genito-urinary abnormalities and mental retardation). The *Small eye* mouse mutants (associated with mouse *Pax-6* mutations) display phenotypes that include eye defects, primarily complete absence of eye structure or defects of the lens, cornea and retina and of the nose and associated olfactory structures (Hogan et al., 1988).

The human *PAX2* gene is expressed in primitive cells of the kidney, ureter, eye, ear and central nervous system (CNS) (Dressler et al., 1990; Nornes et al., 1990). A mutational analysis of *PAX2* in a family with optic nerve colobomas, renal hypoplasia, mild proteinuria and vesicoureteral reflux revealed a single nucleotide deletion, which cause a frameshift of *PAX2* coding region in the octapeptide (Sanyanusin et al., 1995). The phenotype resulting from *PAX2* mutation in this family was very similar to abnormalities that have been reported in Krd mutant mice (Keller et al., 1994).

Mouse *Pax-1* mutations are associated with undulated phenotypes (Balling et al., 1988). The *undulated* mouse shows reduction of the posterior portion of the vertebrae, with increased intervertebral disk spaces, causing a "wavy" spine (Wright, 1947; Carter, 1947).

A property of *Pax* mutations in both human and mouse is that abnormal phenotypic effects accompany the disruption of only one of the normal pair of genes (Hill and Hanson, 1992). Therefore in human, these disorders segregate as autosomal dominant. In mouse, such heterozygous effects are referred to as semidominant, because homozygotes show increased phenotypic severity. These mutations are assumed to be loss-of-function mutations, as the majority of *Pax* mutations are large scale truncations or frameshift that exhibit similar phenotypes as the missense mutations. The term

haploinsufficiency has been used to describe this aspect of the *PAX2, PAX3* and *PAX6* mutations (Glaser et al.,1994, 1995; reviewed by Read, 1995).

Pax Gene Oncogenic Potential

Not only can an insufficient *Pax* dosage lead to a variety of phenotypes, but over-dosage or gain-of-function *Pax* mutations can also cause developmental defects, often tumorigenesis. So far, murine *Pax* genes have been demonstrated to induce tumorigenesis in mice, and various human *PAX* genes have been tentatively implicated in a variety of human cancers.

When Pax genes are expressed in fibroblasts under the control of the cytomegalovirus (CMV) promoter/enhancer, the observed Pax protein overexpression is accompanied by an uncontrolled increase of cell growth in vitro. When injected into nude athymic mice, cells that constitutively overexpress Pax proteins develop into solid tumors. The oncogenic potential of murine Pax genes appears to be dependent on the presence of a functionally active paired domain. For example, the murine *Pax-1 undulated* point mutation in the paired box, which results in a DNA-binding deficient protein, does not have the transformation activity. The absence of the octapeptide or homeodomain does not affect transforming potential. Although Pax genes induce transformation that results in vascularized tumor formation, metastasis was not demonstrated (Maulbecker and Gruss, 1993).

Wilms' tumor, a pediatric renal carcinoma, is a common malignancy in children, occurring in approximately 1 in 10,000 of the population (Hustie, 1993). The presence of both the *PAX2* protein and Wilms' tumor suppresser protein *WT1* has been observed in primary Wilms' tumor (Dressler and Douglass, 1992). It has been demonstrated that WT1 can bind to three high affinity sites in 5' untranslated *PAX2* leader sequence with high affinity, and repress *PAX2* transcription (Ryan et al., 1995). *PAX8* has also been

demonstrated to be expressed in Wilms' tumor (Poleev et al., 1992).

A frequent site of chromosomal rearrangement in pediatric alveolar rhabdomyosarcoma maps to the *PAX3* locus. It has been shown that the common translocation in this type of rhabdomyosarcoma results in a portion of *PAX3* being translocated and forming a fusion protein with a portion of a forkhead gene FKHR (Galili et al., 1993; Shapiro et al., 1993). The fusion protein retains the entire PAX3 DNA-binding domains and only 55% of the forkhead domain. Since the activity of forkhead proteins is dependent on the presence of an intact forkhead domain (Lai et al., 1990), the activity of the PAX3-FKHR fusion protein would appear to be due to the PAX3 DNA binding domains, which may or may not be modulated by the forkhead region of the fusion protein. It has been shown that the PAX3-FKHR fusion protein is a more potent transcriptional activator than the intact PAX3 protein (Fredericks et al., 1995).

PAX5 has also been implicated in the progression of astrocytomas (which account for 60% of all tumors of the human central nervous system) to their most malignant and prognostically unfavored form - glioblastoma multiforme (Stuart et al., 1994).

Clearly, vertebrate development is sensitive to the precise dosage of PAX protein. Why has natural selection managed such a fragile mechanism?

Functions of Pax Genes

Like homeobox (Hox) genes, *Pax* genes encode transcription factors that play important roles in development, as demonstrated by the abundance of mouse and human congenital defects associated with *Pax* gene mutations.

In Drosophila, paired-box-containing genes may have a role in segmentation. For example, the three earliest characterized genes containing paired box, *paired (prd), gooseberry (gsb)* and *gooseberry*

neuro (gsbn), are segmentation genes of the pair-rule and segment-polarity class. The initial activation of the segment-polarity genes *engrailed (en), wingless (wg)*, and *gsb* has been shown to depend on prd at least in every other stripe (Noll, 1993). In addition, *gsbn* and *pox neuro (poxn)* are involved in neurogenesis. Most interestingly, the critical role of the *eyeless (ey)* gene, the Drosophila homolog of *PAX6*, in controlling Drosophila eye formation has been clearly demonstrated. Ectopic *eyeless* expression induces formation of full-fledged eyes in Drosophila wings, legs and other tissues. This suggests it may be a "master control gene" for eye development (Halder et al., 1995).

Mouse *Pax* genes are expressed after somite formation has established the initial segmentation pattern. Therefore, vertebrate Pax genes are unlikely to be involved in primary segmentation of the body axis. Instead, they appear to have tissue-specific roles in specifying positional information (Strachan and Read, 1994). Analysis of *Pax* mutational phenotypes and murine *Pax* expression patterns may lead to a better understanding of the primary functions of *Pax* genes. *Pax-1* and *Pax-9* should have a role in the development of the vertebral column (Dietrich and Gruss, 1995). All other *Pax* genes have a potential role in CNS development (Stuart et al., 1994). In addition, *Pax-2* is important in kidney and eye development (Sanyanusin et al., 1995); *Pax-3* should be involved in neural crest cell patterning and may inhibit myogenic differentiation (Epstein et al., 1995); *Pax-5* is associated with B lymphocyte development and midbrain/hindbrain boundary patterning (Adams et al., 1992); *Pax-6* plays an important role in eye morphogenesis (Halder et al., 1995); *Pax-8* is associated with thyroid development (Zannini et al., 1992). *Pax* mutational phenotypes and functions are summarized in Table 1.

Although the physiological importance of *Pax* genes have been clearly demonstrated, little is known concerning their molecular mechanisms, such as the up-stream regulators or down-stream targets of Pax proteins. Some functional target sequences for *Pax-5, Pax-8* and *Pax-6* have been identified (Pax-5: Barberis et al.,

1989; Kozmik et al., 1992; Waters et al., 1989; Rothman et al., 1991; Williams and Maziels, 1991; Liao et al., 1992; Pax-8: Zannini, 1992; Pax-6: Cvekl et al., 1994, 1995; Chalepakis et al., 1994b; Richardson et al., 1995; Plaza et al., 1995; see figure 4 of Chapter 4). Pax-5 was identified as a B-cell-specific transcription factor and it potentially regulates the CD19 gene, which encodes a B-cell-specific surface-protein. The sea-urchin *Pax-5* homolog, TSAP, regulates two pairs of non-allelic histone genes, H2A-2 and H2B-2. *Pax-8*, which is expressed in the thyroid, binds to and regulates the thyroperoxidase and thyroglobulin genes. Recently, crystallin genes have been proposed to be *Pax6* targets (Cvekl et al., 1994, 1995; Richardson et al., 1995). The study of Pax protein-DNA interactions will provide important information for understanding the molecular mechanism of Pax proteins.

Paired Domain Is Critical for Pax Functioning

Pax proteins vary from 360 to 480 amino-acids in length. The highly conserved 128 amino acid paired domain is located near the N-terminal end of Pax proteins. The functional importance of the paired domain is well demonstrated by the clustering of *Pax* missense mutations inside this domain. Although the majority of *Pax* mutations are large-scale truncating mutations (gene deletion, frameshifting deletion or insertion, splicing site alteration and nonsense mutation), a variety of *Pax* missense mutations has been reported. Most known missense mutations occur in the N-terminal region of the paired domain (Strachan and Read, 1994). In addition, the oncogenic potential of Pax proteins is also dependent on the DNA binding activity of the paired domain, as the Pax-1 undulated mutant protein, which carries a point mutation in the paired domain that impairs DNA binding, can not induce tumor formation (Maulbecker and Gruss, 1993).

Of the nine mouse Pax proteins (Figure 1), five do not encode any other known DNA binding motifs and thus may exclusively use paired domain to bind DNA. The other four Pax proteins (Pax-

3/4/6/7) also contain an intact paired-type homeodomain. The binding of PAX3 and Pax-6 homeodomains to a series of DNA sites containing a single TAAT core, in different sequence contexts, was not detected (Chalepakis et al. 1994; Czerny and Busslinger, 1995). Paired type homeodomains can form dimer upon binding to palindromic DNA sites, which significantly improves its DNA binding activity (Wilson et al., 1993). However, using the optimal binding sites for both paired domain and homeodomain (palindromic site), in the context of full-length Pax-6 protein, the paired domain proved to be more effective, by about 2 orders of magnitude, in DNA binding than the homeodomain (Czerny and Busslinger, 1995). It seems that paired domain plays a dominant role in determining the DNA binding activity of Pax proteins.

The paired domain may also have roles other than specific DNA binding. For example, it has been shown that a region responsible for a strong transcription inhibition activity is located in the first 90 N-terminal amino acids of the mouse Pax-3 protein, which includes the first 57 residues of the paired domain. This region can function as a transcriptional inhibitor independent of the remaining portions of the Pax-3 protein, as it can be transferred onto a heterologous GAL4 DNA-binding domain (Chalepakis et al., 1994).

Paired Domain Is a Novel DNA-binding Motif

Paired domain is a highly conserved DNA-binding domain that does not share any obvious sequence homology with any other DNA binding protein. Thus the study of paired domain-DNA interactions can provide new perspective for understanding the general principles of protein-DNA interactions, in addition to laying the groundwork for understanding the mechanisms Pax proteins use to regulate gene expression during development.

There is evidence indicating that the paired domain is composed of two sub-domains that bind to two half sites in adjacent major grooves on the same side of DNA helix (N-terminal subdomain

binds to 5' half site), and that the N-terminal domain plays a dominant role in the paired domain-DNA interaction (Czerny et al., 1993; Epstein et al., 1994). When aligning the Pax-5 recognition sequences to obtain a binding site consensus, none of the naturally occurring Pax-5 binding sites completely conform to the long consensus sequence. A subset of Pax-5 binding sites, that match better to the consensus in the 5' half than the rest of Pax-5 sites and do not match the 3' sub-site, can be bound by truncated Pax-5 paired domain lacking 36 C-terminal amino acids. The rest of Pax-5 binding sites that match the 5' half of the consensus sequence less well, match better to the 3' half to the consensus sequence (see Figure 4 of Chapter 4). The bipartite structure of both the Pax-5 paired domain and its binding site was directly demonstrated by Pax-5 methylation interference analysis and in vitro mutagenesis of both the Pax-5 paired domain and its recognition sequence. Thus Pax-5 DNA binding sites contain compensatory base changes in their half sites that explain the versatile and seemingly degenerate DNA sequence recognition of Pax-5 protein (Czerny et al., 1993).

What are the structures of the two subdomains of paired domain? What is the relationship of these two domains? What is the structural basis for the specificity of paired domain-DNA interaction? How could single missense mutations in the paired domain lead to the observed phenotypes? Could paired domain-DNA interactions provide new information for understanding the general principles of protein-DNA interactions? These are the questions we hoped to answer by solving a paired domain-DNA complex crystal structure.

Figure Legends

Figure 1

Sub-family classification and structural features of PAX proteins. PD denotes the paired domain, OP the octapeptide and HD the paired-type homeodomain. The three helices in homeodomain are highlighted. The length of proteins and the distance between the structural features are not drawn in proportion. The classification is based on the overall sequence organization (presence of a paired-type homeodomain and an octapeptide motif, location of introns and overall sequence identity) and especially on comparison of the paired box sequences (Walther et al. 1991; Wallin et al. 1993; Stapleton et al. 1993).

Table 1

This table summarize the functions of PAX genes and phenotypes of PAX developmental mutations. Pax proteins in the same subfamily are clustered. WS denotes Waardenburg syndrome. CNS denotes the central neuvous system.

Figure 1.

Classification and Structural Features of Pax Proteins

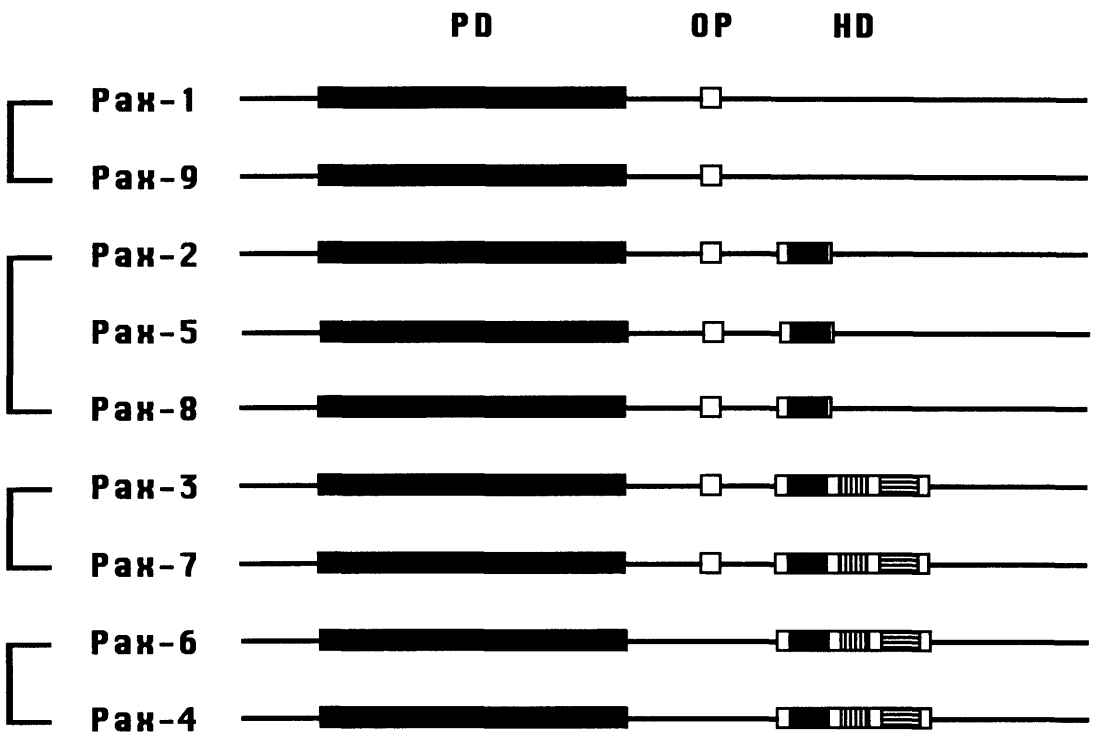

Table 1. PAX Gene Functions and Phenotypes of their Mutations

Gene	Function	Loss of function phenotype	Gain of function phenotype
PAX1	development of vertebral column?	*Mouse*: undulated *Human*: ?	?
PAX9	development of vertebral column?	?	?
PAX2	kidney development, CNS development	*Mouse*: Krd *Human*: kidney & retinal problems	*Mouse*: abnormal kidney development *Human*: role in Wilms'tumor
PAX5 (BSAP)	B-cell development, CNS development	*Mouse*: B-cell & brain abnormalities	*Human*: role in astrocytoma
PAX8	thyroid development, CNS development?	?	*Human*: role in Wilms' tumor?
PAX3	neural crest cell patterning	*Mouse*: Splotch *Human*: WS1, WS3, Craniofacial-deafness-hand syndrome	Rhabdomyosarcoma
PAX7	CNS development?	?	?
PAX4	?	?	?
PAX6	eye development, CNS development	*Mouse*: small-eye *Human*: aniridia, Peters' anomaly, cataracts	?

Chapter 2

Purification, Crystallization and Structural Determination of Prd Paired Domain - DNA complex

When we started to try to solve the structure of paired domain by means of crystallography, very little information about the DNA binding site of Pax proteins was available. Susie Jun and Claude Desplan (Rockefeller University) defined a optimal DNA binding site of Drosophila Prd paired domain by using in vitro selection and amplification of randomized DNA sequences, and that set the basis for our collaboration.

A. Purification

Purification of Drosophila prd Paired Domain

Prd paired domain was initially prepared as a C-terminal fusion with gluthione S-transferase. The chimeric protein was over-expressed in *E. coli* and purified on gluthione-agarose column. However it was very difficult to obtain specific cleavage between gluthione S-transferase and paired domain. Non-specific cleavage also imposed difficulty in purification. Although correct cleavage rate could reach 30% in solutions containing specific DNA site and 50% glycerol (high glycerol may help to stabilize loose domain structure, as reviewed by Sousa, R, and Lafer, E.M., 1990), the final recovery yield was below 0.2-0.3 mg per litre E. coli culture. Thus we tried several new plasmid expression vectors. Among them, a vector with an N-terminal polyhistidine tag, pET14bprdPDB, gave high expression level in soluble phase and its polyhistidine tag could be specifically cleaved, and so was later used to express the *Drosophila* Prd paired domain in E. coli strain BL21(DE3). The protein used in our crystallographic study contains the whole Prd paired domain and and four additional residues (Gly-Ser-His-Met) on the N-terminal end that was introduced from expression vector as part of the polyhistidine tag. All plasmid vectors I have tested were constructed by our collaborator Susie Jun (Rockefeller University)

Cells were grown at 37° and were induced with 0.4 mM isopropyl-ß-D-thiogalactoside (IPTG) when they reached OD_{600}=0.8. Cells were harvested 3 hours after induction, washed with

prechilled phosphate-buffered saline buffer, frozen in a dry-ice/ethanol bath and stored at -80°C. Sonication was carried out in a buffer containing 25 mM Hepes pH 7.6, 0.1M KCl, 0.1% NP-40, 0.3 mg/ml lysozyme, 7 mM 2-mercaptoethanol, 1 μg/ml aprotinin, 1 μg/ml pepstatin, 1 μg/ml benzamidine, and 1 μg/ml sodium metabisulfite. The cell lysate was diluted with solution A (25 mM Hepes pH 7.9, 0.1 M NaCl, 5 mM MgCl2, 15% glycerol, 0.1% NP-40, 7 mM 2-mercaptoethanol) and loaded onto a Ni-NTA column (Novagen). The column was extensively washed with 8 mM imidazole (pH 8.0) in solution A, and then with 40 mM imidazole in solution A; the Prd paired domain was eluted with 100 mM imidazole in solution A. The eluted protein was treated with 0.25U/μl thrombin at 30°C for 15-20 hours to remove the N-terminal polyhistidine tag, and the reaction was stopped by adding 1 mM PMSF to the solution. The Prd paired domain was purified with a Mono-S column (Pharmacia), using a gradient of 0.3 M to 0.7 M NaCl in 40 mM phosphate buffer (pH 6.6), containing 1 mM DTT. Prd paired domain was eluted out by 0.5-0.55 M NaCl. The purified protein gave a single band on an overloaded SDS gel in the absence of reductant. The protein used for crystallization was then purified by gel filtration on a superdex-75 column (Phamacia), with a buffer containing 10 mM bis-tris-propane (pH7.0) and 1 mM DTT. Protein was concentrated by Centricon-3, then frozen by liquid nitrogen and stored at -80°C. In later stage of crystallization, protein purified in this way was further purified by preparative reverse phase HPLC on a Vydac C4 column, and then was lyophilized. Lyophilized proteins were then resuspended by a buffer containing 10mM bis-tris-propane (pH7.5), 1mM DTT, aliquoted, frozen by liquid nitrogen and stored at -80°C. The HPLC/ lyophilization step also function as a concentration step, in this way protein could be concentrated to 22 mg/ml, while it was hard to concentrate protein up to 10 mg/ml by Centriprep-3 or Centricon-3 (Amicon). The HPLC purified protein could produce crystals more reproducibly. The final yield of purification is about 5 mg per litre of *E. coli* culture.

The chemical homogeneity and identity of the purified Prd

paired domain was further confirmed by N-terminal sequencing, amino acid composition analysis, high resolution mass spectrometry (Harvard MicroChem facility), and gel shift experiments.

Purification of DNA oligomers used for crystallization

We used solid-phase phosphoramidite method on an Applied Biosystems DNA/RNA synthesizer 392 for producing all of the DNA oligonucleotides used for crystallization. Individual DNA oligonucleotide strands containing 5-dimethoxytrityl (DMT)-group were purified by preparative reverse-phase HPLC on a Vydac C4 column, using an acetonitrile gradient in 50 mM triethylammonium acetate (pH6.5). The trityl group was cleaved by treatment with 1.1% trifluroacetic acid for 10 min, and the solution was immediately neutrilized by 1.4% triethylamine. Oligomers were then dialyzed extensively against 10 mM triethylammonium bicarbonate (pH7.0) and were then lyophilized. The detrityled oligonucleotides were purified a second time by a C4 reverse-phase column and dialyzed extensively against 10 mM triethylammonium bicarbonate (pH7.0). DNA strands were annealed by heating at 90°C for 10 min and cooling slowly to room temperature. DNA duplexes were stored as freeze-dried aliquots.

The uncoupled failure products were capped by acetylation in each synthesis cycle, and the capped oligos could be easily separated from DNA oligomers with DMT group in reverse-phase HPLC. Thus We kept DMT protecting group after last cycle and then purified oligomers by two runs of reverse-phase HPLC as described above, in order to totally get rid of those uncoupled failure products. However for short DNA oligomers (15mer or shorter) used for crystallization trials, we expected that one-step purified DNA should be sufficiently pure. For example, I obtained paired-DNA complex single crystals with a 15mer DNA oligo. While crystal with DMT-on/two-step purified DNA oligo diffracted 2.5 Å, DMT-off/single step purified diffracted to at least 2.8 Å.

B. Co-crystallization of a Prd Paired Domain - DNA Complex

Selection of DNA Sites and Results of Co-crystallization Trial

When we started our cocrystallization trials, little information was available about the DNA binding specificity of paired domain. The in vitro optimal DNA-binding site of Prd paired domain was deduced from selection and amplification experiments with randomized DNA sequences. The binding site consensus is 12 base pairs long, CGTCACG(G/C)TT(G/C)(A/G). Considering the footprinting of Prd paired domain is 15 base paired long, we decided to search cocrystallization conditions with 14 to 21 base pairs long DNA oligomers, which contains the whole binding site consensus.

It has been repeatly shown that the sequence and length of the DNA oligo used in cocrystallization trials have significant effects on the quality of the cocrystals produced (Jordan et al., 1985; Schultz et al., 1990; Liu et al., 1990; Wolberger et al., 1991). The differences in the DNA length as little as one base pair can dramatically effect the crystal quality. The sequence identity at the 5' and 3' ends of the DNA, in particular the overhanging bases, if any, can also have a large effect on the quality of the crystals. Thus we decided to test a variety of different DNA sequences and lengths in our cocrystallization trials. I first tested the effect of DNA length on the crystallizability of prd paired domain - DNA complex. I synthesized and purified 8 DNA duplexes with different lengths from 14 mer to 21 mer. I was able to obtain microcrystals only with the 15 mer DNA oligomer, after using volatile salt ammonium acetate in the droplet that is neccesary to keep the protein - DNA complex soluble and to obtain any sort of microcrystal. Then I tried 4 other 15 mers with different end bases and/or overhanging bases. With one of the 4 oligomers, which has two overhanging bases (AA/TT), I obtained nice crystals that diffracted to 2.5 Å resolution.

Using Volatile Salts for Crystalliztion

In low ionic, neutral pH, the solubility of prd paired domain - DNA complex is low (lower than 1 mg PrdPD/ml), even with the presence of excessive DNA (which slightly improved the complex solubility). Preliminary studies revealed that the solubility of the Prd paired domain-DNA complex was sensitive to several factors, including ionic strength and pH. High ionic (> 0.25 M NaCl) or alkaline pH (pH> 8.0) can dramatically increase the solubility to above 10 mg PrdPD/ml, with a DNA:protein ratio of 1.5:1.0. However, I was not be able to obtain any ordered solid form, in the high salt (> 0.25 M NaCl) or high pH (> pH 8.0) conditions, with any DNA oligomers I have tried.

At this point, the dynamic light scattering experiment (Ferre-D'Amare and Burley, 1994) indicated that Prd protein-DNA complex is mono-dispersive in solutions containing up to 0.2 M NaCl. In many cases, monodispersity suggests conformational homogeneity. Empirical observations suggest that macromolecules that are monodispersive under "normal" conditions crystallize readily, whereas randomly aggregating or polydispersive systems rarely, if ever, yield crystals (Ferre-D'Amare and Burley, 1994). This result is both encouraging and informative. In the early crystallization trials, the drops initially contains high salt (> 0.25 M NaCl), the salt concentration would go even higher upon equilibrating with reservoir solution containing precipitant. This could cause partial disassociation of protein-DNA complex as indicated by gel-shift. However it seems possible to achieve a soluble mono-dispersive system by using volatile salts.

I then extensively searched the possibility of using volatile salt ammonium acetate and ammonium bicarbonate to cocrystallize prdPD - DNA complex. Ammonium bicarbonate seems not suitable for cocrystallization, because the pH of the droplets containing ammonium bicarbonate tends to go up. The pH of droplets containing ammonium acetate can keep stable around pH 7.0, in a period of several weeks at room temperature, and thus is more useful near

neutral pH. Evaporation of ammonium acetate from droplets decreases the ionic strength in the droplet, and thus drive the PrdPD-DNA complex into supersaturation. The rate of this process depends on the ammonium acetate concentration in the drop solution and in the well solution, as well as the size of the droplet and temperature. Ionic strength is an important determinant of the strength of electrostatic interaction and hydrophobic interaction. It is pretty common that salt can influence the solubility of protein or protein-DNA complex. We expect volatile salt could also be useful for crystallizing other protein or protein-DNA complex. In fact, we have lately obtained several crystal forms of PAX6 paired domain - DNA complex using volatile salt ammonium acetate.

Crystallization Condition

It was interesting that co-crystals could grow in similar conditions and to similar morphology and size in both MPD and PEGs. However crystals grew from MPD could only diffract to about 8 Å resolution, while crystal grew from PEG400 diffracted to 3.2 Å, and crystals from PEG1000 were able to diffract to 2.5 Å resolution. Crystals with the DNA oligo shown in Figure 1d of chapter 3 were grown by the evaporation of volatile salts from the hanging drops. Extensively lyophilized DNA oligomers were resuspended with 10 mM bis-tris-propane (pH 7.0) at a concentration of 1 $O.D._{260}$ per microlitre. Then 1.76 µl of above DNA solution was mixed with 1.81 µl "7.5X buffer" containing 2.25 M ammonium acetate (pH7.0), 0.15 M $MgCl_2$, 37.5 mM DTT, 0.75 mM EDTA. Then 5 µl 22 mg/ml PrdPD was slowly added to above DNA-containing solution while stirring with a pippete tip. Adding DNA to protein or adding protein too quickly would results in some irreversible precipitation. Above DNA-protein mixture was then mixed with equal volume of reservoir solution as the hanging drops and these drops were equilibrated against a reservoir containing 10% PEG 1000 and 5 mM DTT. Crystals grew in 4 to 5 days. Co-crystals diffracting to 2.5 Å resolution grow in orthorhombic space group $P2_12_12_1$, with a=39.6 Å, b=68.6 Å, c=100.5 Å.

C. Structure Determination

Preparation of Heavy Atom Derivative Crystals

We used multiple isomorphous replacement method to solve phase problem. Heavy atoms were introduced into isomorphous crystal by replacing thymine with 5-iodouracil during DNA synthesis.

Iodine atoms in the DNA are not stable upon exposure to light and alkaline conditions. We took special care with the handling of the iodinated DNA oligomers. First we tried to keep oligomers in a dark environment whenever possible, in the whole process of synthesis, purification and crystallization. Secondly, we used milder condition for oligomer deprotection. Iodinated DNA oligomers was deprotected in fresh saturated ammonium hydroxide at room temperature for 20 to 24 hours, then the cap of the vials was opened and kept at room temperature for another 12 hours (to allow ammonium hydroxide to evaporate and to prepare for speed-vac). Thirdly, the trityl-off reaction was controlled with great care. After incubation with 1.1% of trichloracetic acid for 8 minutes, the reaction solution was neutralized immediately with 1.2% triethylamine, and then one tenth volume of 0.5 M bis-tris-propane buffer (pH7.0). The solution was then extensively dialysed against 10 mM TEAB before the second step of HPLC purification. The purity of final iodine-substituted DNA oligomer was confirmed with Mono-Q anion exchange column (Pharmacia LKB, Piscataway, New Jersey), which showed that the molar ratio of iodinated DNA and DNA that has lost iodine was 100 to 1 or higher. I found that it is not necessary to use the more expensive FOB (fast oligonucleotide deprotecting) protection reagent, which uses different protecting groups than the standard CE protection method. First the CE-protected DNA oligomers purified in the way decribed above were fully suitable for making isomorphous heavy atom derivative crystal. Secondly, FOB-protected column was not commercially available.

Thus the 3' end nucleotide is usually CE-protected, and DNA oligomers synthesized with FOB reagent still have to be deprotected in CE-deprotection condition.

We tested a number of these modified DNA oligomers in crystallization trials, and found that substitution of thymine by 5-iodouracil in base pairs 11, 12 or 14 produced isomorphous crystals which were suitable for phasing. (After the structure was solved, we noticed that these three thymine bases are neither contacted by protein from major groove, nor involved in crystal packing). All three derivative crystal forms were isomorphous to native crystal and diffracted to 2.5 Å resolution, same as native crystals.

Data Collection and Reduction

During crystal data collection, there appeared to be gradual changes in the cell dimensions. Most severe changes occured to cell dimension b, which can change from 64.7 Å to 69.3 Å (thus increasing by 7.1%). It is first considered that the change may be caused by temperature fluctuation. We then tried to collect data at constant 10°C and 2°C respectively. The problem persisted. Then we noticed some relationship between the age of the crystal and the the length of b axis. We surmised that the cell dimension change may be caused by the existence of trace amount of ammonium acetate. We eventually solved the cell dimension change problem by the following steps: first, "aging" the crystals for at least two weeks before data collection; second, improving crystallization condition so that the well solution does not contain any ammonium acetate which was originally used for controlling degree of supersaturation; third, mounting crystals without adding any well solution.

All data finally used for structure determination were collected at room temperature on the R-axis image plate system of our laboratory. The crystal have unit cell dimensions of a = 39.6 Å, b =58.6 Å, c = 100.5 Å, and of the orthorhombic space group P $2_12_12_1$.

Initial determination of the lattice parameters was done by collecting 30 frames of oscillation data (Δphi = 1°). The diffraction pattern of these 30 frames were converted to positions in reciprocal space using the conversion programs developed by Mark Rould (extract_peaks.for, peaks_to_reciprocal_space_coordinates.for). By measuring the distances and angles of individual spots in the reciprocal space, using the crystallography graphics program FRODO (Jones, 1978), the primary reciprocal lattice parameters were determined. Then the real space crystal unit cell parameters were deduced. We found at this point that all unit cell angles were very close to 90°, and surmised the crystal belong to a primitive orthorhombic space group. The existence of 2-fold axes in all three directions were confirmed by testing for the presence of the two-fold symmetry operators. Thus we could obtain a full data set by collecting 90° data. After a full native data set was collected, we examined for systematic absences which indicated that we had a space group P $2_12_12_1$. The space group is further confirmed by the solution of the difference Patterson map.

All data sets were reduced using the program DENZO (Z. Otwinowski). Crystal and camera parameters were refined, and intensity mesurements were made by using profile fitting of the recorded spots. Partially recorded reflections were merged and integrated from successive oscillation frames (merge-denzo.for, M. Rould). Data were then scaled using the program SCALEPACK (Z. Otwinowski). We devided merged oscillation frames by 5° wedges, then applied a single scale factor for each wedge. No explicit corrections were made for absorption or crystal decay. Derivative data sets were local scaled to the native data set using the program MAXSCALE (M. Rould).

Structure Determination by MIR Method

Reflections with large intensity differences (>7σ) between the native and derivative data sets were removed from the reflection list (Exorcise.for, M. Rould), because those few reflections strongly

biased the difference Patterson maps. Derivative data sets were local scaled against the native data set again, using MAXSCALE (M. Rould). Isomorphous difference Patterson maps and anomalous difference Patterson maps were calculated for each derivative using the program PROTEIN (Steigemann, 1974). Exorcise and MAXSCALE significantly improved the quality of Patterson maps. At this point our isomorphous difference Patterson maps and anomalous difference Patterson maps showed clearly the heavy atom peaks in the Harker sections (Figure 1). We then picked initial heavy atom sites corresponding to heavy atom peaks in the Harker sections using the program HASSP (Terwilliger et al., 1987). HASSP is an independent program which systematically searches the difference Patterson function and pick up potential heavy atom sites with large values for both self- and cross-vector positions.

The refinement of heavy atom parameters was carried out using the program REFINE from CCP4 package (The SERC Collaborative Computing Project No.4, a Suite of Programs for Protein Crystallography [Distributed from Daresbury Laboratory, Warrington WA4 4AD, UK, 1979]). After refining the heavy atom parameters (positions, occupancy, and thermal parameters) for every derivative, we used the refined heavy-atom positions to fix the origin of the unit cell with respect to the positions of the heavy atom sites from the three derivatives. With the heavy atom sites initially refined for every derivative, we used difference Fourier methods to check the correctness of the heavy atom sites (Henderson and Moffat, 1971). After refining the heavy atom parameters for every derivative individually, we did cross-phased refinement using the program PHARE from CCP4 package (SERC, 1979) in order to reduce bias (Blow and Mattews, 1973). With this program, the parameters of one derivatives are refined while the other two derivatives are used to calculate phases. After every derivative was refined twice by cross-phase refinement, we generated an initial MIR map with the mean figure of merit 0.59 at 2.5 Å resolution. We then used a procedure (Rould et al., 1992) to improve the phase quality. This procedure decouples heavy atom

parameter refinement from the calculation of parent phases by first solvent-flattening (Wang, 1985) the initial MIR map in order to generate new solvent-flattened phases. These new phases, in turn, are used in the second round of refinement of heavy atom parameters. New MIR phases are not updated until the convergence of the refinement. The new MIR map (mean fom = 0.71) was subject to another round of solvent flattening to give the final MIR electron density map (Figure 6 of chapter 3, mean fom = 0.79). All of the DNA was clearly resolved in this map, as were almost all the sidechains and mainchain carbonyl groups of the N-terminal domain of the protein (Figure 6 of chapter 3). The electron density for the C-terminal domain was not as good (it is packed less rigidly in the crystal), but about half of the sidechains of this globular sub-domain were clear. The initial model was built using TOM FRODO in Silicon Graphics computer (Israel, M., Chirino, A. J. and Cambillau, C. M., personal communication). The initial idealized B-form DNA was generated using the program Insight.

Structure Refinement

The initial model was subjected to multiple rounds of positional refinement (Brünger, 1992a) and manual adjustment. Refinement was monitored by following the free R-factor to avoid overbuilding (Brünger, 1992b). In later stages of refinement, tightly restrained individual B-factors were used. Local scaling of the observed and calculated structure factors (using a minimum neighborhood of 100 reflections and excluding the reflection being scaled) was also done to correct for absorption and anisotropic diffraction. In the final cycle, 13 water molecules were included in the model. Every water molecule added forms at least two hydrogen bonds with the paired-DNA complex, and has a B-factor lower than 50.0. Although most structural features including water molecules were cleared resolved in the initial unbiaed MIR map, all of the key contacts and the key features of the complex were further confirmed by checking simulated annealing omit maps (Hodel et al., 1992). About 30% of the sidechains of the C-terminal domain could not be

built with confidence and were modeled as alanines; the first 5 and last 4 residues of the polypeptide also were omitted. (A few of these N-terminal residues were ones introduced during cloning, and thus our model includes residues 2-124 of the paired domain.) Our current model has an R factor of 23.4% and a free R factor of 28.4% with good stereochemistry (Table 1). All phi and psi angles, except for residues 78 (in the linker) and 91 (in an extended loop), are in allowed regions of the Ramachandran plot (Figure 2).

Figure Legends

Figure 1

This figure shows three sections of the isomorphous difference Patterson map, corresponding to Harker section u=1/2, v=1/2, and w=1/2, calculated from the native data set and the dIU(11) data set (see Chapter 3 Table 1). The map was generated by the program PROTEIN (Steigemann, 1975), and used data after local scaling, from 20 to 2.8 Å resolution. The contours of the maps start at 1 sigma and are in increment of 1 sigma. The peaks representing the single iodine atom are clear in this map.

Figure 2

Ramachandran plot. This figure shows the location in Phi, Psi space of each amino acid residue from the final model of the Prd paired domain-DNA complex. The angle phi (the dihedral angle about N-Cα bond) is shown on the abscissa, and the angle psi (the dihedral angle about the Cα-C bond) is shown on the ordinate. All phi (ϕ) and psi (φ) angles, except for residue 78 (in the linker) and 91 (in the extended loop), are in allowed regions of the Ramachandran plot. Coordinates in boxes indicate glycine residues.

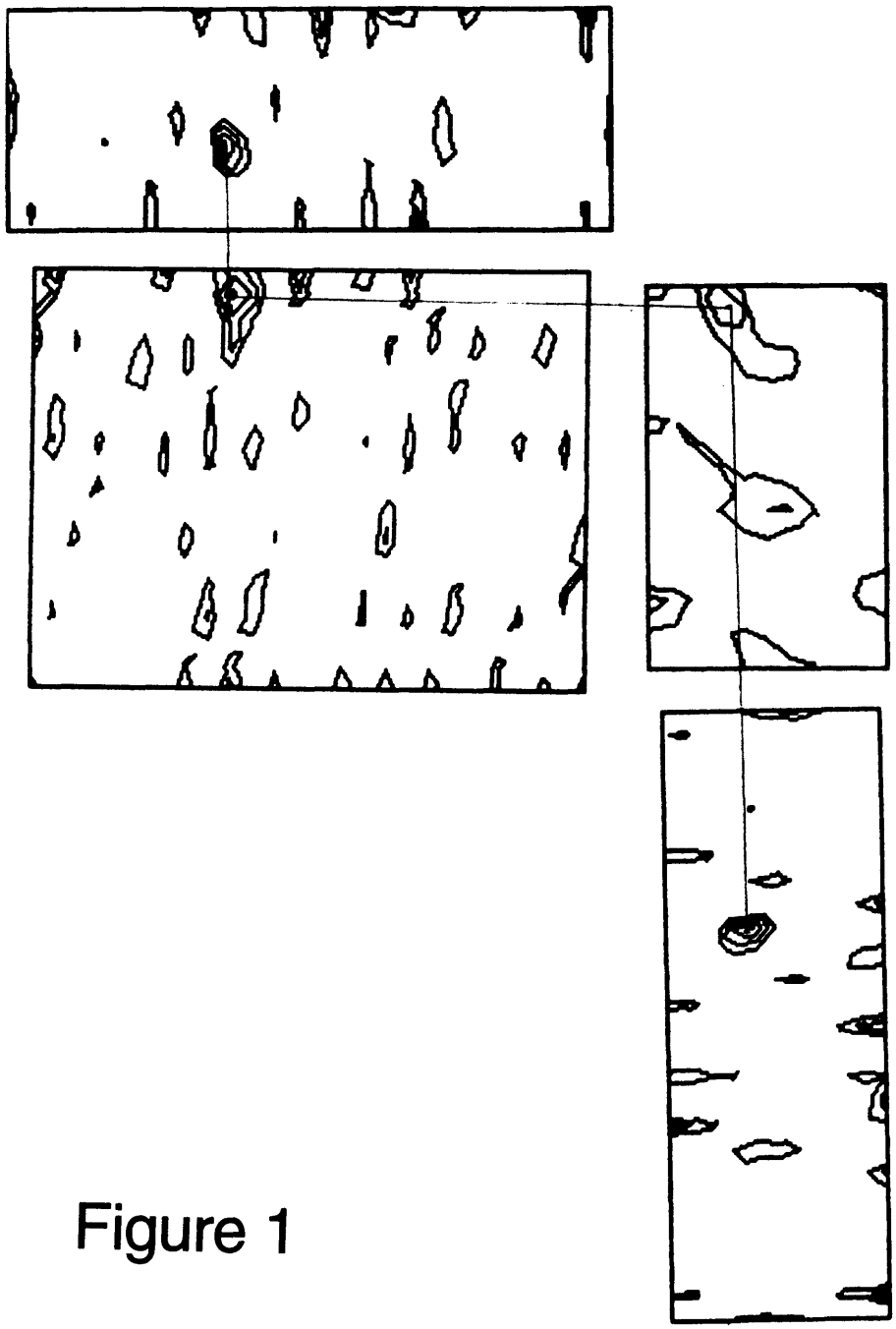

Figure 1

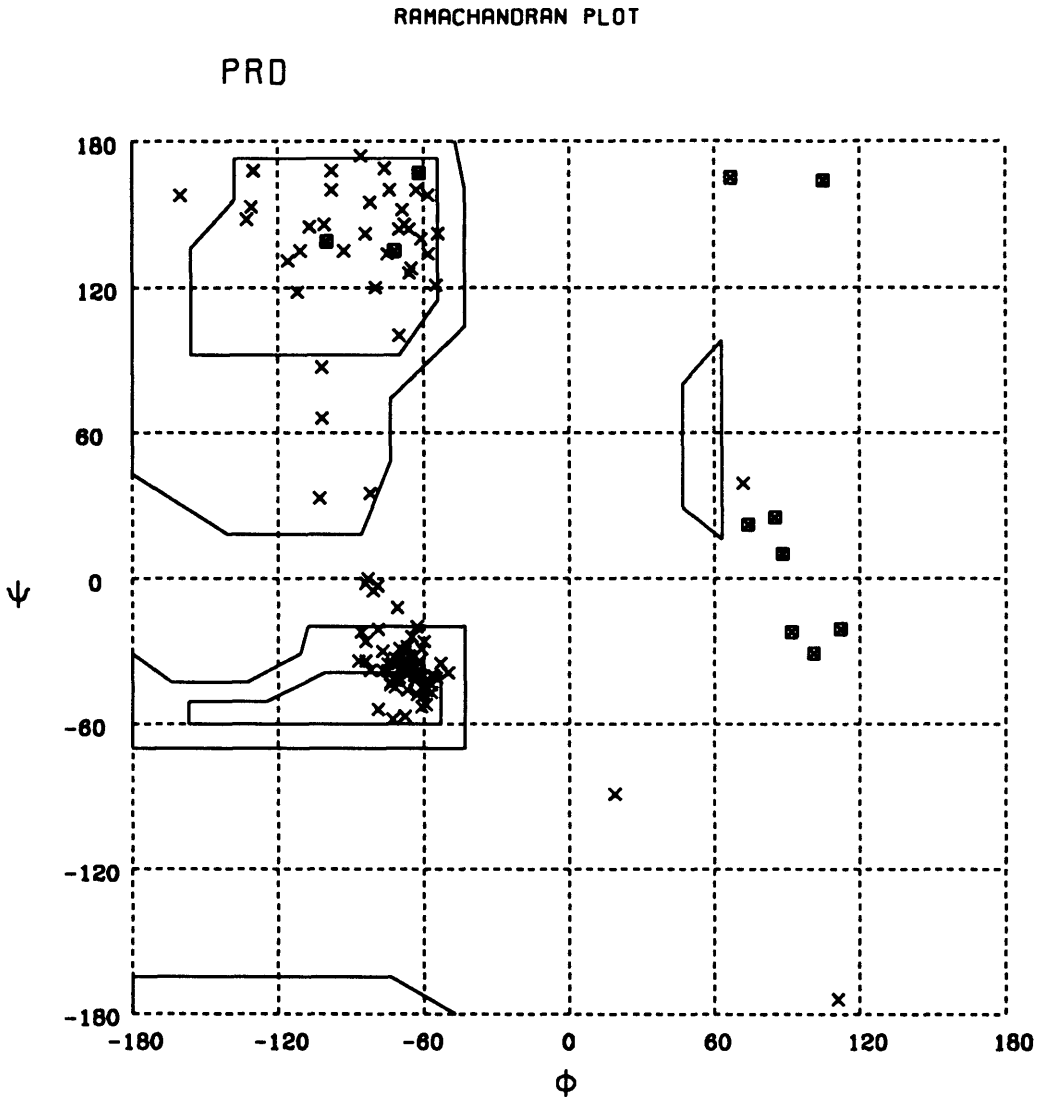
RAMACHANDRAN PLOT
PRO
180
120
60
0
-60
-120
-180
ψ
-180
-120
-60
0
60
120
180
Φ

Figure 2

Chapter 3

Crystal Structure of a Paired Domain-DNA Complex at 2.5 Å Resolution Reveals Structural Basis for Pax Developmental Mutations

Summary

The 2.5 Å resolution structure of a co-crystal containing the paired domain from the *Drosophila* Paired protein and a 15 bp site shows structually independent N-terminal and C-terminal sub-domains. Each of these domains contains a helical region resembling the homeodomain and the Hin recombinase. The N-terminal domain makes extensive DNA contacts, using a novel β-turn motif that binds in the minor groove and a helix-turn-helix unit with a docking arrangement surprisingly similar to that of the λ repressor. The C-terminal domain is not essential for Prd binding and does not contact the optimized site. All known developmental missense mutations in the paired box of mammalian *Pax* genes map to the N-terminal sub-domain, and most of them are found at the protein - DNA interface.

Introduction

The paired domain is a conserved DNA-binding domain (Treisman et al., 1991; Chalepakis et al., 1991) found in a set of transcription factors (Pax proteins, Figure 1a) that play important roles in development (Gruss and Walther, 1992). This 128 amino acid domain was first identified in the *Drosophila paired (prd)* and *gooseberry* genes (Bopp et al., 1986) and often is found in association with a homeodomain (Walther et al., 1991). Numerous paired domain proteins are known, and nine *PAX* genes have been identified in the human genome (Walther et al., 1991; Stapleton et al., 1993; Wallin et al., 1993; Figure 1a). A number of murine and human developmental mutants are known to have alterations in specific *Pax* genes, and several of these involve missense mutations in the paired domain (reviewed by Gruss and Walther, 1992; Strachan and Read, 1994; Figure 1b). Mutations in the human *PAX3* and *PAX6* genes cause Waardenburg's syndrome (Tassabehji et al., 1992; Baldwin et al., 1992) and aniridia (Ton et al., 1991; Hill et al., 1991;

Glaser et al., 1992), respectively. The *Pax* genes also appear to have oncogenic potential: overexpression of *Pax* genes can lead to transformation in cell culture and *in vivo*, and this oncogenic potential is dependent on the presence of a functional paired domain (Maulbecker et al., 1993). A chromosomal translocation of *PAX3* is implicated in the generation of a myosarcoma (Barr et al., 1993; Galili et al., 1993; Shapiro et al., 1993).

Only a few of the physiological targets of the Pax proteins have been identified (Czerny et al., 1993), but optimal binding sites have been selected from randomized DNA for the paired domains of Prd, Pax-2, Pax-6, and Pax-8 (Figure 1c; Epstein et al., 1994a; Jun and Desplan, manuscript in preparation), and it has been shown that these sites can mediate transactivation in cell culture assays. These optimized binding sites, which share a common core sequence, are relatively long (13-20 bp), but they appear to be recognized by monomers of the paired domain (Treisman et al., 1991; Chalepakis et al., 1991; Czerny et al., 1993; Epstein et al., 1994a). Genetic and biochemical studies have indicated that the 128 amino acid paired domain has a bipartite structure and that the N- and C-terminal sub-domains bind to distinct regions of the DNA consensus sites defined for the Pax-5 and Pax-6 proteins (Czerny et al., 1993; Epstein et al., 1994).

To understand the role of the paired domain in DNA recognition and gene regulation, we have crystallized and solved the structure of a complex that contains the paired domain from the *Drosophila* Paired (Prd) protein with a 15 bp duplex containing an optimized binding site (Figure 1d). The structure of this complex reveals how a β-turn can be used for minor groove recognition, gives important new information about the docking of helix-turn-helix units and provides a structural basis for understanding PAX developmental mutants.

Results and Discussion

Overall Arrangement of the Paired Domain-DNA Complex

The co-crystal structure shows that the paired domain actually includes two structurally independent globular domains (Figure 2). The N-terminal domain contains: 1) a short region of antiparallel β-sheet followed by a type II β-turn; 2) three α-helices with a fold that resembles the homeodomain and the Hin recombinase; and 3) an extended C-terminal tail. The C-terminal domain is somewhat smaller. It contains three α-helices, and this helical unit also has a fold resembling the homeodomain and the Hin recombinase.

The binding site chosen for the crystallographic studies (Figure 1d) was defined by using *in vitro* selection and amplification of randomized DNA sequences (Figure 1c; Jun and Desplan, manuscript in preparation) and it is very similar to the optimized sites defined for other paired domains. The crystal structure shows that the N-terminal region of the paired domain makes extensive contacts with this 15 bp optimized binding site, and several different secondary structures participate in recognition. A β-sheet (residues 4-6 and 10-12) grips the sugar-phosphate backbone of the DNA, and this is immediately followed by a β-turn that makes critical base contacts in the minor groove (residues 13-16, τ2 in figure 1a; Figures 2, 3). The first helical region (residues 20-60) contains a HTH motif: Helix 2 makes extensive phosphate contacts and helix 3 binds in the major groove (Figures 2, 4). The C-terminal tail (residues 65-72) of this domain also makes minor groove contacts near those made by the β-turn (Figure 2).

There is a short linker (residues 73-78) between the N-terminal and C-terminal domains; the structure shows no protein-protein contacts between these globular domains. The C-terminal domain does not make any DNA contacts with our optimized binding

site (see discussion), and all of the known missense mutations in the paired domains map to this N-terminal sub-domain. However, biochemical studies suggest that the C-terminal domain may have a significant role in the DNA-binding of other paired domains such as Pax-5 and Pax-6. The structure of the C-terminal domain and similarities with the Hin recombinase suggest how the C-terminal domain may contact DNA in those other systems.

Minor Groove Contacts from the β-turn

The N-terminal portion of the paired domain contains a type II β-turn that fits directly into the minor groove of the DNA (Figure 3). The primary sequence of this region is conserved in the Pax proteins, and several of the known Pax developmental mutations map to this β-turn. In the Prd paired domain, this critical turn includes Ile 13, Asn 14, Gly 15 and Arg 16, and this turn contacts base pairs 9-11 of the binding site (Figures 2, 3, 5). Contacts made by the β-turn include: a hydrogen bond between the Asn 14 side chain and the N2 of the guanine at bp 9; van der Waals contacts between Gly 15 and the cytosine at bp 9; a hydrogen bond between the carbonyl oxygen of Gly 15 and the N2 of the guanine at bp 10; van der Waals contacts between Arg 16 and the sugar phosphate backbone; and a water-mediated contact between Arg 16 and the O2 of the thymine at bp 11 (Figures 3, 5).

The docking of this β-turn appears to be stabilized by protein-protein and protein-DNA contacts from flanking regions. Thus a short antiparallel β-sheet (residues 4-6 and 10-12) contacts one strand of the DNA backbone and the loop between the two strands of this β-sheet (residue 6-10) interacts with residues 40, 44 and 45 in the HTH unit (Figure 2b). The docking of the β-turn also is constrained by Pro 17, Leu 18, and Pro 19, which interact with the DNA backbone. Finally, we note that the β-turn and the β-sheet are held against the C-terminal tail (residues 65-72, see below) by a hydrophobic interface, and these substructures contact adjacent regions of the minor groove (Figures 2, 6b).

Major Groove Contacts by the N-terminal HTH Motif

The helical portion of the N-terminal domain, which begins just a few residues after this critical β-turn, contains three α-helices (residues 20-32, 37-43, and 47-60). This helical unit has a fold that superimposes well on the homeodomain and on the Hin recombinase: helix 1 and helix 2 pack against each other in an antiparallel arrangement and are roughly perpendicular to helix 3. Helix 3, the "recognition helix," fits directly into the major groove, and side chains from this helix contact base pairs 4-8 of the binding site (Figures 2, 4, 5). Ser 46, which is the residue immediately preceding this α-helix, makes van der Waals contacts with the thymine at bp 7. His 47, which is the first residue in the recognition helix, forms a hydrogen bond with the guanine at bp 4. Continuing along helix 3, we see that Gly 48 and Ser 51 make van der Waals contacts with the methyl group of the thymine at bp 5. Similarly, Cys 49 contacts the methyl of the thymine at bp 7. Lys 52 bridges two phosphates and contacts the N7 of guanine at bp 8 (Figure 4). There are several well-ordered water molecules at the protein-DNA interface, and these also may play a role in recognition.

This helical unit also makes extensive contacts with the sugar phosphate backbones (Figure 4). Helix 1, which runs across the major groove, contributes a phosphate contact from Arg 23 but this helix is too far from the DNA to make any other contacts. Additional backbone contacts are made by Arg 35 and Pro 36, which are in the turn between helix 1 and helix 2 (Figures 2, 4). Other backbone contacts from this region involve: Cys 37 and Arg 41 from helix 2; Val 45 and Ser 46 from the turn between helices 2 and 3; and Cys 49, Ser 51 and Lys 52 from helix 3.

C-terminal Tail from the N-terminal Domain Binds in the Minor Groove

The N-terminal domain has a C-terminal tail (residues 65-72) that binds in the minor groove. Conserved residues at the end of helix 3 help fix the position of the extended polypeptide chain. Thr 60 (which is found in all paired domains) helps cap helix 3, and this is followed by an invariant Gly. Residue 63, which is always an Ile or a Leu, anchors the tail in a hydrophobic pocket. In addition, the backbone carbonyl of this residue forms a hydrogen bond with the side chain of the invariant Arg 23 residue, and this directs the polypeptide strand towards the minor groove.

Residues 65-67 run parallel to, and make contacts with, one strand of the DNA backbone. Residues 68-72, which are invariant in all paired domains, fit directly into the minor groove. In particular: Ile 68 makes hydrophobic contacts with Pro 17 and turns the polypeptide chain towards the bottom of the minor groove. While the precise interactions are not clear in this region, Gly 69, Gly 70, and Ser 71 run along the minor groove of base pairs 12-14. The subsequent region (residues 73-78), which links the N-terminal and C-terminal domains, is visible in our electron density map, but these residues are not well ordered.

Structure of the C-terminal Domain

The C-terminal domain, like the N-terminal domain, contains three α-helices (residues 79-88, 96-106, and 117-124) and has a fold which closely resembles that of the homeodomain and the Hin recombinase. However, this C-terminal domain does not contact the optimized binding site used for cocrystallization. This region also appears more flexible and/or disordered than the N-terminal domain, presumably because it is not constrained by DNA contacts or by extensive crystal packing contacts.

The C-terminal domain includes helices 4 through 6, with helices 5 and 6 resembling a HTH unit. This C-terminal domain can be superimposed reasonably well on the engrailed homeodomain

(rms distance = 1.73 Å for 30 C_α's), the Hin recombinase (rms distance = 1.79 Å for 31 C_α's), and the N-terminal domain of paired domain (rms distance = 1.67 Å for 31 C_α's in the helical regions). However, in comparison with these other proteins, the C-terminal domain of paired has longer "loops" or "turns" between the helices (Figures 1a, 2a). There are seven residues in the turn between helix 4 and helix 5, and there are ten residues in the loop between helix 5 and helix 6.

DNA Conformation

Analyzing the DNA structure with the program of Lavery and Sklenar (Lavery and Sklenar, 1988; Ravishanker et al., 1989) shows that the overall structure of the paired binding site corresponds to that expected for B-DNA. It has an average helical twist of 34.4° (10.5 bp per turn) and an average rise of 3.4 Å per base pair. It has been suggested that paired domains bend DNA when binding their specific DNA sites (Chalepakis et al., 1994), and we see a 20° bend in the region where the β-turn fits into the minor groove (Figure 3a). The localized bend involves a large roll between bp 8 and bp 9, and this may help to accommodate the conserved Phe 12 side chain in the minor groove. There also are interesting variations in groove width. The minor groove is widened in the region recognized by the C-terminal tail (residues 65-72). Most of the major groove has a relatively normal width (~12 Å), but it is surprisingly narrow (8.8-9.9 Å) in the region where helix 3 binds.

Structural Basis of *Pax* Developmental Mutants

The structure reported here is consistent with all of the biochemical data that is available about paired domain-DNA interactions and provides a clear structural basis for understanding missense mutations that result in developmental abnormalities. Biochemical and genetic studies had correctly anticipated that the paired domain would have discrete N-terminal and C-terminal sub-domains (Czerny et al., 1993; Epstein et al., 1994b). Several studies

had indicated that the N-terminal domain provided the most important contacts and actually was sufficient for DNA binding (Treisman et al., 1991; Chalepakis et al., 1991; Czerny et al., 1993). Noting the location of conserved residues and the similarities in the optimized binding sites makes it clear that the structure and DNA docking of the N-terminal domain is highly conserved in the pax family. Comparing the structure with the available sequence data shows that all of the hydrophobic contacts that stabilize the protein and all but one of the DNA contacts are made by residues that are absolutely conserved among all paired domains (Figure 1a). Position 47 is the only variable residue at the protein-DNA interface, but changes at this position correlate with known differences in the optimal binding sites. His 47 recognizes a guanine in Prd, Pax-2, Pax-5 or Pax-8 (Czerny et al., 1993; Epstein et al., 1994a; Jun and Desplan, manuscript in preparation), while Pax-6 has an asparagine at residue 47 and prefers a thymine at the corresponding position in the binding site (Epstein et al., 1994a; Figure 1c). Thus it appears that residue 47 plays an important role in the differential specificity of the Pax proteins.

There also is a remarkable correlation between the observed DNA contacts and the location of missense mutations that result in developmental abnormalities in mice and humans. The mouse developmental mutant *undulated*, which exhibits malformations in the vertebral column, has a missense mutation (Gly 15 -> Ser) (Balling et al., 1988) in the β-turn that contacts the minor groove. Biochemical studies have shown that this mutation dramatically reduces the DNA binding affinity of the Pax-1 protein (Chalepakis et al., 1991), and this Ser also disrupts DNA binding when inserted into the Prd protein (Treisman et al., 1991). The structure shows that this residue lies at the bottom of the minor groove and is too close to accommodate any side chain other than a glycine. Introducing a Gly -> Ser mutation would require the backbone to move and would disrupt other contacts that the β-turn makes in the minor groove. Several of the PAX3 point mutations found in Waardenburg's syndrome patients (Asn 14 -> His; Pro 17 -> Leu; Figure 1b) (Baldwin

et al., 1992; Hoth et al., 1993) also are located in or near this β-turn and further emphasize the importance of the contacts made by the turn. Several other missense mutations map to the N-terminal helical unit, and the structure also provides a basis for understanding these mutants. For example, one form of Waardenburg's syndrome involves a Gly 48 -> Ala mutation (WS .15; Figure 1c) (Tassabehji et al., 1993), and it appears that introducing an alanine at this position would give unfavorable van der Waals contacts or disrupt the docking of the helix-turn-helix unit on the DNA. Two other mutations (Peters' of PAX6 and Bu35 of PAX-3, Figure 1b) (Hoth et al., 1993; Hanson et al., 1994) change the conserved Arg 23 residue which normally contacts both the phosphate backbone and the main chain carbonyl of residue 63. Obviously, introducing Gly or Leu at position 23 would disrupt these contacts. When considering the Pax missense mutations, it is interesting to note that almost all involve changes in residues that contact the DNA (Figure 1b). A priori it would have seemed possible that many of the mutations would disrupt folding (many other Pax mutations involve frameshifts or large deletions), but the missense mutations clearly cluster at the protein-DNA interface. It also is interesting that all the missense mutations map to the N-terminal domain, again indicating that this domain has a very important role in recognition and regulation.

Role of the C-terminal Domain

The C-terminal domain does not make any DNA contacts in the cocrystal structure, and all the available data suggest that this domain is not essential for the Drosophila Prd protein. Thus we note that: 1) The DNA site used for cocrystallization includes all the conserved bases in the optimized binding site. Binding site selections, repeated after the crystal structure was known, were unable to find any sequence preferences outside of the original consensus site we had used for cocrystallization (Jun and Desplan, manuscript in preparation). 2) Methylation interference experiments - using our consensus site embedded in a larger DNA fragment - do

not give any evidence of contacts with neighboring bases (Jun and Desplan, manuscript in preparation). 3) Previous studies had indicated that the first 80 residues of the Prd paired domain were sufficient for site-specific DNA binding (Treisman et al., 1991). 4) Experiments in Drosophila using an ectopic expression assay demonstrated that the C-terminal domain of Prd does not have an essential role in vivo: The Prd protein can still function in vivo when the C-terminal portion of the Prd paired domain is deleted (Cai et al., 1994). 5) A deletion of the C-terminal domain from the Prd paired domain has been shown to have little effect on *Prd* function: Prd mutant flies can be rescued to viability with a *Prd* transgene lacking the C-terminal domain, but exhibit a complete Prd mutant phenotype when the N-terminal domain is disrupted by G15S (*undulated*) mutation (Bertuccioli et al., submitted).

Although the C-terminal domain is not required for the Prd paired protein, there are other Pax proteins in which the C-terminal domain clearly plays an important role in site-specific recognition (Czerny et al., 1993; Epstein et al., 1994a and 1994b). [In considering these differences, one should note that the sequence of the C-terminal domain is significantly less well conserved than the sequence of the N-terminal domain (Figure 1a).] In the case of the Pax-5 protein, interactions between the C-terminal domain and the DNA were demonstrated by methylation interference analysis and by *in vitro* mutagenesis of both the paired domain and its binding site (Czerny et al., 1993). It has also been shown that Pax-6 gives a 26 bp DNase I footprint (Epstein et al., 1994a). Finally studies of a PAX6 splicing variant PAX6-5a also have shown that the C-terminal domain can - after disruption of the N-terminal domain - recognize a distinct set of binding sites (Epstein et al., 1994b).

The structure of the C-terminal domain - which resembles the helical portion of the N-terminal domain, the homeodomain, and the Hin recombinase - certainly is consistent with its having a role in DNA recognition. Helices 5 and 6 form a helix-turn-helix unit that other Pax proteins may use for DNA binding. The rather long loop

between helices 5 and 6 may not be a problem, since studies of other HTH domains have shown that large insertions can be tolerated in the "turn" between the helices (Klemm et al., 1994; Brennan, 1993; Finney, 1990). It also seems plausible that the last four residues of the paired domain (residues 125-128), which are disordered in our electron density maps, may become ordered upon DNA binding (or may be ordered in the context of the full-length protein) and thus may extend the recognition helix. There are numerous examples, including the recognition helices of some homeodomains, where such disorder -> order transitions are coupled with DNA recognition (Qian et al., 1989; Spolar and Record, 1994). Although genetic and biochemical data indicate that the C-terminal portion of the Prd paired domain does not make any critical contacts with the DNA, our structure allows us to predict how the C-terminal domain of Pax-5 and Pax-6 may contact the DNA (Figure 7). As explained in the legend of Figure 7, this model is based on: 1) structural similarities between the C-terminal domain of the Prd protein and the Hin recombinase, 2) the amino acid sequence similarities between Hin and those members of the Pax family which use the C-terminal domain in DNA recognition (see legend to Figure 7), and on 3) modeling constraints imposed by the length of the linker and by the position of the additional base pairs recognized by Pax-5 and Pax-6. In our model (Figure 7), the C-terminal domain of paired binds like Hin, and there is an approximate two-fold axis relating the N-terminal and C-terminal sub-domains. The linker between the subdomains lies in the minor groove, thus extending the minor groove contacts seen in the co-crystal structure. The recognition helix of the C-terminal domain (helix 6) bind in the major groove and is positioned to interact with base pairs 16-20 of the optimized binding sites for Pax-5 and Pax-6.

The β-turn DNA binding motif

Previous studies of protein-DNA complexes have shown how α-helices, β-sheets and regions of extended peptide chain can be used for site-specific recognition of DNA. This is the first

structure to show how a β-turn can play a critical role in protein-DNA recognition. In the paired domain, the β-turn (which is rigidly anchored by neighboring regions of the protein) reaches into the minor groove of the DNA to form direct base-specific hydrogen bonds with guanines 9 and 10, and a water-mediated contact with thymine 11. (It also is interesting to note - as discussed above - that there is a 20° bend in the region contacted by this β-turn.) All paired domains studied show a strong preference for guanine at position 9, and for a cytosine or guanine at position 10 (Epstein et al., 1994a; Czerny et al., 1993; Jun and Desplan, manuscript in preparation; Fig. 1c). The structure provides an explanation for this specificity: the side chain of the conserved asparagine 14 forms a hydrogen bond with the 2-amino group of guanine 9, and the main chain carbonyl of glycine 15 forms a hydrogen bond with the 2-amino group of guanine 10 (Fig. 3b). The hydrogen bond contact with the 2-amino group can readily distinguish guanine from adenine and thymine, which do not have hydrogen bond donors in the minor groove. In Pax-5, a point mutation changing guanine to thymine at position 10 of its binding site decreases the binding affinity by about 40 fold, the largest observed affinity loss in the binding site saturation mutagenesis experiment (Czerny et al., 1993). Cytosine is allowed at position 10 since the GC -> CG change only gives small directional and positional differences in the hydrogen bond with the 2-amino group (Seeman et al.). The biological importance of the β-turn/DNA contacts is well demonstrated by the clustering of Pax point mutations in and adjacent to the β-turn (Fig. 1b).

The β unit that preceeds the N-terminal HTH unit of paired domain and the C-terminal tail that follows are critical for recognition (Treisman et al., 1991; Chalepakis et al., 1991; Czerny et al., 1993). Several other HTH proteins use flanking regions to contact the minor groove. Specifically: 1) the position of the critical β-turn in the paired domain corresponds with the position of the N-terminal arm in the engrailed homeodomain (Kissinger et al., 1990); 2) the Hin recombinase has both an N-terminal arm and a C-terminal tail that contact the DNA (Feng et al., 1994); and 3) the

helical region of HNF3 also has flanking β units (Clark et al., 1993). However, comparison of these β units reveals that the structures and DNA contacts of these other proteins are significantly different, and the paired domain provides the first example of how a β-turn can be used for minor groove recognition of DNA. [The closest analogue may involve a β-turn in glutaminyl-tRNA synthetase that interacts with the minor groove of tRNA. This also has a hydrogen bond between a carbonyl oxygen from the protein backbone and the 2-amino of a guanine (Rould et al., 1989)].

Paired Folds like a Homeodomain but Docks on DNA like λ Repressor

The overall fold of both the N- and C- terminal helical regions of paired resemble the fold of the homeodomain (Kissinger et al., 1990; Qian et al., 1989) and are remarkably similar to the fold of the Hin recombinase (Feng et al., 1994). In comparing the N-terminal region of the paired domain with these other proteins, we find that helices 1, 2 and 3 of the paired domain can be superimposed on the engrailed homeodomain with an rms distance of 1.71Å for 43 C_α's (with two gaps) and can be superimposed on the Hin recombinase with an rms distance of 1.28 Å for 38 contiguous C_α's.

The homeodomain and the λ repressor have been shown to bind their DNA sites in fundamentally different ways (Kissinger et al., 1990; Otting et al., 1990). Residues near the N-terminal end of the recognition helix make critical contacts in the λ repressor-operator complex, while the critical residues in homeodomain-DNA complexes are near the center of an extended recognition helix (Jordan and Pabo; 1988; Qian et al., 1989; Kissinger et al., 1990; Wolberger et al., 1991; Klemm et al., 1994; Fig. 7). The paired domain provides an interesting "missing link" in these comparisons. The docking of the paired HTH unit is distinctly different from the homeodomain but is surprisingly similar to that of Hin and the λ repressor (Fig. 7). Like the λ repressor, the first helix of the paired domain HTH unit (helix 2) fits partway into the DNA major groove, and the N-terminal end of this helix contacts the sugar-phosphate backbone of the DNA. It

appears that the length of helix 2 may be particularly important in distinguishing the alternative docking arrangement seen with the homeodomains: homeodomains have several additional residues at the N-terminus of helix 2, and these would collide with the DNA backbone if the HTH unit docked in the same way as λ, Hin and Prd. Curiously, helix 3 of paired domain (the "recognition helix") fits more deeply into the major groove than do other known recognition helices, and the glycine at position 48 facilitates this close approach. The paired structure helps us understand these family/subfamily relationships and superimposing the complexes in this way (Figure 7) highlights the differences in the way that the helix-turn-helix units are used.

Conclusions

The crystal structure, in conjunction with the available biochemical and genetic data, reveals the key features of paired domain-DNA interactions and provides a structural basis for understanding the known Pax developmental mutants. In particular, we conclude that:

The paired domain contains two structurally independent, globular sub-domains. The N-terminal domain is most highly conserved and makes very important contacts with the DNA. A β-turn near the start of this domain makes critical contacts in the minor groove, and a helix-turn-helix unit makes critical contacts in the major groove.

The structure and contacts of this N-terminal domain are relevant for understanding the entire family of Pax proteins. Residues that form the hydrophobic interior and residues that contact the DNA are remarkably conserved. All of the known point mutations mapping to the paired domain involve changes in the N-terminal sub-domain, and most of these change critical residues at the protein-DNA interface.

For this particular protein - the Prd paired domain - the genetic and biochemical data indicate that the C-terminal domain does not play any essential role in DNA recognition. The structure is consistent with these observations, as the N-terminal domain makes all of the contacts with the optimized binding site. However, the structure of the C-terminal domain and the way that it is tethered to the rest of the complex suggest how this domain may be used to contact the DNA in other paired domain-DNA complexes. In particular, sequence similarities and structural homology suggest that the C-terminal domain may also dock like Hin, giving an overall paired domain/ DNA complex with an approximate two-fold axis relating the N-terminal and C-terminal domains in the complex.

Further crystallographic studies will be needed to understand the precise role of the C-terminal domain in other complexes, but this co-crystal structure provides a firm basis for understanding the fundamental principles of paired domain-DNA interactions and for understanding the known Pax developmental mutations.

Experimental Procedures

A plasmid expression vector with an N-terminal polyhistidine tag, pET14bprdPDB (S. J. and C. D., manuscript in preparation), was used to express the *Drosophila* Prd paired domain in E. coli strain BL21(DE3). Cells were grown at 37° and were induced with 0.4 mM isopropyl-ß-D-thiogalactoside (IPTG) when they reached OD_{600}=0.8. Cells were harvested 3 hours after induction, washed with prechilled phosphate-buffered saline buffer, frozen in a dry-ice/ethanol bath and stored at -80°C. Sonication was carried out in a buffer containing 25 mM Hepes pH 7.6, 0.1M KCl, 0.1% NP-40, 0.3 mg/ml lysozyme, 7 mM 2-mercaptoethanol, 1 µg/ml aprotinin, 1 µg/ml pepstatin, 1 µg/ml benzamidine, and 1 µg/ml sodium metabisulfite. The cell lysate was diluted with solution A (25 mM Hepes pH 7.9, 0.1 M NaCl, 5 mM MgCl2, 15% glycerol, 0.1% NP-40, 7 mM 2-mercaptoethanol) and loaded onto a Ni-NTA column (Novagen). The column was extensively washed with 8 mM imidazole (pH 8.0) in solution A, and then with 40 mM imidazole in solution A; the Prd paired domain was eluted with 100 mM imidazole in solution A. The eluted protein was treated with 0.25U/µl thrombin at 30°C for 15-20 hours to remove the N-terminal polyhistidine tag, and the reaction was stopped by adding 1 mM PMSF to the solution. The Prd paired domain was purified with a Mono-S column (Pharmacia), using a gradient of 0.3 M to 0.7 M NaCl in 40 mM phosphate buffer (pH 6.6), containing 1 mM DTT. The purified protein gave a single band on an overloaded SDS gel in the absence of reductant. The protein used for crystallization was further purified by gel filtration and by reverse phase HPLC, and then was lyophilized and stored at -80°C. The chemical homogeneity and identity of the purified Prd paired domain was further confirmed by N-terminal sequencing, amino acid analysis, mass spectrometry, and gel shift experiments. DNA oligonucleotides used for crystallization were purified as described elsewhere (Klemm et al., 1994).

Preliminary studies revealed that the solubility of the Prd paired domain - DNA complex was very sensitive to ionic strength.

Crystals with the DNA oligo shown in Figure 1d were grown by the evaporation of volatile salts from the hanging drops. Drops initially contained 0.49 mM Prd paired domain, 0.62 mM of the DNA duplex, 0.15-0.2 M ammonium acetate (pH 7.0), 10 mM bis-tris-propane (pH 7.0), 10 mM $MgCl_2$, 0.1 mM EDTA, 5 mM DTT and 0.5% PEG 1000, and these drops were equilibrated against a reservoir containing 10% PEG 1000 and 5% DTT. Crystals grew in 4 to 5 days, but there appeared to be gradual changes in the cell dimensions, and crystals were allowed to "age" for about two weeks before being used for data collection.

Co-crystals diffracting to 2.5 Å resolution grow in orthorhombic space group $P2_12_12_1$, with a=39.6 Å, b=68.6 Å, c=100.5 Å. Data were collected at room temperature on an R-axis image plate system, and reduced using DENZO and SCALEPACK (Z. Otwinowski, personal communication). Derivative data sets were local scaled to the native data set using MAXSCALE (M. A. R.), and heavy atom sites were determined with the program HASSP (Terwilliger et al., 1987). Refinement of heavy atom parameters was carried out using REFINE from CCP4 (The SERC Collaborative Computing Project No.4, a Suite of Programs for Protein Crystallography [Distributed from Daresbury Laboratory, Warrington WA4 4AD, UK, 1979]), followed by cross-phased refinement using PHARE (CCP4). The initial MIR map (mean figure of merit 0.59) was solvent flattened (Wang, 1985), and the heavy atom parameters were then refined using these solvent flattened phases (Rould et al., 1992). The new MIR map (mean fom = 0.71) was subject to another round of solvent flattening to give the final electron density map (Figure 6, mean fom = 0.79). All of the DNA was clearly resolved in this map, as were almost all the sidechains and mainchain carbonyl groups of the N-terminal domain of the protein (Figure 6). The electron density for the C-terminal domain was not as good (it is packed less rigidly in the crystal), but about half of the sidechains of this globular sub-domain were clear. The initial model was built using TOM FRODO (Israel, M., Chirino, A. J. and Cambillau, C. M., personal communication) and subject to multiple rounds of

positional refinement (Brünger, 1992a) and manual adjustment. Refinement was monitored by following the free R-factor to avoid overbuilding (Brünger, 1992b). In later stages of refinement, tightly restrained individual B-factors were used. Local scaling of the observed and calculated structure factors (using a minimum neighborhood of 100 reflections and excluding the reflection being scaled) was also done to correct for absorption and anisotropic diffraction. In the final cycle, 16 water molecules were included in the model. All of the key contacts and the key features of the complex were confirmed by checking simulated annealing omit maps (Hodel et al., 1992). About 30% of the sidechains of the C-terminal domain could not be built with confidence and were modeled as alanines; the first 5 and last 4 residues of the polypeptide also were omitted. (A few of these N-terminal residues were ones introduced during cloning, and thus our model includes residues 2-124 of the paired domain.) Our current model has an R factor of 23.4% and a free R factor of 28.4% with good stereochemistry (Table 1). All phi and psi angles, except for residues 78 (in the linker) and 91 (in an extended loop), are in allowed regions of the Ramachandran plot.

Acknowledgements

This work was supported by NIH grant GM-31471 (C. O. P.) and by the Howard Hughes Medical Institute, and used equipment purchased with support from the PEW Charitable Trusts. We thank Richard Maas and Jonathan Epstein for many helpful discussions and for comments on this manuscript; Lena Nekludova for help in analyzing the DNA structure and the docking arrangement and for help with the figure 8; Amy Dunn for help in preparing this manuscript; Guojun Sheng and members of the Pabo lab for helpful discussions and for critical reading of the manuscript. M.A.R., C.D., and C.O.P. are in the Howard Hughes Medical Institute. S.J. is supported by William O. Baker Fellowship from Mellon Foundation. Coordinates are being deposited with the Brookhaven Data Bank. While they are being processed, interested scientists may obtain a set of coordinates by sending an e-mail message to pabo@pabo1.mit.edu.

Figure Legends

Figure 1

Paired domains and their DNA binding sites. **a.** The sequence and secondary structure of the Prd paired domain are shown at the top, and sequences of paired domains from representative proteins are shown below in one letter code (Walther et al., 1991; Stapleton et al., 1993). Dashes indicate the same amino acid as Prd, and dots indicate gaps in the sequence. The numbering corresponds to that of the Prd paired domain. The protein used in our crystallographic study contains the whole Prd paired domain and four additional residues (Gly-Ser-His-Met) on the N-terminal end that were introduced from the expression vector. Invariant residues found in all paired domains are shown below the set of sequences. DNA contacts are indicated on the last two lines, with the first line used to indicate contacts with the sugar phosphate backbone (p), and the second line used to indicate base contacts (M --> major groove contact, m --> minor groove contact). **b**. Missense mutations in paired domains that are associated with developmental abnormalities in mice and in humans (Tassabehji et al., 1992; Baldwin et al., 1992; Balling et al., 1988; Hoth et al., 1993; Tassabehji et al., 1993; Hanson et al., 1993; Vogan et al., 1993). The tilde (~) symbols denote residues different from Prd to PAX3, or PAX6, or Pax-1. Only partial sequences are shown since all known missense mutations of the paired domains map to this region. **c**. DNA binding sites of paired domains. Consensus binding sites for the paired domains of Prd, Pax-2, and Pax-6 were deduced from in vitro selection and amplification experiments (Epstein et al., 1994a; Jun and Desplan, manuscript in preparation). That of Pax-5 was deduced from alignment of functional promotor sequences (Czerny et al., 1993). The numbering scheme corresponds to that used in Figure 1d. **d**. DNA oligonucleotide used for cocrystallization.

Figure 2

Overview of the paired domain-DNA complex. **a**. Stereo view with ribbons drawn through the Cα's of the protein and through the phosphate backbone of the DNA strands. The paired domain is in yellow, and the DNA is in blue. The Pax missense mutations, which all map to the N-terminal domain, are indicated by the red dots at the Cα atoms of residue 9, 14, 15, 17, 23 and 48. This figure was generated with Insight II software from Biosym. **b**. Sketch of the complex in a similar orientation. Cylinders indicate α-helices, and arrows indicate β-sheets. The critical β-turn (residues 13-16) is shaded.

Figure 3

DNA recognition in the minor groove by the β-turn. **a.** DNA backbone contacts made by residues that flank the β-turn and help position it in the minor groove. (The backbone of residues 13-16 has been shaded.) Hydrogen bonds from peptide backbone amides as well as asparagine and glutamine sidechains hold the short antiparallel β-sheet against the DNA. Phenylalanine 12 and proline 17 form hydrophobic surfaces which pack against the ribose rings. The DNA is bent where the β-turn inserts into the minor groove. **b.** Residues of the β-turn participate in recognition of base pairs 9, 10 and 11 via minor groove contacts. The sidechain of Asn 14 and the peptide carbonyl of Gly 15 form hydrogen bonds with the 2-amino groups of guanines 9 and 10. Arg 16 makes a water mediated contact with the thymine at bp 11. **c.** Overview of the novel β-turn motif seen in the paired complex. The Cα trace of residues 4-12 and 17-18 is shown in yellow. The trace of residues 13-16 and the side chains of Asn 14, Gly 15, Arg 16 are shown in red. Base pairs 9, 10 and 11, which are contacted by the β-turn, are shown in white.

Figure 4

Hydrogen bonds between the N-terminal helical unit (residues 20-60) and the DNA. Most of these hydrogen bonding interactions involve contacts with the DNA backbone. There are two important

sidechain-base interactions: The sidechain of His 47 forms a hydrogen bond with the O6 of the guanine at bp 4 and Lys 52 hydrogen bonds to the N7 of the guanine at bp 8.

Figure 5

Sketch summarizing hydrogen bonding interactions between the Prd paired domain and DNA. The DNA is represented as a cylindrical projection. Circles denote the phosphates, and hatched circles indicate positions where there are bonds with the sugar phosphate backbone. (MC denotes peptide main chain.)

Figure 6

Section of the original 2.5 Å resolution solvent-flattened MIR electron density map showing the interface between the HTH unit and the DNA. The protein is in yellow, the DNA in red, and the electron density is shown in blue. The map is contoured at 1.8 rms above the average electron density.

Figure 7

Model indicating how the C-terminal domain of Pax-5 and Pax-6 may contact DNA. The N-terminal domain (shown in purple) binds as observed in our crystal structure. Our model for the overall docking arrangement of the C-terminal region (shown in red) is based on sequence and structural homology with the Hin recombinase. Two regions of sequence homology suggested this model: 1) The linker between the two domains of Prd (residues 70-77, GSKPRIAT) is similar to the N-terminal arm of the Hin recombinase (residues 139-145, GRPRAIT). Since the N-terminal arm of Hin binds in the minor groove, we used it as a guide when modeling residues 71-78 of the paired linker. 2) Sequence homology between the recognition helix of Hin (residues 173-179, VSTLYR) and the helix 6 region of Pax-6 (residues 117-123, VSSINR) suggests that these helices may have

similar binding modes, and the Hin complex was used as a guide for docking residues 79-124 from Prd. The base pairs in the corresponding region of the optimized site recognized by Pax-6 (Epstein et al.,1994a; base pairs 16-20 of Figure 1c) are highlighted in gray.

Figure 8

Stereo view of the two distinct modes of DNA docking used by 1) homeodomain proteins and by 2) the λ-repressor, Prd N-terminal domain and Hin recombinase. Complexes were aligned by superimposing 18 C_α's common to all the HTH units, and therefore the docking arrangements can be compared by comparing the position of the corresponding DNA duplexes. The docking arrangement for the paired N-terminal domain (purple) is quite similar to that of the λ repressor and the Hin recombinase (both are in blue). Docking arrangements for the engrailed and α2 homeodomains (both in red) appear to define a separate class of docking arrangements. [Note: Differences in the lengths of the helices are not apparent in this figure since we only show regions that are common to the set of helical units.]

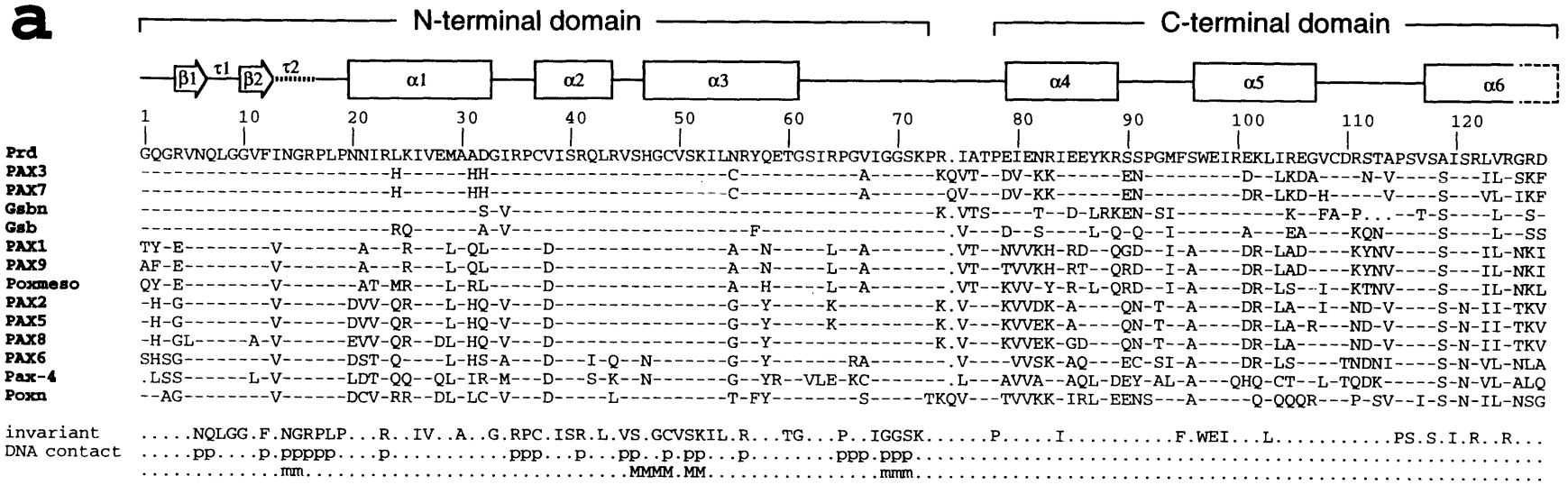
a
N-terminal domain
C-terminal domain
β1 τ1 β2 τ2 α1 α2 α3 α4 α5 α6
1 10 20 30 40 50 60 70 80 90 100 110 120
Prd GQGRVNQLGGVFINGRPLPNNIRLKIVEMAADGIRPCVISRQLRVSHGCVSKILNRYQETGSIRPGVIGGSKPR.IATPEIENRIEEYKRSSPGMFSWEIREKLIREGVCDRSTAPSVSAISRLVRGRD
PAX3 ----------H------HH----------C----------A------KQVT--DV-KK------EN---------D--LKDA----N-V----S---IL-SKF
PAX7 ----------H------HH----------C----------A-------QV---DV-KK------EN---------DR-LKD-H-----V----S---VL-IKF
Gsbn ----------S-V----------K.VTS----T--D-LRKEN-SI----------K--FA-P...--T-S----L--S-
Gsb ----------RQ------A-V----------F----------.V---D--S----L-Q-Q--I------A---EA----KQN-----S----L--SS
PAX1 TY-E--------V-------A---R---L-QL-----D----------A--N-----L--A--------.VT--NVVKH-RD--QGD--I-A----DR-LAD----KYNV----S---IL-NKI
PAX9 AF-E--------V-------A---R---L-QL-----D----------A--N-----L--A--------.VT--TVVKH-RT--QRD--I-A----DR-LAD----KYNV----S---IL-NKI
Poxmeso QY-E--------V-------AT-MR---L-RL-----D----------A--H-----L--A--------.VT--KVV-Y-R-L-QRD--I-A----DR-LS--I--KTNV----S---IL-NKL
PAX2 -H-G--------V------DVV-QR---L-HQ-V---D----------G--Y-----K---------K.V---KVVDK-A----QN-T--A----DR-LA--I--ND-V----S-N-II-TKV
PAX5 -H-G--------V------DVV-QR---L-HQ-V---D----------G--Y-----K---------K.V---KVVEK-A----QN-T--A----DR-LA-R---ND-V----S-N-II-TKV
PAX8 -H-GL-----A-V------EVV-QR--DL-HQ-V---D----------G--Y----------K.V---KVVEK-GD---QN-T--A----DR-LA-----ND-V----S-N-II-TKV
PAX6 SHSG--------V------DST-Q----L-HS-A---D---I-Q--N-------G--Y--------RA--------.V----VVSK-AQ---EC-SI-A----DR-LS----TNDNI----S-N-VL-NLA
Pax-4 .LSS------L-V------LDT-QQ--QL-IR-M---D---S-K--N-------G--YR--VLE-KC--------.L---AVVA--AQL-DEY-AL-A---QHQ-CT--L-TQDK-----S-N-VL-ALQ
Poxn --AG--------V------DCV-RR--DL-LC-V---D-----L----------T-FY--------S-----TKQV---TVVKK-IRL-EENS---A-----Q-QQQR---P-SV--I-S-N-IL-NSG
invariant .....NQLGG.F.NGRPLP...R..IV..A..G.RPC.ISR.L.VS.GCVSKIL.R...TG...P..IGGSK.....P.....I.....F.WEI...L.....PS.S.I.R..R...
DNA contact .....pp....p.ppppp....p.....ppp...p...pp..p.pp...p.....ppp.ppp
mm MMMM.MM mmm
b
1 10 20 30 40 50
Prd GQGRVNQLGGVFINGRPLPNNIRLKIVEMAADGIRPCVISRQLRVSHGCV
PAX3(Sp^d) --------R----------
PAX3(BU47) ----------H----------
PAX3(BU26) ----------L----------
PAX3(BU35) ----------L----------
PAX3(WS.15) ----------A--
Pax-1(un) ----------S----------
PAX6(Peters') ----------G----------
c
1 3 5 7 9 11 13
Prd PD --CGTCACG(G/C)TT(G/C)(A/G)
Pax-2 PD -T-GTCA(T/C)GC(A/G)TGA
Pax-6 PD A--TTCACGC(A/T)T(G/C)A-T(G/T)(A/C)-(T/C)
Pax-5 -(T/C)(G/C)GT(C/T)(A/C)CGC--CA-TG--(C/T)
d
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
AACGTCACGGTTGAC
GCAGTGCCAACTGTT

Figure 1

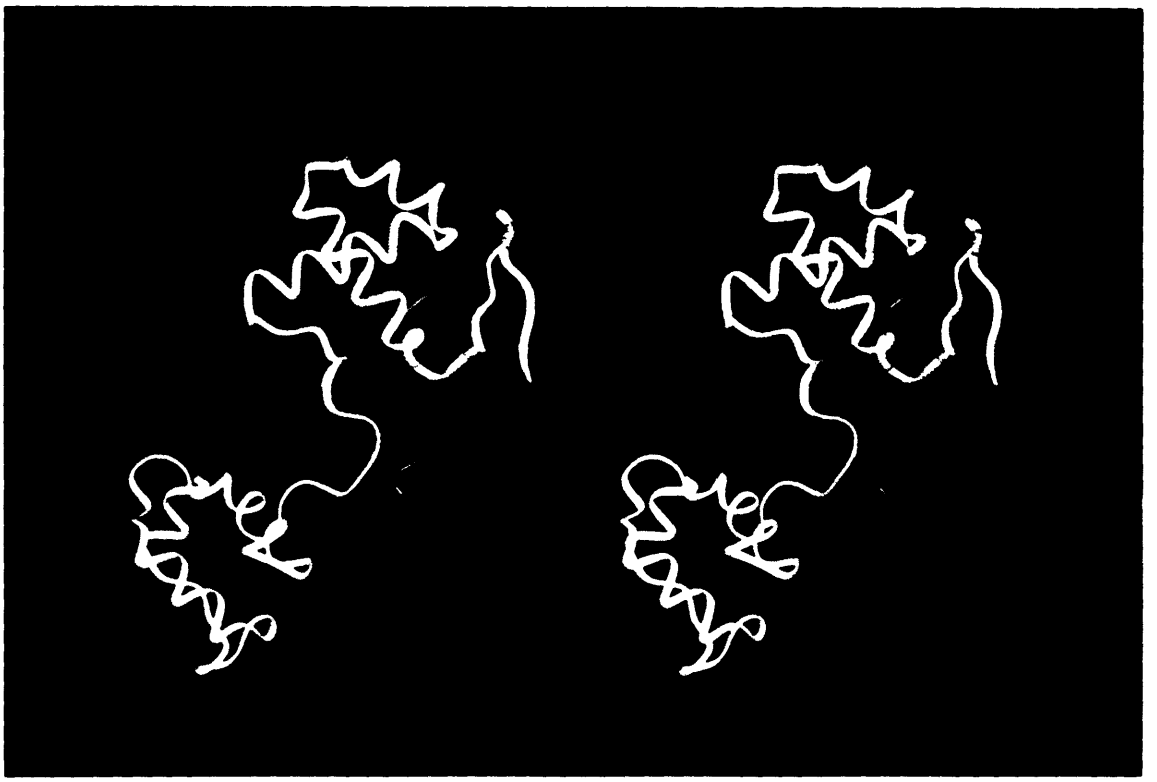

Figure 2a

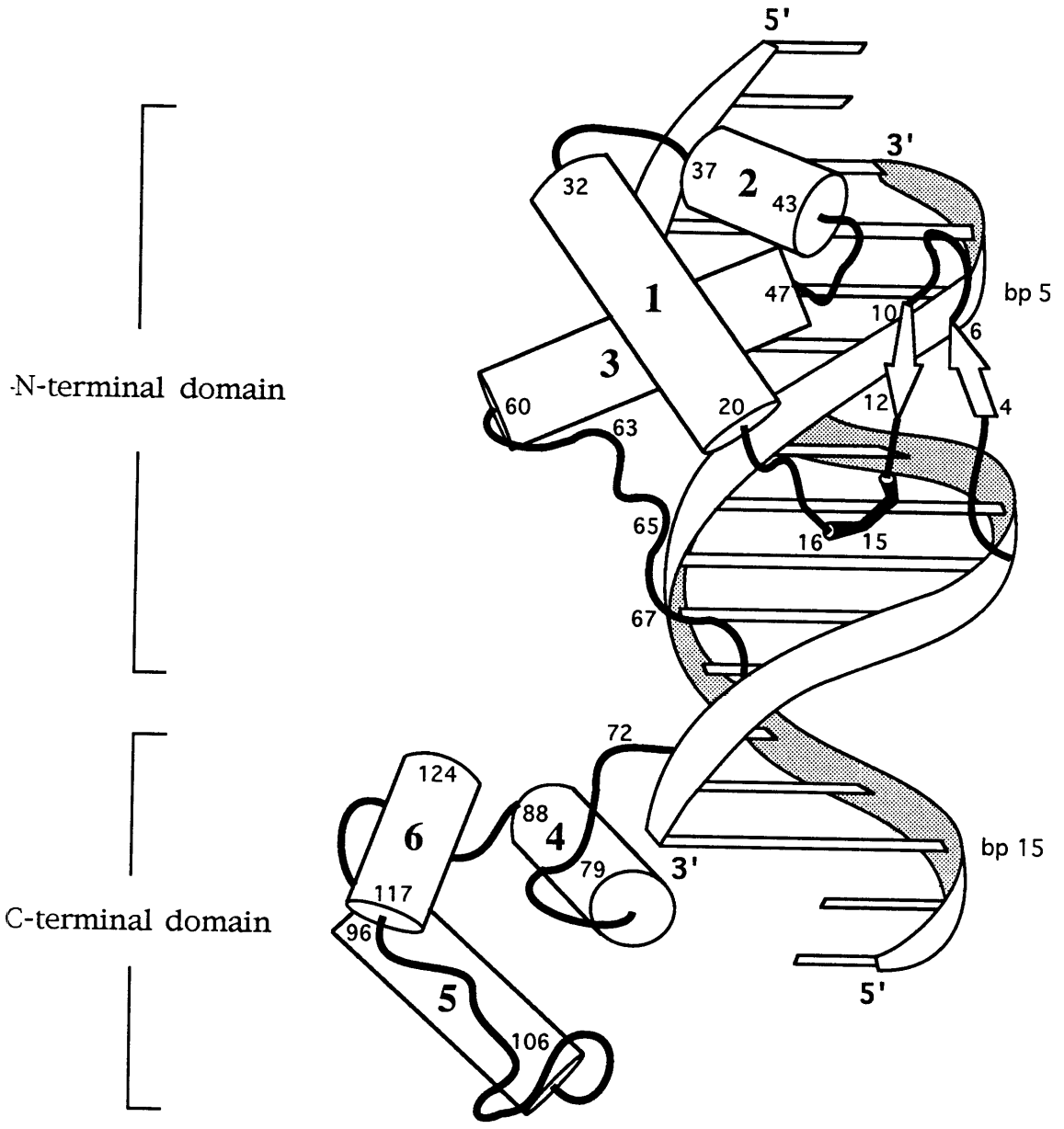
N-terminal domain
C-terminal domain
5'
3'
32
37
2
43
1
47
10
6
bp 5
3
60
20
12
4
63
65
16
15
67
72
124
6
88
4
bp 15
79
3'
117
96
5
106
5'

Figure 2b

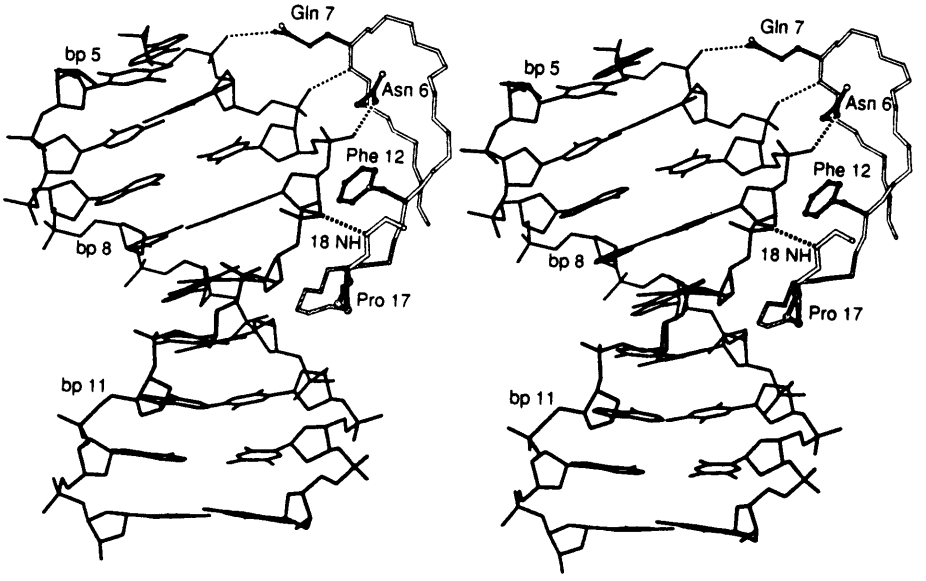
Gln 7
bp 5
Asn 6
Phe 12
bp 8
18 NH
Pro 17
bp 11
Gln 7
bp 5
Asn 6
Phe 12
bp 8
18 NH
Pro 17
bp 11

Figure 3a

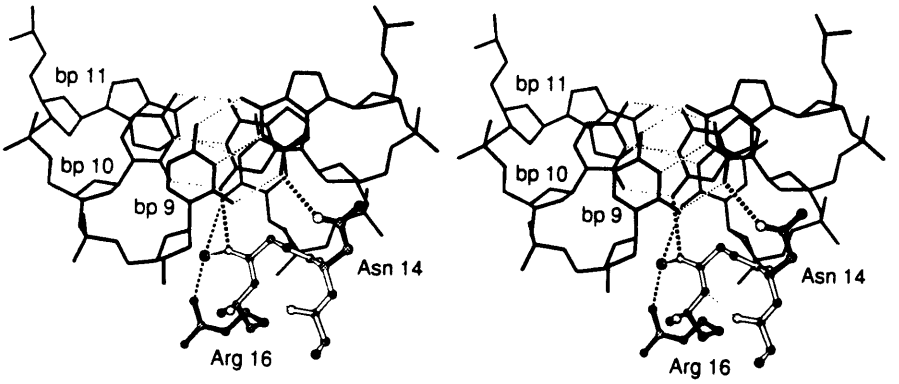

Figure 3b

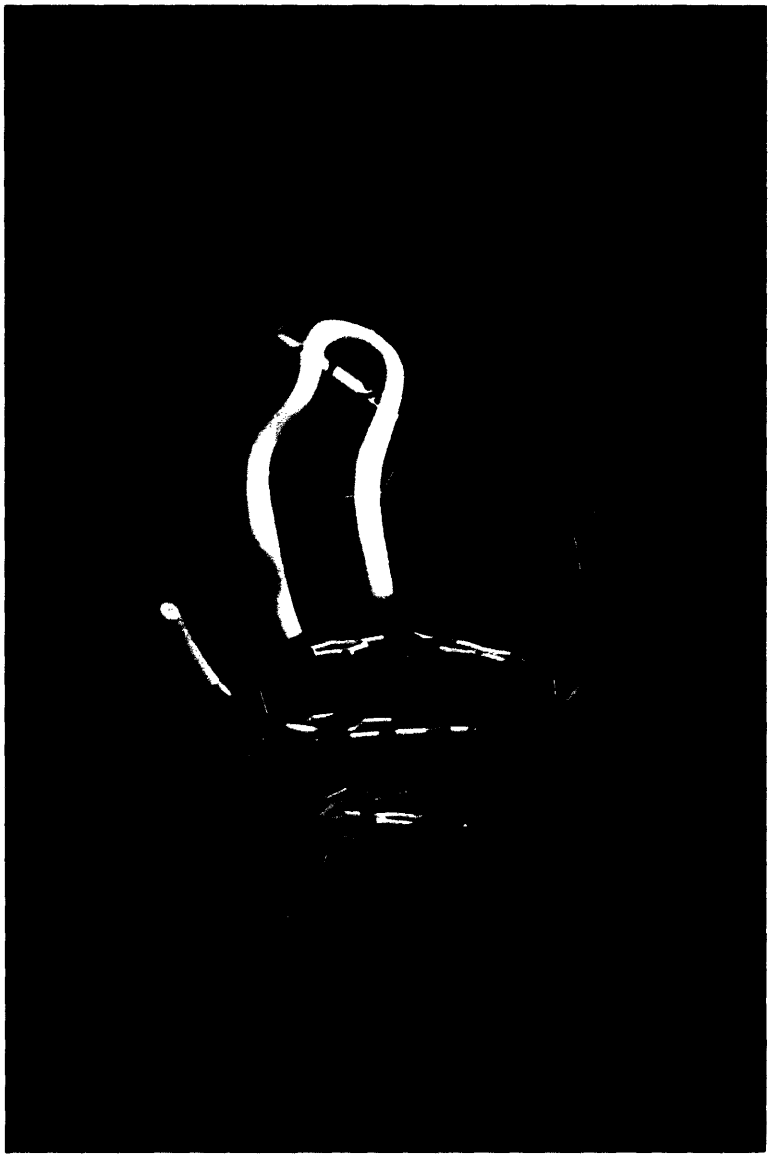

Figure 3c

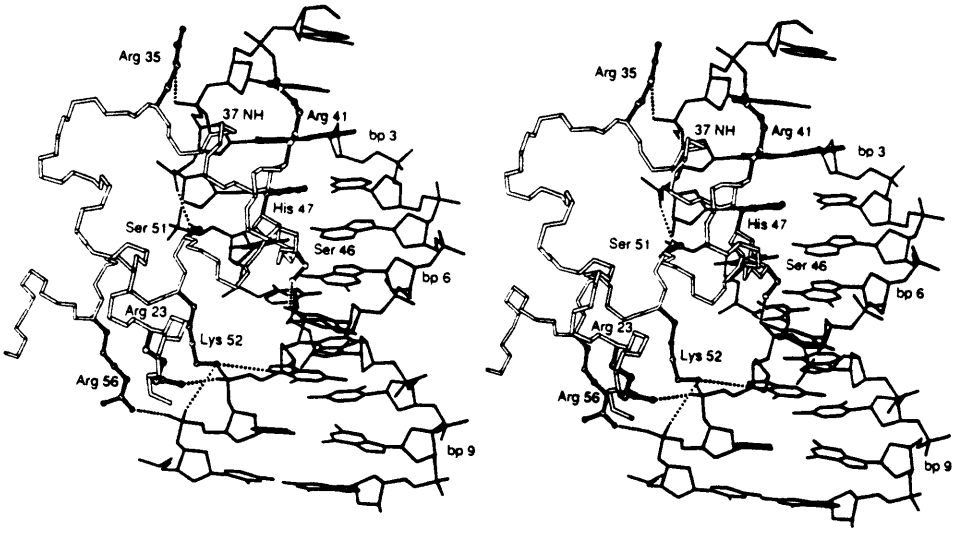
Arg 35
37 NH
Arg 41
bp 3
His 47
Ser 51
Ser 46
bp 6
Arg 23
Lys 52
Arg 56
bp 9
Arg 35
37 NH
Arg 41
bp 3
His 47
Ser 51
Ser 46
bp 6
Arg 23
Lys 52
Arg 56
bp 9

Figure 4

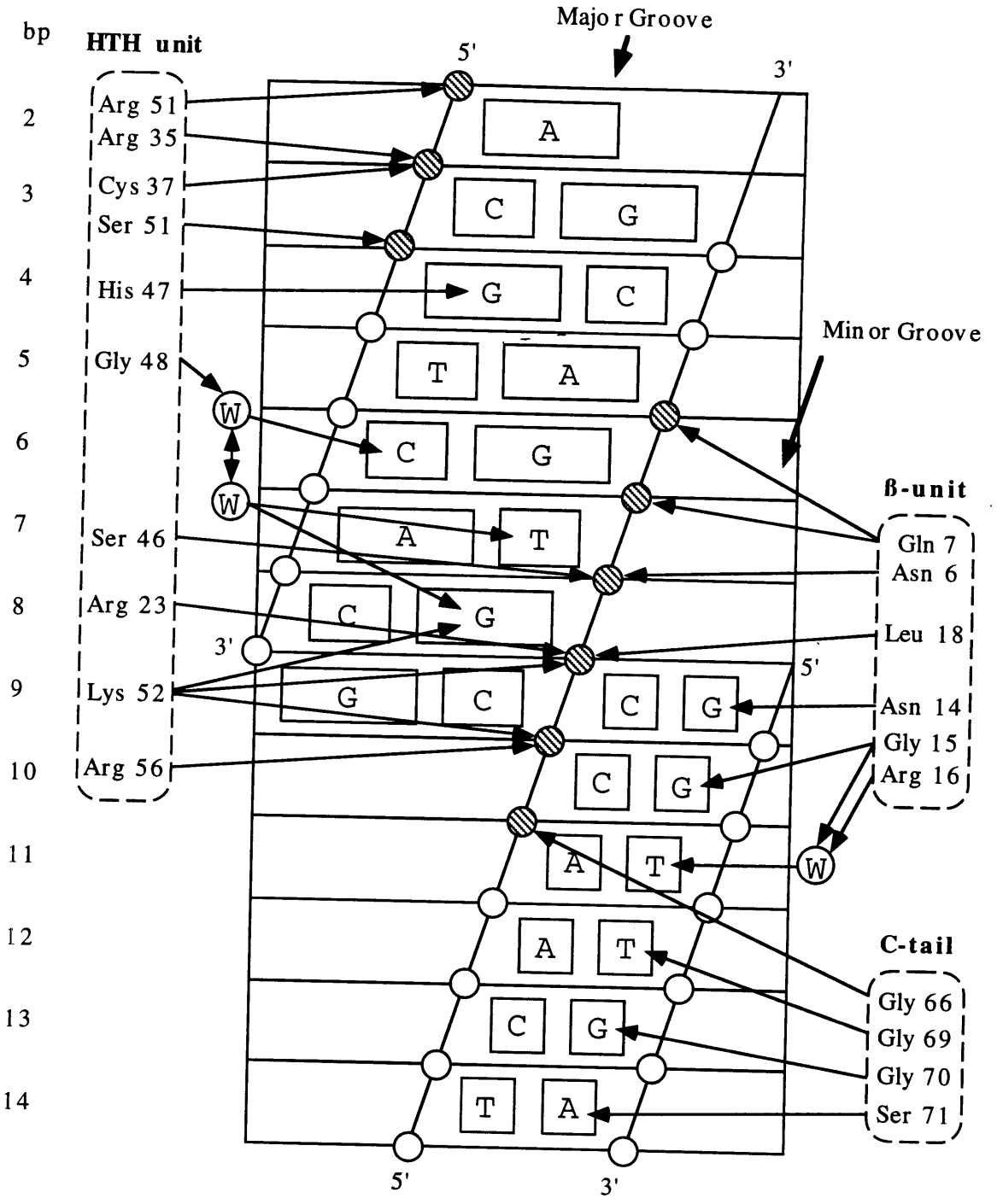
bp
HTH unit
Major Groove
5'
3'
Arg 51
Arg 35
Cys 37
Ser 51
His 47
Gly 48
Ser 46
Arg 23
Lys 52
Arg 56
2
3
4
5
6
7
8
9
10
11
12
13
14
A
C
G
T
W
Minor Groove
ß-unit
Gln 7
Asn 6
Leu 18
Asn 14
Gly 15
Arg 16
C-tail
Gly 66
Gly 69
Gly 70
Ser 71

Figure 5

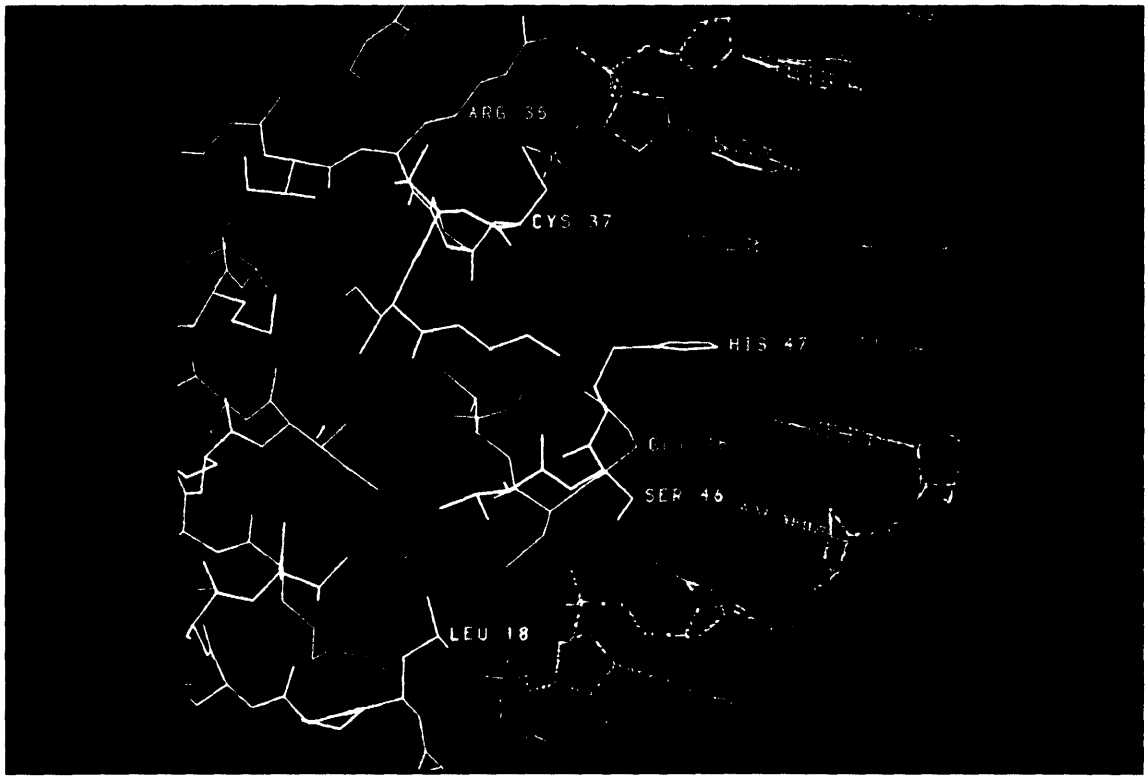

Figure 6

Figure 7

Figure 8

Table 1. MIR Phasing and Crystallographic Refinement Statistics

	Native	dIU(11)	dIU(14)	dIU(12+14)
Resolution (Å)	2.5	2.5	2.5	2.5
Measured reflections	40156	19477	32699	31129
Unique reflections	9285	7446	9822	9566
Data coverage (%)	92.5	66.0	98.4	95.7
Rsym	7.5	7.2	8.2	6.8
Cullis R factor		0.69	0.66	0.47
Phasing power		1.63	1.66	3.24

Refinement	
Resolution (Å)	20-2.5
R factor	0.234
Free R factor	0.284
Non-Hydrogen Atoms	1509
rms ΔB for bonded atoms ($Å^2$)	2.2

Deviations from Ideal Stereochemistry	Protein	DNA
rms bond length (Å)	0.005	0.017
rms bond angles (°)	1.1	3.5

Designations for the derivative data sets indicate the base(s) at which 5-Iodo-uracil was substituted for thymine.

Rsym = $\Sigma_h\Sigma_i|I_{h,i}-I_h| / \Sigma_h\Sigma_i I_{h,i}$ where I_h is the mean intensity of the i observations of reflection h.

Cullis R factor = $\Sigma||F_{PH}\pm F_P|-F_{H,calc}| / \Sigma|F_{PH}\pm F_P|$ (centric reflections only)

Phasing Power = $\sqrt{[\Sigma F_{h,calc}^2 / \Sigma(F_{PH,obs}-F_{PH,calc})^2]}$

Free R factor = $\Sigma||F_{obs}|-|F_{calc}|| / \Sigma|F_{obs}|$, for a 10% subset of all reflections that were never used in crystallographic refinement (Brünger, 1992b).

R factor = same as Free R factor, but only for the remaining 90% of the reflections used in crystallographic refinement (Brünger, 1992b).

Ideal stereochemical parameters for protein refinement are from (Engh and Huber, 1991); for DNA, ideal parameters are from PARAM11X.DNA of the standard XPLOR library (Brünger, 1992a).

Chapter 4

Structural Basis of the Specificity: Pax Binding Sites, Protein-DNA Contacts, and Pax Developmental Mutations

Although some implications and analysis of the paired domain-DNA complex were mentioned in Chapter 3, space limitations prevented more detailed discussion. In the first section of this chapter we will provide a more explicit discussion about the DNA binding sites and the DNA binding specificity of the paired domain. In the second section, we will further discuss the structural basis of *Pax* developmental mutations.

Structural Basis of the Specificity of Paired Domain-DNA Interaction

DNA-binding Sites of Paired Domain

DNA binding sites for the paired domain have been identified both by selection from random DNA and analysis of natural promoters. Optimal binding sites have been selected for paired domains of Pax-2, PAX6 and Prd (Epstein et al. 1994a; S. Jun and C. Desplan, manuscript in prepration). Several functional target sequences have been identified for the Pax subfamily containing Pax-2, Pax-5 and Pax-8. The *CD-19* gene, which encodes for a B-cell surface protein, was discovered to be a mouse Pax-5 (BSAP) target gene (Kozmik et al. 1992). Additional Pax-5 target sequences were identified in the *blk* promoter (Zwollo and Desiderio 1994) and in the vicinity of the immunoglobulin heavy-chain gene switch regions (αS-1, Waters et al. 1989; 5'Sγ2a, Liao et al. 1992; Sγ1, Williams and Maizels 1991; Iε, Rothman et al. 1991). TSAP, the sea urchin Pax-5 homolog, binds to and regulates each of the promoters of two pairs of nonallelic histone *H2A-2* and *H2B-2* genes in sea urchin (Barberis et al. 1989). Pax-8 was found to bind to and regulates the *N-CAM* genes (Holst et al. 1994), the *thyroglobulin* gene (Tg) and the *thyroperoxidase* (TPO) gene (Zannini et al. 1992). Recently, a number of DNA sites have been implicated as potential Pax-6 targets, including the promotor sites of the *αA-crystallin* genes (Cvekl et al., 1994), the *δ1-crystallin* gene (Cvekl et al., 1995), the *ζ-crystallin* gene (Richardson et al., 1995), the gene of neural cell adhesion molecule L1 (Chalepakis et al. 1994b) and P1 and P0 sites of *PAX-*

QNR (Plaza et al., 1995). Based on above sequences and the alignment described by Czerny and Busslinger (1995), we compiled the consensus binding sites for the Pax subfamily containing Pax-2/5/8 (Figure 4).

Pax proteins have been divided into different subfamilies (see discussion in chapter 1). This classification is also nicely reflected at the level of DNA binding specificity. When we compare above *in vitro* selected sites and *in vivo* functional sites, it is clear that: 1) the 5' sub-sites bound to the N-subdomain of paired domain are more conserved, in agreement with the observation that the N-subdomain plays a dominant role in paired domain-DNA interaction; 2) the major difference in binding specificity among various Pax sub-families is at base pair 4. While PAX6 obviously prefers a thymine, Prd, Pax-2, Pax-5 and Pax-8 all prefer a guanine in this position; 3) none of the *in vivo* functional sites completely conform to the consensus sequence.

Based on their ability to bind to a truncated Pax-5 peptide lacking 36 carboxyl amino acids in the paired domain, Pax-5 functional sites were grouped into two classes (Czerny et al., 1993). Comparison of class II recognition sequences, which interact with Pax-5 paired domain lacking a functional C-terminal domain, revealed that all of these sequences are very similar in 5' sub-site. Class I sequences, which are recognized by intact Pax-5 but not the C-terminal truncated Pax-5 paired domain, match less well to 5' sub-site but contain an invariant TG dinucleotide in 3' sub-site, whereas the same two nucleotides are absent in all class II sites. Some Pax proteins, such as Prd, PAX3, PAX7, gsb and poxn, may bind exclusively to class II DNA binding sites, since the longer class I binding sites for these Pax proteins have never been identified.

Paired Domain-DNA Binding Specificity as Observed in Prd Structure

The crystal structure of Prd paired domain-DNA complex strongly suggested that residue 47 plays an important role in the

differential specificity of the Pax proteins. The residue 47 is the only variable residue in the protein-DNA interface and contacts the base pair 4, which shows the only major difference in binding specificity among the various Pax sub-families (Xu et al., 1995). Recently an independent peptide-swapping and mutagenesis study has shown that simutaneous mutation in 3 varible residues in the paired domain (residue 42, 44, 47), from those of Pax-6 to corresponding residue of Pax-5, is sufficient to switch the DNA binding specificity from Pax-6 to Pax-5 (primarily the preference at the divergent position 4) (Czerny et al. 1995). (Mutagenesis in each single position was not tested). This result corresponds beautifully to our crystal structure.

The invariant amino acid Lys 52 contacts two phosphates and the N7 of guanine at base pair 8. Mutation of this guanine to cytosine or thymine would eliminate this hydrogen bond.

In base pair 9 and 10, GC are strongly prefered by all Pax proteins tested (Figure 4). In the Prd structure, the side chain of the conserved asparagine 14 forms a hydrogen bond with the 2-amino group of guanine 9, and the main chain carbonyl of glycine 15 forms a hydrogen bond with the 2-amino group of guanine 10. The hydrogen bond contact with the 2-amino group can readily distinguish guanine from adenine and thymine, which do not have hydrogen bond donors in the minor groove. The change between G and C may be allowed at position 9 and 10 since it only gives small directional and positional differences in the hydrogen bond with the 2-amino group (Seeman et al., 1976). Several lines of evidence indicate that the contacts between base pair 9 and 10 and the β-turn in the minor groove are critical to paired domain-DNA interaction: 1) In Pax-5, a point mutation changing guanine to thymine at position 10 of its binding site decreases the binding affinity by about 40-fold, the largest observed affinity loss in the binding site saturation mutagenesis experiment (Czerny et al., 1993). 2) A binding site mutagenesis experiment with the Prd optimal site also showed similar results (Susie, J., personal communication). 3) Base pair 9 and 10 are

absolutely conserved in Pax-5 class II binding sites, and base pair 10 is the only invariant position in all Pax-5 sites (Czerny et al., 1993). 4) A potential Pax-6 functional site, L1-170, contains a paired domain site, which matchs the optimal Pax-6 consensus except position 9 and 10 (T and A in L1-170 site) (Chalepakis et al. 1994b; Epstein et al. 1994b). Pax-6 paired domain gel-shift band with this site is basically undetectable (intact Pax-6 protein can bind to this site because of the affinity compensation by the Pax-6 homeodomain). When position 9 and 10 of this site were mutated to CC ("L1 Hox-mut" site), Pax-6 paired domain alone showed a strong shift band (Chalepakis et al., 1994). Furthermore, *Pax-1* undulated mutation (G15S), which presumably would disrupt the β-turn-DNA contacts, showed dramatically reduced DNA binding affinity (Chalepakis et al., 1991).

In our Prd structure, the N-terminal domain forms extensive contacts with DNA backbone. However the recognition helix only form two direct side chain-side chain hydrogen bonds with the base pairs 4 and 8 in the major groove, respectively. In addition, the recognition helix forms van der Waals contacts with base pairs 5, 6 and 7, and two water-mediated hydrogen bonds with base pairs 6 and 7. Correspondingly, we see significant variations in these three positions in the functional sites. Alignment of the binding sites indicates that thymine and guanine are both preferred in position 5, cytosine and thymine are almost equally represented in position 6, while adenine is more preferable to cytosine in base pair 7. It is very interesting to note that among the known class II sites which presumably provide optimal sub-sites for the N-terminal domain, position 5, 6 and 7 are variable while position 4, 8, 9 and 10 are absolutely conserved (Czerny et al. 1993). More structural, mutagenic and thermodynamic studies will be required to fully understand the binding specificity in these three positions, as the local DNA structure and water-mediated contacts may be important for the recognition.

While base pairs 16 to 20 may be recognized by the C-terminal

domain of PAX6 and Pax-5, binding specificity at base pairs 12, 13 and 14 may be exclusively provided by contacts made by the linker between two domains. The N-terminal domain conformation in our Prd structure should account for the interactions of residue 2 to 68 with DNA in all Pax proteins. The Ile 68 forms hydrophobic interface with N-terminal globular domain and leads the peptide chain into the bottom of the minor groove. Residue 69-72, which travels in the bottom of minor groove and make contacts with base pairs, are not as clear as the N-terminal globular region in our electron density map. The conformation of this region may also be affected by the position of the C-terminal domain, which does not make DNA contacts in our Prd structure. Thus the explanation of the linker-DNA binding specificity awaits the determination of a paired domain-DNA complex crystal structure with a DNA site containing both 5' sub-site and 3' sub-site.

Structural Basis of *Pax* Naturally-occurring Developmental Mutations

It has become apparent that *PAX* genes are frequent targets for pathological mutations. Mutations in three mouse *Pax* genes, *Pax-1*, *Pax-3* and *Pax-6,* are known to produce the undulated, Splotch and Small-eye mutant phenotypes, respectively. In addtion, mutations in human *PAX3* cause Waardenburg syndrome type 1 (WS1) and type 3 (WS3), and mutations in *PAX6* cause familial and sporadic aniridia and Peters' anomaly.

After a PAX3-Waardenburg syndrome (WS) linkage study organized by Waardenburg Syndrome Consortium (England), some 40 *PAX3* mutations have been characterized, and it was reported that all WS1 families and none of the WS2 families are linked to *PAX3* mutations (Farrer et al. 1994; Read 1995). These include a variety of truncating mutations (gene deletion, frameshifting deletion or insertion, splicing site alteration and nonsense mutation), which are scattered across the gene. There are roughly equal number of missense mutations, largely in the N-terminal half of the paired box. Except *Splotch delayed* (Sp^d), all mutations discovered so far in mouse Pax-3 gene are truncating mutations. It seems that mouse Pax-3 dosage-dependency is not as sensitive as human PAX3.

PAX6 mutations accounts for most, if not all, cases of autosomal dominant aniridia, for both familiar and sporadic cases. *PAX6* mutations are mostly truncating mutations, and are distributed uniformly across the whole *PAX6* gene.

Several lines of evidence suggest that virtually all of the PAX3 and PAX6 mutations are expected to produce inactive proteins. First, mutations expected to allow expression of some product (missense mutations, small in-frame deletion) produce similar phenotypes to mutations expected to inactivate gene expression completely (complete gene deletion, frame-shifting mutations early in the reading frame) (Strachan and Read, 1994). Second, the

severity of the phenotype can often be correlated with the residual gene expression expected. For example, of the three characterized undulated alleles, the missense mutation undulated (*un*) has the least severe phenotype, the partial gene deletion un^{ex} gives a more severe phenotype, and the most severe allele, un^{s}, has a complete deletion of the *Pax-1* gene (Balling et al., 1988; Chalepakis et al., 1991). Finally, *in vitro* DNA binding studies of the mutant PAX3 proteins and our crystal structure support the loss-of-function model (Chalepakis et al., 1994a; see discussion below) However it is important to test the possibility that truncated proteins make antagonistic interactions with wild type protein or some cellular target, since truncated PAX3 forms are produced normally in some tissues by alternative splicing (Tsukamoto et al. 1994) and dorminant negative effects have been noted in Drosophila ectopically expressing a paired transgene with a precise deletion of the paired domain (Morrissey et al. 1991). In addition, possible specific dominant-negative effects has been suggested to *PAX3* mutations (in the position 14 of paired domain) associated with WS type 3 (Read 1995; A.P. Read, personal communication).

The majority of *Pax* missense mutations are human *PAX3* mutations which are associated with Waardenburg syndrome type 1 or type 3. Missense mutations played an important role in understanding the associations between *Pax* mutations and developmental phenotypes. Furthermore, *Pax* missense mutations involves important functional regions of the protein, and are invaluable for studying the molecular mechanism of specific cellular and physiological functions of Pax proteins. All missense mutation reported are observed in positions that are invariant in all paired domains. Most of them are expected to be loss-of-function mutations, and most are in the N-terminal half of paired domains. Our crystal structure provides the structural basis for understanding these point mutations. Interestingly, in our structure, most missense mutations in the paired domain are found in protein-DNA interface, and involve changes that would be expected to disrupt the DNA contacts. In particular, missense mutations are clustered

in or near the β-turn in the minor groove which plays a very important role in paired domain-DNA interaction (Figure 5).

Pax-1 Mutation and Vertebral Column Development

The mouse developmental mutant *undulated* has a missense mutation in the *Pax-1* gene (Gly 15 -> Ser) (Balling et al., 1988) in the β-turn that contacts the minor groove. Our structure shows that this residue lies at the bottom of the minor groove and is too close to accommodate any residue larger than a glycine. Introducing a Gly -> Ser mutation would require the backbone to move and would disrupt other contacts that the β-turn makes in the minor groove (Figure 6).

PAX3 Loss-of-function Missense Mutations and Waardenburg Syndrome Type 1 and 3

Several of the *PAX3* point mutations found in Waardenburg's syndrome patients (Asn 14 -> His; Asn 14 -> Lys; Pro 17 -> Leu; Phe 12 -> Leu; Figure 1b) (Baldwin et al., 1992; Hoth et al., 1993) also are located in or near this β-turn and further emphasize the importance of the contacts made by the turn. Mutations in position 14 seem to be associated with WS type 3, and will be discussed later. F12 and P17 are two very interesting residues. They both pack against the backbone of one strand of DNA. Here Pro 17-ribose 9 interactions take part in fixing the C-terminal end of the β-turn as it exits from the minor groove (Figure 3A of Chapter3). Pro 17 also forms hydrophobic interfaces with Pro 65 and Ile 68, and may help to hold the C-terminal tail of the N-subdomain in the minor groove. The P17L mutation may significantly weaken these interactions and thus destabilize the β-turn-DNA contacts and the DNA contacts made by the linker between two paired subdomains in the minor groove. It has been shown that this mutation abolishes PAX3 DNA binding activity (Czerny et al. 1993; Chalepakis et al., 1994a). The invariant Phe 12 side chain ring fits between the DNA backbone and the N-terminal β-sheet and may play a role in stabilizing the whole N-

terminal β-unit structure, and may play a special role in stabilizing the N-terminus of the β-turn in the minor groove (Figure 3A of Chapter 3; Figure 9). In addition, there is 20° local bend in the region where the β-turn fits into minor groove. This bend involves a large roll angle between base pair 8 and 9 (Table 1), and this may help to accomodate the conserved Phe 12 side chain in the minor groove. The DNA binding activity of P12L has not been reported, but our structure leads us to expect that it would has reduced DNA binding activity.

Several other missense mutations map to the N-terminal helical unit, and the structure also provides a basis for understanding these mutants. For example, one form of Waardenburg's syndrome involves a Gly 48 -> Ala mutation (WS .15 of *PAX3*; Figure 1c) (Tassabehji et al., 1993). Gly 48 is located at the bottom of the recognition helix, and it appears that introducing an alanine at this position would give unfavorable van der Waals contacts or disrupt the docking of the helix-turn-helix unit on the DNA (Figure 7). Another mutation (Bu35 of PAX-3, Figure 1b) (Hoth et al., 1993) changes the conserved Arg 23 residue which normally contacts both the phosphate backbone and the main chain carbonyl of residue 63. Obviously, introducing Leu at position 23 would disrupt these contacts (figure 8A and 8B).

All mutations discussed above are observed in heterozygotic individuals. Recently, the first case of human homozygotic *PAX3* mutation was reported: a new-born baby affected with a very severe form of WS type 3 was found to carry homozygotic missense mutation at position 51 of paired domain (S51F). (Patients with WS1 and upper-limb defects are classified as affected with WS type 3 (WS3), or Klein-Waardenburg syndrome.) The observation that the PAX3 homozygote in humans may allow life at least in early infancy and does not cause neural tube defects was not expected, since, in all the mutations known in mice (Splotch), homozygosity has led to severe neural tube defects and intrauterine or neonatal death. Genetic studies of the child's family, which for generations has been

affected by WS1, suggest that heterozygotic effect of S51F mutation is relatively mild. Our structure is consistent with the observed mild phenotype of this mutation. In our Prd structure, this conserved Ser is located in the recognition helix with its side chain forming a hydrogen bond with the DNA backbone. S51F mutation would disrupt the DNA backbone contacts and may thus result in a weaker DNA binding.

Very recently, other missense mutations in *PAX3* paired domain have been observed in patients affected with WS1 (Farrer et al. 1994; Read 1995), but more detailed information about position and identity of the mutation will be needed for discussion. There is also a missense mutation (*Splotch delayed*) in mouse *Pax-3* gene (Vogan et al. 1993). This mutation (G9R of paired domain) is located in the β-bulge (residue 6-10) between the N-terminal β-sheet. This β-bulge makes contacts with both the DNA backbone and the turn between helix 2 and helix 3 of paired domain. The ϕ and φ angle of this residue is in a region only favorable for glycine. Introducing an arginine side chain in this position would result in conformational changes in this β-bulge. Although glycine 9 does not directly contact DNA, it may affect DNA binding by interfering with the β-unit structure and/or the β-unit-HTH unit relationship. It has been shown that G9R mutation reduces the binding affinity of Pax-3 paired domain to the e5 site (Underhill et al., 1995). In addition, two missense mutations in the recognition helix of PAX3 homeodomain have been reported (Lalwani et al., 1995). First mutation V47F (numbered as in homeodomain) may prevent the proper recognition helix docking in the major groove. The second mutation R53G occurs to a conserved arginine which makes a DNA backbone contact (Kissinger et al., 1991).

Potential Gain-of-function *PAX3* Mutations and Waardenburg Syndrome Type 3

Interestingly, two heterozygotic point mutations in position 14 of the PAX3 paired domain can cause especially severe

phenotypes of Waardenburg syndrome (Read, A. P., personal communication). These two families have extra features beyond standard WS type 1. An N14H mutation exists in a family having WS type 3 (Hoth et al., 1993). Another mutation, N14K, has been found in a family featuring craniofacial-deafness-hand syndrome. A dominant-negative effect of these changes at Asn 14 seems a likely explanation (Read, 1995).

Asn 14 is the second residue of the β-turn in the minor groove, and makes a hydrogen bond with the 2-amino group of guanine 9. An examination of the structural environment of this residue in the paired domain-DNA complex suggests that N14H and N14K may bind to DNA with different specificity, instead of losing binding activity. The new His and Lys sidechains can be fit in the minor groove, making contacts with bp 8 and/or bp 9, while the rest of the complex retains the same conformation as the native prd paired domain-DNA complex. Obviously, other more radical changes in folding and docking connot be excluded at this stage, and it would be very interesting to know the DNA binding specificity of these two mutant proteins.

PAX6 Mutation and Aniridia/ Peters' Anomaly

Two point mutations have been reported for the *PAX6* gene. One mutation form (paired domain Arg 23 -> Gly) is associated with Peters' anomaly (Hanson et al. 1994). This mutation is expected to lose its DNA binding activity for the same reason discussed for *PAX3*(BU35) mutation (Arg 23 -> Leu) which cause Waardenburg syndrome. Another mutation associated with aniridia was found in the conserved region before the PAX6 homeodomain. In this patient, an arginine, which is conserved among all paired and homeo box-contaning genes, is substituted by a tryptophan (Hanson et al., 1993). It has been shown that this residue is located inside a nuclear localization signal, and is required for PAX6 nuclear localization. This PAX6 mutant protein has been found to be located in cytoplasm (Glaser et al., 1995). Since nuclear localization is presumably

required for PAX6 activity, mutations that disrupt the nuclear localization will be loss-of-function mutations.

Haploinsufficiency and PAX Phenotypes

Analysis of *Pax* mutations suggest that disrupting one copy of the *Pax* gene gives the observed phonotypic changes. Thus haploinsufficiency is the major pathological mechanism of WS and aniridia, but different tissues show differential sensitivity. The phenotypes of heterozygotes tend to involve tissues other than the central nervous system. Possibly the various Pax genes can complement each other in regions where several are expressed. This is clearly shown by Waardenburg syndrome type 1. Almost all affected people have dystopia canthorum, a mild facial malformation. Affected regions are derived from the neural crest, a PAX3-expressing tissue. About two thirds of patients show some patchy pigmentary disturbance of the eyes, hairs or skin, and a similar proportion have some degree of hearing loss. Both these features reflect faulty differentiation or migration of melanocyte precursors from the neural crest. A very few PAX3 heterozygotes have neural tube defects and a very few have contractures or hypoplasia of the upper limbs (WS 3). The end-organ sensitivity to PAX3 dosage is also reflected in the abundance of missense mutations associated with Waardenburg syndrome, as partially functional protein still cause a phenotype in heterozygote. *PAX6* mutations almost always truncate the protein. The dosage sensitivity to PAX6 seems lower than PAX3. Alternatively, missense mutations in *PAX6* may results in phenotypes diagnostically separate from aniridia.

Figure Legends

Figure 1

Stereo overview of the Prd paired domain - DNA complex. Ribbons are drawn through the Cα's of the protein and through the phosphate backbone of the DNA strand. The paired domain is in yellow, and the DNA is in blue. The complex is in the same orientation as Figure 2 of Chapter 3.

Figure 2

Stereo view of the paired N-terminal domain (residues 2-60), showing a view looking down helix 3 which makes contacts in the DNA major groove. The protein is in yellow, while the β-turn (residue 13-16) is in red. The β-turn contacts DNA in the minor groove. The 15mer DNA oligomer is in blue. It can also been seen from this view that helix 2 of paired domain fits partway into the DNA major groove, and the N-terminal end of this helix contacts the sugar-phosphate backbone.

Figure 3

Superposition of N-terminal paired domain with the engrailed homeodomain and with the DNA binding domain of Hin recombinase. **a.** Stereo view showing the superposition of Cα atoms of the engrailed homeodomain (residues 4 to 56) with the paired N-terminal domain (residues 15 to 61). The rms distance is 1.71Å for 43 C_a's (with two gaps). Paired is in yellow, and engrailed is in red. **b.** Stereo view showing the superposition of Cα atoms of the Hin recombinase (residues 144 to 181) with the paired N-terminal domain (residues 18 to 55). The rms distance is 1.28 Å for 38 contiguous C_a's. Paired is in yellow, and Hin is in purple.

Figure 4

DNA recognition sequences of the Pax subfamily containing Pax-2/5/8. Pax-5 recognition sequences are divided into two classes: class I requires the C-terminal domain for binding; class II only requires the N-terminal domain. "Pax-5 con" is the consensus of these Pax-5 recognition sequences. "Pax-2 con" is the *in vitro* selected binding site (Epstein et al. 1994). TPO and Tg are Pax-8 binding sites. The Pax-5 and Pax-8 recognition sequences are aligned as discussed by Czerny et al.(1993, 1995). In the Prd crystal structure, the N-terminal paired domain contacts base pairs 4 to 11 (HTH unit contacting bp 4 to 8, and β-turn motif contacting bp 9 to 11). In our Pax-5/PAX6 model, C-terminal domain interacts with base pairs 16 to 20, while the linker connecting the two domains contacts intervening base pairs in the minor groove.

Figure 5

Overview of the locations of Pax missense mutations in the structure. Only the paired N-terminal domain is shown. The Pax missense mutations are indicated by the red dots at the Cα atoms of residue 9, 14, 15, 17, 23 and 48. Except residue 9, all these residues are at protein-DNA interface and make important DNA contacts. Furthermore, mutations are clustered in or near the β-turn in the minor groove (residue 13 to 16) (two recently reported mutations at residue 12 and 14, which are associated with Waardenburg's syndrome, are not shown). Warrdenburg's syndrome and Peters' anomaly are human congenital disorders, and *Splotch* and *undulated* are mouse developmental phenotypes.

Figure 6

Structural basis for *Pax-1* undulated mutation (G15S). The structure shows that glycine 15 is located at the bottom of minor groove, and is too close to the DNA to accommodate any residue other than a glycine. The two N2 atoms of adjacent guanines in the minor groove are in gray. The ribbon representing protein is drawn through Cα atoms of residues 4 to 18.

Figure 7

Structural basis for *PAX3* Waardenburg syndrome mutation (G48A). In the Prd structure, invariant glycine 48 is the second residue of the recognition helix and is located at the bottom of major groove, and is too close to the DNA to accommodate any residue other than a glycine.

Figure 8

Structural basis for Pax mutations at residue 23 of the paired domain. PAX3 mutation R23L cause Waardenburg's syndrome, and PAX6 mutation R23G is associated with Peters' anomaly. **a.** a view looking along the recognition helix in the major groove. The DNA is shown by space-filling model. The base pairs are in cyan, while DNA backbone is in blue. The protein is shown by yellow ribbon. The side chain of arginine 23 is shown in red. Shown in gray are the phosphate oxygen atom in DNA backbone and the carbonyl group of residue 53 of paired domain, which form hydrogen bond contacts with the guanido group of arginine side chain. Hydrogen bonds are indicated by solid white lines. Residue 1 to 20 of the paired domain have been omitted for clarity. **b.** another view of the complex, showing the same interactions.

Figure 9

Side chain of Phe 12 fits in between the N-terminal β-sheet and the DNA backbone, and may play a role in stabilizing the β-turn in the minor groove. The DNA is in blue, while the three base pairs contacted by the β-turn are in white. The Cα trace of residue 4 to 18 of paired domain is in yellow while the β-turn is shown in red. A PAX3 mutation, F12L, is associated with Waardenburg's syndrome.

Table 1

DNA structure parameters.

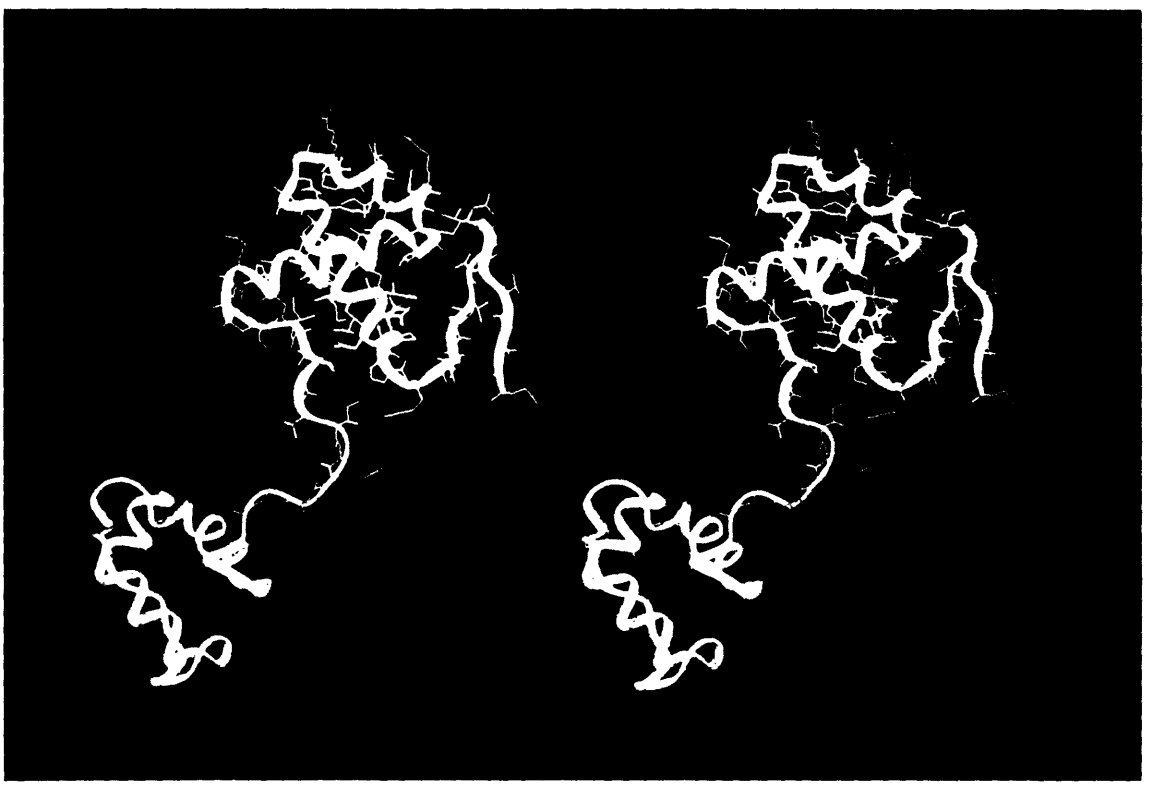

Figure 1

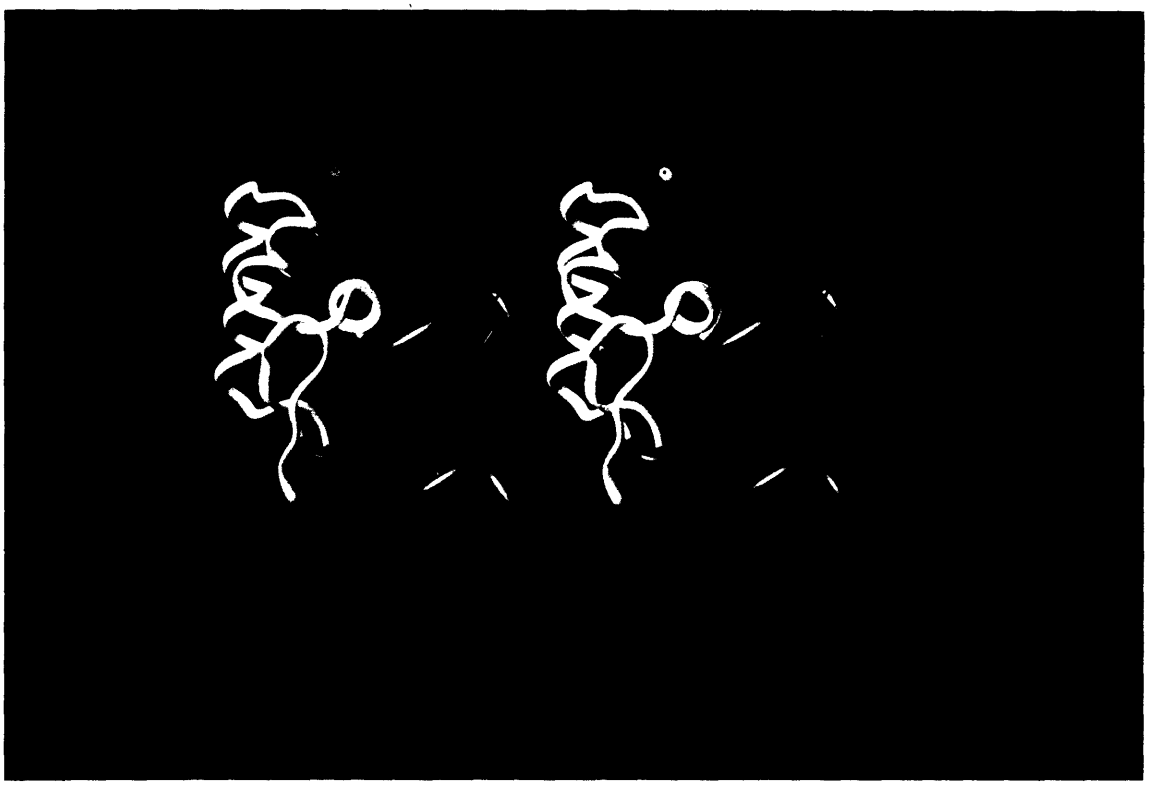

Figure 2

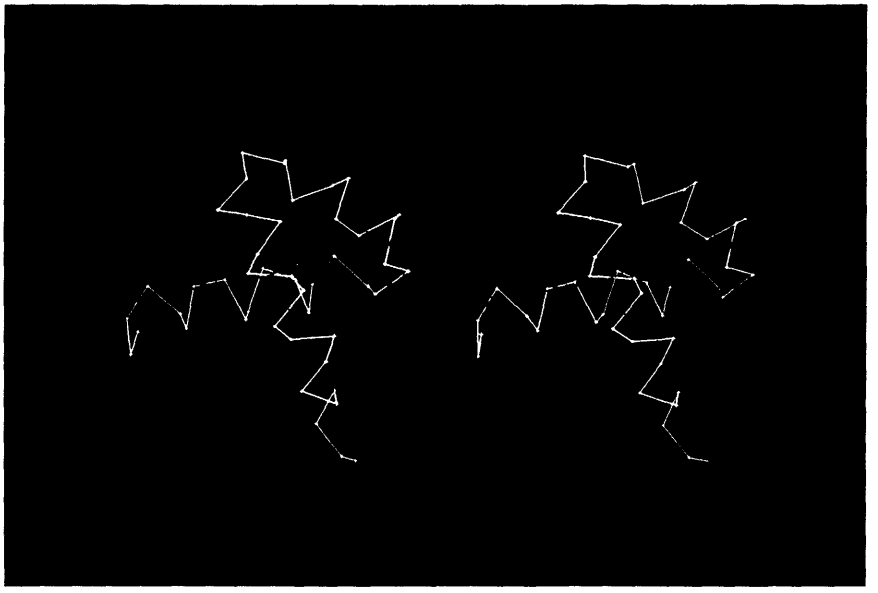

Figure 3a

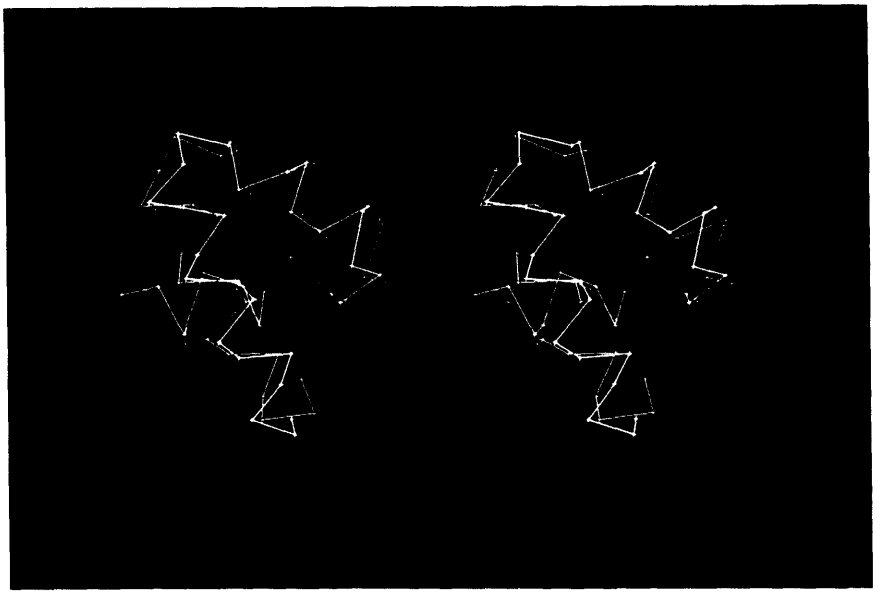

Figure 3b

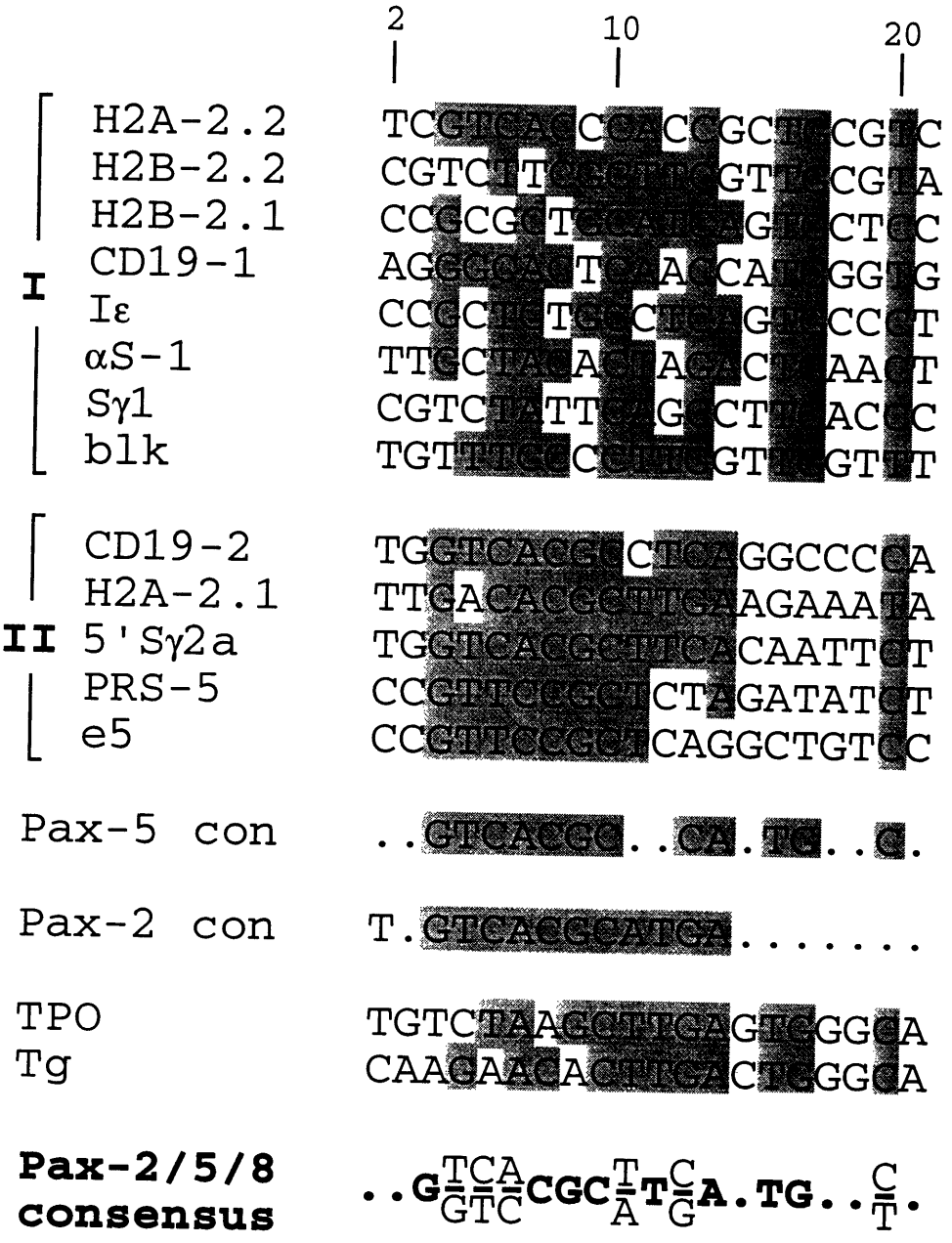
2
10
20
I
H2A-2.2
H2B-2.2
H2B-2.1
CD19-1
Iε
αS-1
Sγ1
blk
II
CD19-2
H2A-2.1
5'Sγ2a
PRS-5
e5
Pax-5 con
Pax-2 con
TPO
Tg
Pax-2/5/8 consensus

Figure 4

Figure 5

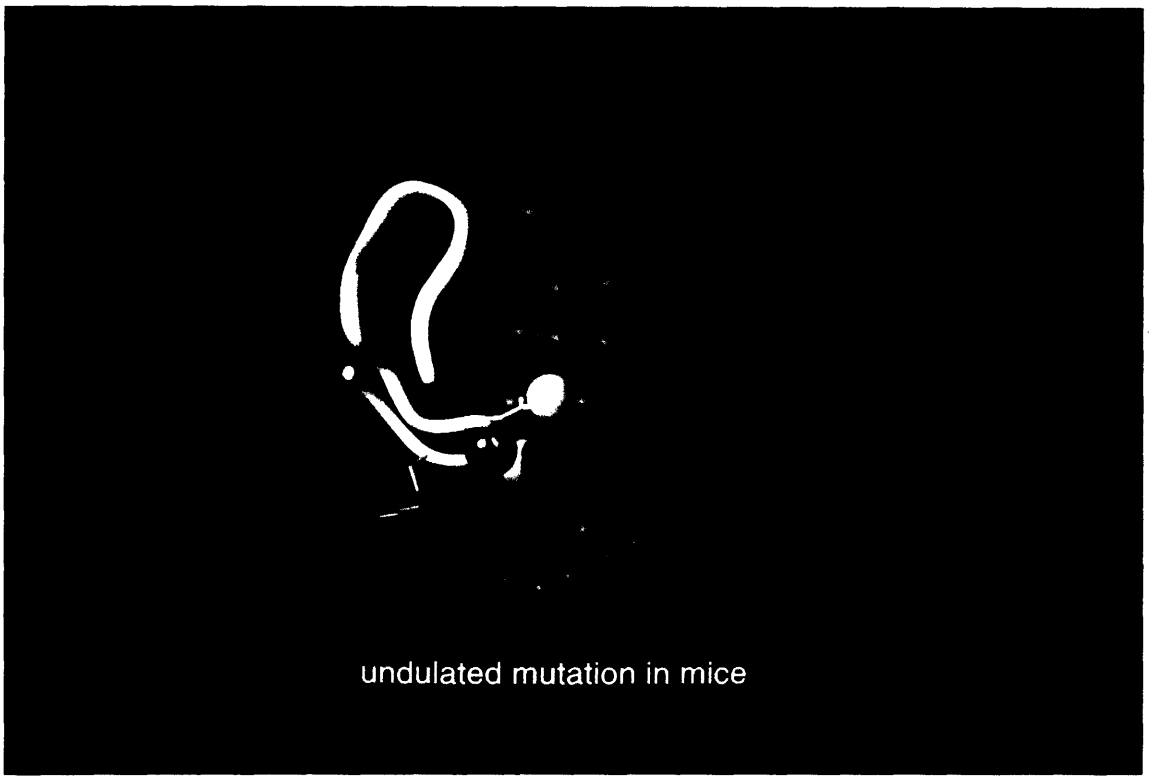

Figure 6

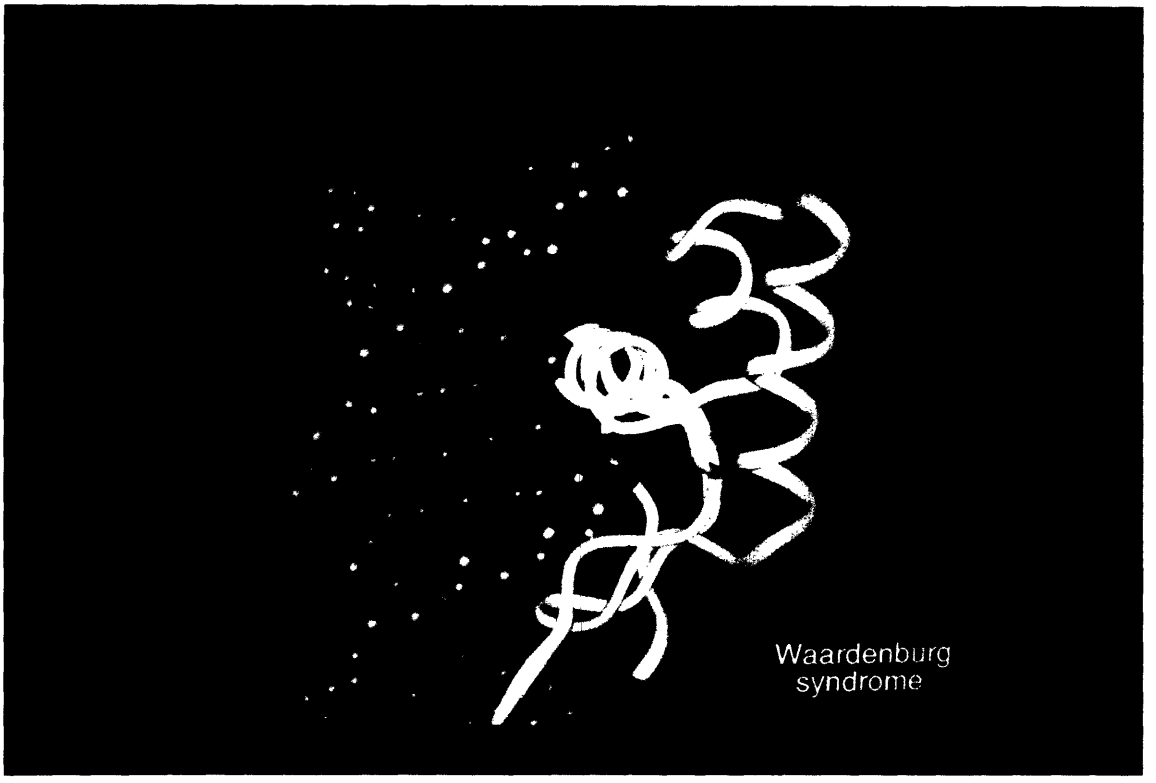

Figure 7

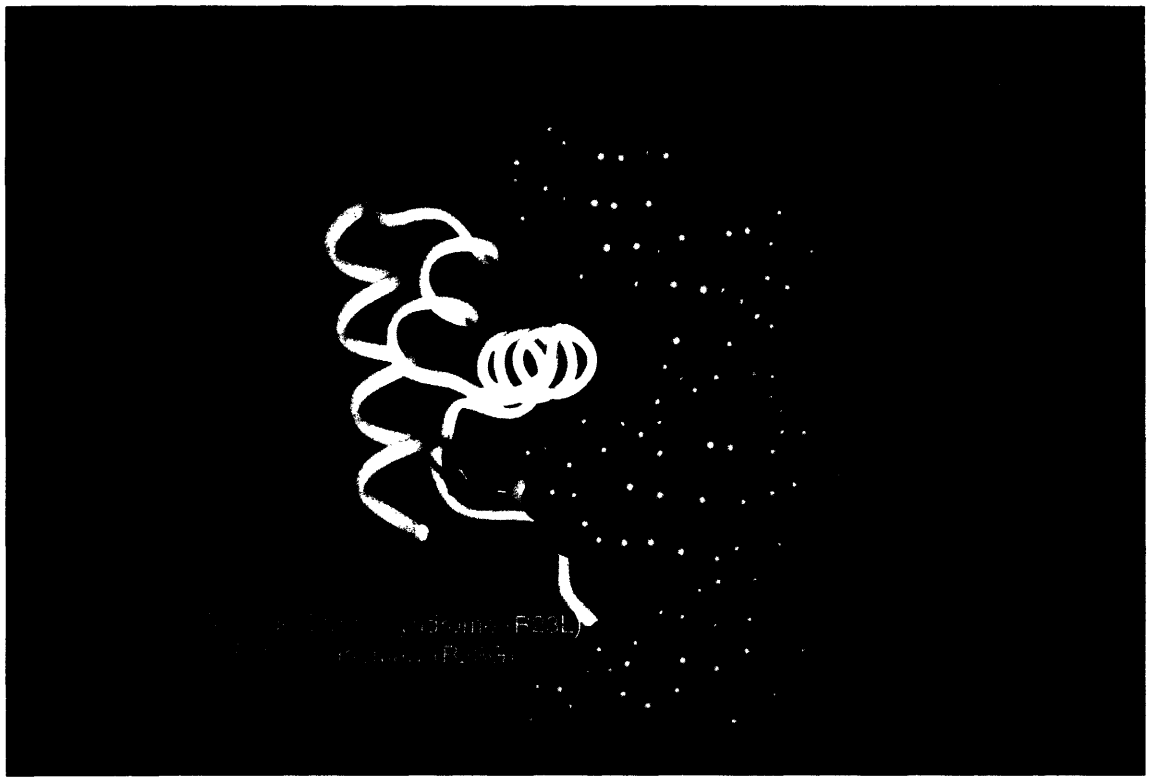

Figure 8a

Figure 8b

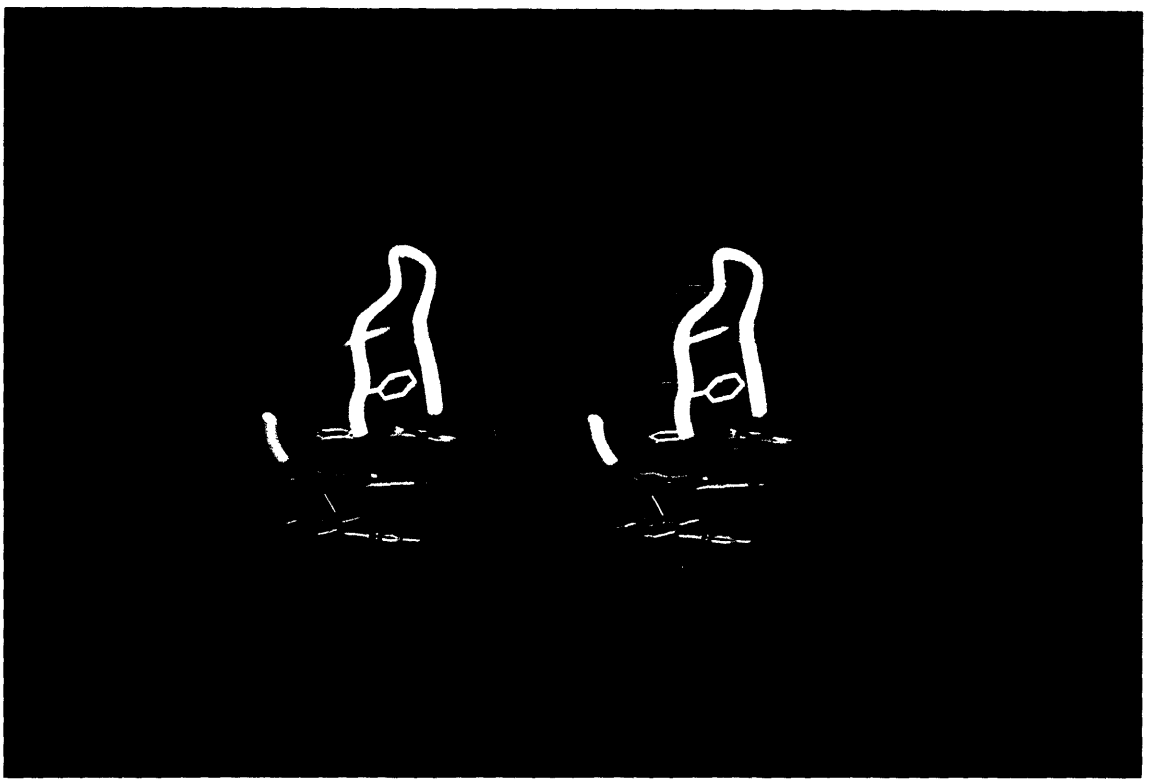

Figure 9

Table 1. Local helical parameters for the DNA site.

Base pairs	Helical twist	Rise/bp	Tilt	Roll	Propeller twist
3 C:G	---------->				-16.11
	36.54	3.61	-5.58	-3.37	
4 G:C	---------->				-14.81
	34.48	3.10	5.67	2.71	
5 T:A	---------->				-18.47
	31.84	3.38	-0.59	2.55	
6 C:G	---------->				-7.89
	41.52	3.60	-4.83	0.21	
7 A:T	---------->				-23.11
	30.05	3.30	-2.76	-8.40	
8 C:G	---------->				-11.54
	38.18	3.07	-5.50	18.63	
9 G:C	---------->				-18.69
	30.63	4.34	0.10	2.80	
10 G:C	---------->				-16.31
	33.54	3.36	3.14	3.50	
11 T:A	---------->				-11.82
	34.00	3.39	3.99	-2.20	
12 T:A	---------->				-19.86
	34.04	3.10	2.65	-0.40	
13 G:C	---------->				-6.71
	37.85	3.23	-2.91	4.40	
14 A:T	---------->				-19.95
	30.35	3.41	1.59	-4.15	
15 C:G	---------->				-17.02
Average	34.42	3.41	-0.39	1.25	-15.56

Chapter 5

Purification and Crystallization of Human PAX6 Paired Domain - DNA Complex

The crystal structure of the Prd paired domain - DNA complex provided the basic information on the structure of paired domain and how it uses the N-terminal domain for DNA recognition. However, for some other paired domains, such as these of PAX6 and Pax-5, the C-terminal paired domain also play an important role in paired domain-DNA interaction (Czerny et al. 1993; Epstein et al. 1994a). We are very interested in solving the cocrystal structure of the PAX6 paired domain complexed with a DNA oligomer containing the binding site for both the N-terminal and C-terminal domains. We are collaborating on this project with Dr. Richard Maas' group at Harvard Medical School. We expect this new structure will provide new information about: 1) how the C-terminal domain interacts with DNA; 2) how the linker region between two domains interacts with DNA; and 3) what the relationship is between two sub-domains in the paired domain. In addition, studies of the PAX6 paired domain will strengthen our understanding of the N-terminal domain and may reveal subtle structural variations for paired domains from different subfamilies.

The human *PAX6* gene controls early events in the morphogenesis of the brain and eye. These events include the induction of the lens and the development of cerebral cortex, topics that have fascinated developmental biologists since the beginning of this century (Grainger 1992; McConnell 1991; Glaser et al. 1995). *Pax-6* is primarily expressed in the developing central nervous system, and in the nose and the developing eye, including the lens (Walther and Gruss 1991; Puelles and Rubenstein 1993; Kioussi and Gruss 1994). The critical role of *PAX6* in vertebrate eye development has been demonstrated by the association of PAX6 mutations with three human eye diseases and one mouse developmental anomaly. Heterozygotic human PAX6 gene mutations are associated with Aniridia (An), Peters' anomaly and cataracts development (Glaser et al. 1992, 1994, 1995; Jordan et al. 1992; Hanson et al. 1993, 1994). Heterozygotic murine *PAX6* mutations is associated with Small eye (Sey) (Glaser et al. 1990; van der Meer-de Jong et al. 1990; Matsuo et al. 1993; Schmahl et al. 1993; Krauss et

al. 1992). Homozygotic *PAX6* mutations cause central nervous system defects in human and mouse. In both species, there are widespread abnormalies in the differentiation and migration of primitive neuronal cells, particularly in the celebral cortex, which are consistent with the normal expression pattern of *PAX6* (Hodgson and Saunders 1980; Hogan et al. 1986; Schmahl et al. 1993).

PAX6 homologues have also been cloned from mouse, rat, zebrafish, quail, chicken, axolotl salamander, sea urchin, Drosophila and C elegans (Walther and Gruss 1991b; Matsuo et al. 1993; Krauss et al. 1991;Quiring et al. 1994; Chisholm and Horvitz, submitted; Glaser at al. 1995; et al.). The comparison of *PAX6* sequence revealed a high degree of evolutionary conservation. For example, the Drosophila homolog of *PAX6*, *eyeless*, shows 94% amino acid sequence identity to vertebrate *Pax-6* in the paired domain, and 90% sequence identity to vertebrate *Pax-6* in the homeodomain (Quiring et al. 1994). Spontaneous mutations of the *eyeless* gene have been observed to affect gene expression in the eye primordia, causing its characteristic phenotype, the partial or complete absence of the compound eyes. It is remarkable that differentiation of organs as different as the eye of flies and humans may be under the control of a homologous gene cascade. Phylogenetic studies on the structure and development of eyes led to the proposal that eyes have evolved independently many times (perhaps as many as three or four dozens) (Salvini-Plawen and Mayr 1977). The finding of a highly homologous molecule functioning as a key regulator of eye morphogenesis in flies and vertebrates strongly argues for a common developmental origin (Quiring et al. 1994). The even more striking experiment is the induction of ectopic eyes by targeted expression of the *eyeless* genes in Drosophila. When *eyeless* was expressed in various imaginal disc primordia of Drosophila, ectopic eyes were induced on the wings, the legs, and on the antennae. The ectopic eyes appeared morphologically normal and consisted of groups of fully differentiated ommatidia with a complete set of photoreceptor cells (Halder et al.1995). This experiment strongly suggest that *eyeless* is the master control gene of Drosophila eye formation. Because

homologous genes are present in vertebrates, ascidians, insects, cephalopods, and nemerteans, it has been suggested that *Pax-6* may function as a master control gene throughout the metazoa.

The number of genes required for Drosophila eye morphogenesis has been estimated on the basis of enhancer detection lines that show reporter gene expression in the eye imaginal discs posterior to the morphogenetic furrow during eye differentiation (Halder et al. 1995). It has been estimated that more than 2500 genes are involved in Drosophila eye morphogenesis (of course, many of these genes may also be expressed in other tissues and cell types). The fact that *eyeless* gene can single-handedly turn on the program of Drosophila eye morphogenesis in different imaginal discs indicate that most of these genes are under the direct or indirect control of the *eyeless* gene. Obviously, it would be of great interest to understand how PAX6 can recognize the promoter regions of its target genes. We believed that the crystal structure of PAX6 paired domain-DNA complex would contribute significantly to this goal.

Purification of PAX6 Paired Domain

PAX6 paired domain was first purified as polyhistidine-tagged protein by Dr. Jonathan Epstein in Dr. Richard Maas' group (Harvard Medical School), but I was unable to specifically cut off the polyhistidine-tag. Then I worked out the protocol to purify protein from a T7 expression vector pET16b-Pd-DNB without any fusion protein, which was provided by Dr. Jonathan Epstein. The expression vector encodes the intact human PAX6 paired domain and 6 extra residues (Met-Asp-Pro-Met-Gln-Asn) on the N-terminus as a cloning artifact. Large scale preparation (10 litres of E coli. in LB broth) of the protein was done with the fermentor. Cells were grown at 37°C to a concentration of OD_{600} = 0.6, and then induced with 0.4 mM IPTG for three hours. Cells were spun down and then washed with prechilled phosphate-buffered-saline. Then cells were resuspended to a volume of 150 ml with a prechilled buffer containing 40 mM

Hepes (pH7.5), 50 mM NaCl, 1 mM EDTA, 2 mM DTT, 1 mM PMSF, 1μg/ml aprotinin, 1μg/ml pepstatin, 1μg/ml benzamidine, and 1μg/ml sodium metabisulfite. Cells were broken by passing French Press twice. Cell lysate was collected by centrifugation (GSA rotor, 12,000 rpm, 20 minute, at 4°C). Cell lysate was diluted by equal volume of a buffer containing 40 mM Hepes (pH 7.5),50 mM NaCl, 1 mM EDTA. Then add more NaCl into above diluted cell lysate to bring the final NaCl concentration to 0.2 M. Polyethyleneimine (PEI) precipitation is used next to get rid of nucleic acid, and also as a major purification step. In the cold room, 5 % PEI (pH7.6) stock solution was very slowly (in a period of about 10 minute) added into cell lysate (15 ml/ 1 litre E coli. prep.) during vigorous stirring. The solution was kept stirring in the cold room for another 30 minutes. Precipitation was spun down (GSA rotor, 12 krpm, 30 min.). The PEI supernatant was diluted by equal volume of 40 mM phosphate buffer (pH6.5), 1 mM DTT before loading to the column. The PAX6 paired domain was then purified with a S-sepharose Fast Flow (Pharmacia) column HR 10/10, using a gradient of 0.1 M to 0.3 M NaCl in 40 mM phosphate buffer (pH6.5), containing 1 mM DTT. PAX6 paired domain was eluted out by 0.2-0.25 M NaCl. At this stage the SDS gel showed a dominant PAX6-PD band and several weak contaminant bands. PAX6-PD was further purified by DNA cellulose column. The elution buffer contains 40 mM Hepes (pH7.6), 0.1 % NP-40, 1 mM EDTA, and 1 mM DTT. After loading PAX6-PD sample to the column in low salt (< 50 mM NaCl) condition, the column was washed by 10 bed volumes of 50 mM NaCl and 100 mM NaCl in elution buffer, then PAX6-PD was eluted out by another 10 bed volumes of 0.15-0.2 M NaCl in elution buffer. A few very weak contaminant bands was only visible when more than 20 ug PAX6-PD sample was loaded in a single lane. As a final step of purification and also a concentration step, reverse phase HPLC with a vydac C4 column was used to eliminate those minor contaminants. Protein was then lyophilized, resuspended by a buffer containing 10 mM Hepes (pH7.6), 1 mM DTT, aliquoted, frozen by liquid nitrogen, and stored at -80°C. The chemical homogeneity and identity of the purified PAX6 paired domain was further confirmed by N-terminal sequencing, and high resolution mass

spectrometry (Harvard MicroChem Facility, Cambridge). The protein showed one sharp shift band in gel-shift electrophoresis. Because of concerns about denaturation and unfolding, Eric Xu lately also tried to use milder columns (specific DNA affinity column and heparin column) to replace the last-step reverse phase HPLC column. However PAX6-PD purified in alternative ways produced crystals of similar quality. The procedures for purifying the PAX6 paired domain are summarized in Figure 1.

Crystallization of PAX6 Paired Domain - DNA Complex

The consensus DNA binding sites of PAX6 and Pax-5 are significantly longer than that of prd. In our model, PAX6 paired domain covers 20 base pairs in total. To ensure that C-terminal domain has enough space to settle down on DNA, we decided to test PAX6-PD - DNA complex crystallizability with DNA oligomers 20 base pairs or longer. We first tested 5 DNA duplexes, 20mer to 24 mer. For each oligomer, we tested MPD, PEG 400 and PEG 3350 as precipitants; tried different salt concentrations (0-0.2 M); and tested several pHs (pH4.6 - pH8.6). The effects of divalent cation (especially magnesium), spermine or spermidine, and cobaltic hexa-amine chloride, were also investigated. We also tested each oligomer with volatile salt ammonium acetate. It seems that longer DNA oligomers tend to be better candidates. We then tested a 25mer and a 26mer and several other 21mer to 24mers. While most DNA oligomers produced microcrystals or big crystals with some defects, we obtained two nice-looking crystal forms with the 25mer. Crystal form1 (rod-like) diffracted to 3.0 Å resolution in the DNA axis and 4.5Å in the directions perpendicular to DNA axis. Crystal form 2 (diamond-like) diffracted to 3.2 Å and 4.5 Å in corresponding directions. We also obtained nice crystals with the 26mer which later also diffracted to about same resolution. At this stage, Eric Xu joined me and we were able to test more crystallization conditions. We together tried cocrystallization with a series of different 25mers and 26mers with different overhanging bases or protein docking phases. So far, we have tried 32 DNA

duplexes and we have obtained at least 8 beautiful-looking crystal forms with 7 DNA duplexes (all of them are 25mer or 26mer). The morphology of one of these crystal forms was shown in Figure 2. The composition of the crystals has been analysed, and it has been clearly shown that crystals contain both PAX6 paired domain and the DNA oligomer used. Primary crystallographic tests showed that although several crystal forms diffract to about the same resolution as the first 25mer crystal form we obtained, none of them diffracted significantly better.

Except for the first 25mer crystal form, variations in the co-crystallization conditions have not yet been extensively screened for most oligomers mentioned above. It is plausible that the diffraction resolution can be improved by the fine-tuning of crystallization conditions. It also would be very interesting to systematically analyse the cell dimensions of every crystal form (using the data processing program DENZO). It may provide useful information for catagory possible crystal packing forms, which in turn may reflect the potential of improving a specific crystal form or a new crystal form.

PAX6 paired domain contains four cysteine residues. Three of them are conserved, the new one is in the turn between helix 4 and helix 5. All PAX6-PD co-crystallization overhanging drops were set with the presence of 5 to 10 mM DTT, as used in the PrdPD co-crystallization. Potential protein oxidization was also tested with SDS gel electrophoresis with the absence of reductant (β–ME or DTT), which showed that no protein-crosslinking was detectable in the cystallization drop two weeks after it was set. Chemical homogeneity of both protein and DNA were also satisfactory, as the purity of protein was shown by SDS-gel and mass spectrometry and the purity of DNA shown by a single band in denaturing DNA gel and a single sharp shift band in gel-retardation experiment. We believe that the chemical homogeneity of both the protein and DNA oligomers is not a major problem for more ordered crystal packing.

If further crystallization trials still can not significantly improve the diffraction resolution, it is worth studying whether the complex is conformationally homogeneous. First, the DNA oligomers in our crystals are at least 25 nucleotide long. The longer the DNA, the higher the flexibility for the DNA. In at least one case, protein-DNA cocrystals containing DNA oligomers longer than 25 bp has been produced and diffracted to high resolution (CAP-DNA complex, Schultz et al., 1991). In that case, protein forms dimer upon binding to DNA, and there is a large dimeric interface. The direct interactions between proteins may help to stabilize the DNA conformation. However in the paired domain - DNA complex, two sub-domains do not interact with each other, and the region of DNA contacted by the flexible linker may provide a major source of conformational flexibility. Second, it has been shown that PAX6 paired domain is partially unfolded before binding to DNA (Epstein et al., 1994a). The footprinting corresponding to C-terminal domain is relatively weak comparing to that of N-terminal domain, and the *in vitro* DNA binding consensus is also significantly weaker than that of N-terminal domain. It is possible that C-terminal domain may still be partially unfolded even after binding to DNA, or that the C-terminal domain only binds weakly to the DNA. To increase the chance of proper folding of the C-terminal domain, we usually incubated the protein-DNA mix for at least 30 minutes before setting the tray. We also tried to crystallize the complex in conditions containing up to 30% glycerol which may help to stabilize flexible domains. However the crystal diffraction power was not significantly improved. It would be interesting to know how well the C-terminal domain bound to the DNA (Kd measurements), and how stable the DNA-bound C-terminal domain is (Tm measurement). (These experiments are currently in progress in Dr. Richard Maas' laboratory).

In addition to trying other DNA oligomers and other crystallization conditions, a major parameter yet to be tried is crystallization in lower temperature, either setting the tray in the cold room or flash-freeze the crystal during data collection in

cryogenic conditions. The flexibility of both protein and DNA would be lower in lower temperature. Both methods, particularly cryogenic data collection, have been used in many cases to improve crystal diffraction resolution. Finally, if DNA conformational heterogeneity is the major problem, cocrystallization with shorter DNA oligomers (20mer or 21mer) could help to produce better crystals.

In summary, we have worked out the way to purify high-quality PAX6 paired domain in large quantity, and have obtained several promising cocrystal forms. There is still many options to vary to improve the crystal diffraction resolution. Eric Xu, a postdoctoral fellow in the laboratory, is continuing this project.

Future Directions of Structural Study of Pax protein-DNA Interaction

The crystal structure of PAX6 paired domain-DNA complex will provide structural basis for understanding remaining questions about paired domain-DNA interaction. Important directions for further structural studies concerning Pax protein-DNA interaction include: 1) structure of a protein-DNA complex of intact PAX6 protein, or a PAX6 fragment containing both paired domain and homeodomain; and 2) protein-DNA complex structure of alternatively splicing forms of Pax proteins, in particular, PAX6-5a.

The Pax proteins of two subfamilies, including PAX3, PAX4, PAX6 and PAX7, contain a second DNA binding domain other than paired domain, a paired-type homeodomain. As paired domain contains two homeodomain-like subdomains, these Pax protein contain three "homeodomains" in one protein. While the linker between two domains of paired domain is short and conserved, the linker between paired domain and homeodomain is divergent in both length and sequence. The role of the linker is still unclear. It does not activate transcription when fused to GAL4 (Glaser et al. 1994), but may function as a flexible hinge, since it is enriched in residues

such as glycine that have few bulky side chain.

Like other homeodomains, paired-type homeodomains are expected to contain three α–helices. The third helix contacts the DNA binding site in the major groove (Kissinger et al. 1990), and contains a serine in the ninth position. This serine is characteristic of paired-type homeodomain and plays an important roles in determining binding specificity (Treisman et al. 1989). Several members of this homeodomain sub-family have been shown to dimerize upon binding DNA (Wilson et al. 1993; Czerny and Busslinger, 1995; W.Schafer et al. 1994). While the in vitro selected optimal binding site for Prd homeodomain is a palindromic P2 site (TAATPyGATTA), the preferential site for Pax-6 homeodomain is a palindromic P3 site (TAATGCGATTA) (Czerny and Busslinger, 1995). The dimeric binding mode of paired-type homeodomain has been shown in the context of full length Pax-3, Pax-6 and Pax-7 proteins (Czerny and Busslinger, 1995; W.Scafer et al. 1994). More interestingly, Pax-3 and Pax-7, which show partial overlapping expression profile, are capable of forming heterodimer by binding to the P2 site (W.Scafer et al. 1994). Very recently, the cocrystal structure of Drosophila Prd homeodomain on the palindromic P2 site has been solved (Wilsonet al., personal communication).

Although hetero-dimerization on DNA site provides a potential mechanism for combinatorial control of gene expression by different PAX proteins, its biological role is to be confirmed, as no *in vivo* palindromic site has been found yet. The binding of PAX3 and Pax-6 homeodomains to a series of DNA sites containing a single TAAT core, in different sequence context, was not detected in gel-shift assays (Chalipakis et al. 1994; Czerny and Busslinger, 1995). Dimerization on palindromic sites can significantly enhance their DNA binding activity. However, using the optimal binding sites for both paired domain and homeodomain (a optimal palindromic P3 site), in the context of full-length Pax-6 protein, the paired domain proved to be more effective, by about 2 orders of magnitude, in DNA binding than the homeodomain. Thus it seems that the primary DNA-

binding mode of Pax homeodomains involves binding together with the paired domain. Actually, it has been shown that the appearance of a homeodomain sub-site near a poor paired domain site can compensate for a weak paired domain-DNA interaction. For example, intact Pax-6 protein, but neither the Pax-6 paired domain nor the Pax-6 homeodomain, can bind to the DNA site L1-170, which contains a very weak paired domain sub-site (A and T, instead of G and C, appear in base pair 9 and 10, which were contacted by the β-turn in the minor groove in our Prd structure) (Chalipakis et al. 1994b; see discussion in Chapter 4).

DNA binding site selection for a peptide containing both Prd paired domain and homeodomain showed a prefered alignment between paired domain and homeodomain, in which a conseved TAAT sequence (presumably the homeodomain binding sub-site) appears immediately 5' before the conserved sequence corrsponding to the paired domain binding site (Jun et al., personal communication). Our Prd paired domain-DNA complex crystal structure made it possible to model the structure of complex containing Prd paired domain, homeodomain and the selected optimal DNA binding site (Jun et al., personal communication). If this binding mode is biologically important, a crystal structure containing optimal DNA site and a PAX6 fragment contaning both paired domain and homeodomain would be very interesting. It will provide information about: 1) the spatial alignment of paired domain and homeodomain, their interactions and structural basis of their DNA binding cooperativity, if there is any; 2) the structure of linker between paired domain and homeodomain, their interactions with the two domains; 3) the structure of the short conserved regions flanking Pax homeodomains, which may be important for Pax function.

Another important direction is alternative splicing forms of Pax proteins. Several *PAX* gene alternative splicing forms have been observed. While Pax-2, PAX3 and Pax-8 isoforms preserve the intact paired domain (Dressler and Douglass 1992; Ward et al. 1994; Kozmik et al. 1993), PAX6 and PAX7 isoforms contain insertion in

the paired domain. PAX7 has an alternative form containing two extra residues before residue 117 of paired domain (the first residue of the proposed recognition helix of the C-terminal domain) (W.Schafer et al. 1994). This insertion extends the long "turn" of C-terminal HTH motif of PAX7 paired domain, and is not expected to interfere with the functioning of paired domain (see discussion of Chapter 3). The PAX6 alternative splicing form PAX6-5a, which has a 14 amino acid insertion before residue 47 of PAX6 paired domain, is particularly interesting. This insertion has been observed in the PAX6 of many species, and its sequence is evolutionally conserved. The 5a-insertion disrupts the original DNA binding activity of N-terminal paired domain, and the DNA binding activity of PAX6-5a paired domain can be simulated by deletion of 30 amino terminal residues from the PAX6 or Pax-2 paired domain. This insertion dramatically alters the DNA binding specificity, causing the paired domain to contact DNA primarily via its C-terminal half and facilitating cooperative dimerization between paired domains upon DNA binding (Epstein et al. 1994b). The functional importance of this PAX6 isoform is underscored by a PAX6 splicing site mutation which changes the ratio of two isoforms and causes a distinct human ocular syndrome (Epstein et al. 1994b). A crystal structure of PAX6-5a - DNA complex will reveal the structure of this isoform, and explain its DNA binding specificity and dimerization cooperativity upon DNA binding.

In summary, our Prd paired domain-DNA complex crystal structure, in conjunction with the available biochemical and genetic data, revealed the key features of the paired domain-DNA interactions and provided a structural basis for Pax developmental mutations. We have also made very exciting progress in crystallizing PAX6 paired domain-DNA complex which is expected to reveal the features of C-terminal paired domain-DNA interactions. Pax proteins play important roles in development. The structural studies of the Pax protein-DNA complex would be invaluable for understanding both the mechanisms of development and the general principles of protein-DNA recognition.

Figure Legends

Figure 1

Flow chart for PAX6-PD purification. The branching of procedure represents alternative ways for final step purification.

Figure 2

One of the crystal forms of PAX6-PD/DNA complex. The size of the crystal shown in this picture is about 0.45 X 0.45 X 0.15 mm.

PAX6-PD Purification procedures:

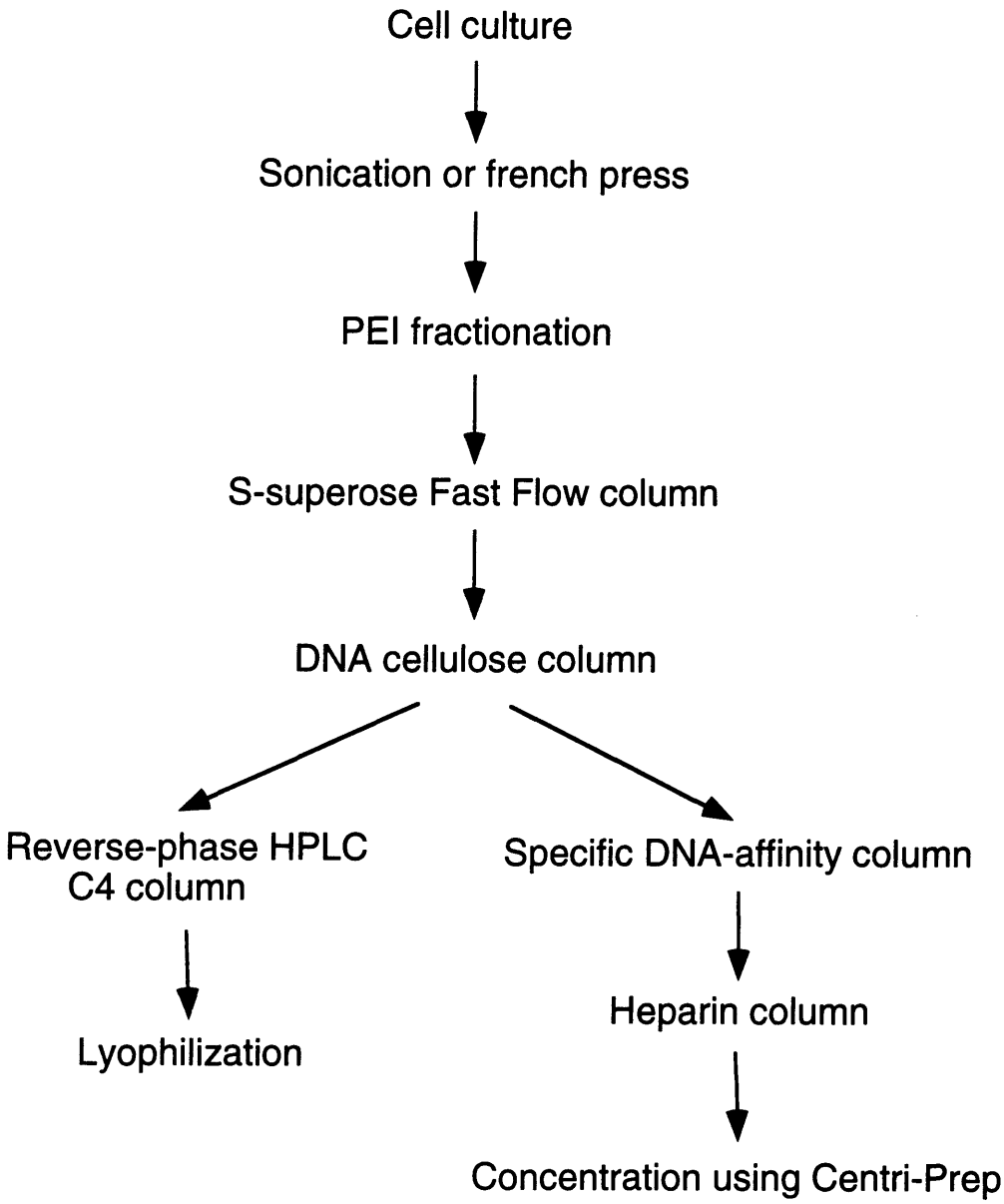

Figure 1

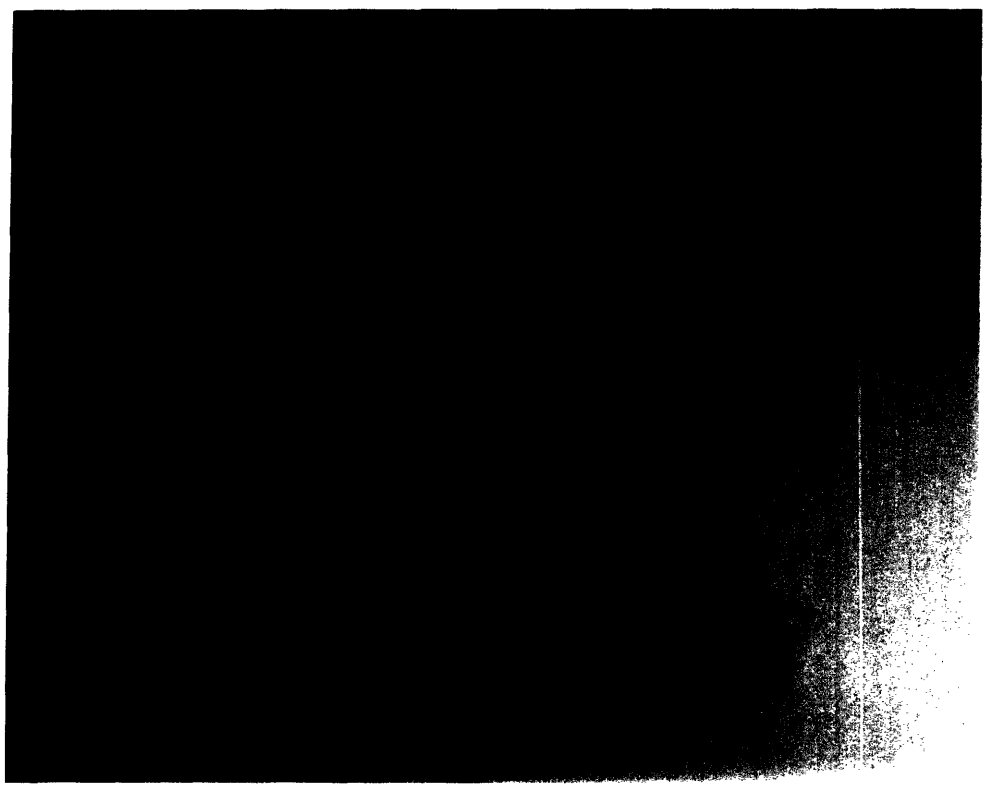

Figure 2

References

Adams, B., Dorfler, P., Aguzzi, A., Kozmik, Z., Urbanek, P., Maurerfogy, I. and Busslinger, M. (1992). *Pax-5* encodes the transcription factor BSAP and is expressed in B-lymphocytes, the developing CNS, and adult testis. Genes Dev. 6, 1589-1607.

Anderson, J., Ptashne, J. and Harrison, S. C. (1984). Cocrystals of the DNA-binding domain of phage 434 repressor and a synthetic phage 434 operator. Proc. Natl. Acad. Sci. USA 81, 1307-1311.

Balling, R., Deutsch, U. and Gruss, P. (1988). *undulated*, a mutation affecting the development of the mouse skeleton, has a point mutation in the paired box of *Pax-1*. Cell 55, 531-535.

Barberis, A., Superti-Furga, G., Vitelli, L., Kemler, I. and Busslinger, M. (1989). Developmental and tissue-specific regulation of a novel transcription factor of the sea urchin. Genes Dev. 3, 663-675.

Barr, F. G., Galili, N., Holick, J., Biegel, J. A., Rovera, G., and Emanuel, B. S. (1993). Rearrangement of the *PAX3* paired box gene in the paediatric solid tumour alveolar rhabdomyosarcoma. Nature Genet. 3, 113-117.

Baumgartner, S., Bopp, D., Burri, M. and Noll, M. (1987). Structure of two genes at gooseberry locus related to the paired gene and their spatial expression during Drosophila embryogenesis. Genes Dev. 1, 1257-1267.

Blow, D. M. and Matthew, B. W. (1973). Parameter refinement in the multiple isomorphous replacement method. Acta Cryst. A29, 56-62.

Bopp, D., Burri, M., Baumgertner, S., Frigerio, G. and Noll, M. (1986). Conservation of a large protein domain in the segmentation gene *paired* and in functionally related genes of *Drosophila*. Cell 47, 1033-1040.

Bopp, D., Jamet, E., Baumgartner, S., Burri, M. and Noll, M. (1989). Isolation of two tissue-specific Drosophila Paired box genes, *Pox meso* and *Pox neuro*. EMBO J. 8, 3447-3457.

Brennan, R. G. (1993). The winged-helix DNA-binding motif: another helix-turn-helix takeoff. Cell 74, 773-776.

Brünger, A. T. (1992a). X-PLOR Manual Version 3.0. (New Haven, Connecticut: Yale University Press).

Brünger, A. T. (1992b). The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature 355, 472-474.

Burri, M., Tromvoukis, Y., Bopp, D., Frigerio, G. and Noll, M. (1989). Conservation of the paired domain in metazoans and its structure in three isolated human genes. EMBO J. 8, 1183-1190.

Cai, J., Lan, Y., Appel, L. F. and Weir, M. (1994). Dissection of the *Drosophila* Paired protein: functional requirements for conserved motifs. Mechan. Dev. 47, 139-150.

Carter, T. C. (1947). A new linkage in the house mouse: undulated and agouti. Heredity 1, 367-372.

Chalepakis, G., Fritsch, R., Fickenscher, H., Deutsch, U., Goulding, M., and Gruss, P. (1991). The molecular basis of the *undulated/Pax-1* mutation. Cell 66, 873-884.

Chalepakis, G., Goulding, M., Read, A., Strachan, T. and Gruss, P. (1994a). The molecular basis of splotch and Waardenburg *Pax-3* mutations. Proc. Natl. Acad. Sci. USA 125, 417-425.

Chalepakis, G., Wijnholds, J., Giese, P., Schachner, M. and Gruss, P. (1994b). Characterization of Pax-6 and Hoxa-1 binding to the

promoter region of the neural adhesion molecule L1. DNA Cell. Biol. 13, 891-900.

Chalepakis, G., Wijnholds, J., and Gruss, P. (1994c). Pax-3 - DNA interaction: flexibility in the DNA binding and induction of DNA conformational changes by paired domains. Nucleic Acids Res. 22, 3131-3137.

Clark, K. L., Halay E. D., Lai, E. and Burley, S. K. (1993). Co-crystal structure of the HNF-3/*fork head* DNA-recognition motif resembles histone H5. Nature 364, 412-420.

Cvekl, A., Sax, C. M., Li, X., Bresnick, E. H.. and Piatigorsky, J. (1994). A complex array of positive and negative elements regulates the chicken *αA-crystallin* gene: involvement of Pax-6, USF, CREB and/or CREM, and AP-1 proteins. Mol. Cell. Biol. 14, 7363-7376.

Cvekl, A., Sax, C. M., Bresnick, E. H. and Piatigorsky, J. (1995). Pax-6 and lens-specific transcription of the chicken *δ1-crystallin* gene. Proc. Natl. Acad. Sci. USA 92, 4681-4685.

Czerny, T., Schaffner, G. and Busslinger, M. (1993). DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site. Genes Dev. 7, 2048-2061.

Czerny, T. and Busslinger, M. (1995). DNA-binding and transactivation properties of Pax-6: three amino acids in the paired domain are responsible for the different sequence recognition of Pax-6 and BSAP (Pax-5). Mol. Cell. Biol. 15, 2858-2871.

da-Silva, E. O. (1991). Waardenburg syndrome I: a clinical and genetic study of two large Brazilian kindreds, and literature review. Am. J. Med. Genet. 40, 65-74.

Deutsch, U., Dressler, G. R. and Gruss, P. (1988). *Pax-1*, a member of a paired box homologous murine gene family, is expressed in segmented structures during development. Cell 53, 617-625.

Dietrich, S. and Gruss, P. (1995). undulated phenotypes suggest a role of *Pax-1* for the development of vertebral and extravertebral structures. Dev. Biol. 167, 529-548.

Dressler, G. R., Deutsch, U., Chowdhury, K., Nornes, K. and Gruss, P. (1990). *Pax2*, a new murine paired box-containing gene and its expression in the developing excretory system. Development 109, 787-795.

Dressler, G. R. and Douglass, E. C. (1992). Pax-2 is a DNA-binding protein expressed in embryonic kidney and Wilms tumor. Proc. Natl. Acad. Sci. USA 89, 1179-83.

Engh, R. R. and Huber, R. (1991). Accurate bond and angle parameters for X-ray protein-structure refinement. Acta Crystallogr. A47, 392-400.

Epstein, J. A, Cai, J., Glaser, T., Jepeal, L. & Maas, R. L. (1994a). Identification of a Pax paired domain recognition sequence and evidence for DNA-dependent conformational changes. J. Biol. Chem. 269, 8355-8361.

Epstein, J. A., Glaser, T., Cai, J., Jepeal, L., Walton, D. S., and Maas, R. L. (1994b). Two independent and interactive DNA-binding subdomains of the Pax6 paired domain are regulated by alternative splicing. Genes Dev. 17, 2022-2034.

Epstein, J. A., Lam, P., Jepeal, L., Maas, R. L. and Shapiro, D. N. (1995). Pax3 inhibits myogenic differentiation of cultured myoblast cells. J. Biol. Chem. 19, 11719-11722.

Farrer, L. A., Arnos, K. S., Asher, J. M., Baldwin, C. T., et al. and Read, A. P. (1994). Locus heterogeneity for Waardenburgsyndrome is predictive of clinical subtypes. Am. J. Hum. Genet. 55, 728-737.

Ferre-D'Amare, A. R. and Burley, S. K. (1994). Use of dynamic light scattering to assess crystallizability of macromolecules and macromolecular assemblies. Structure 2, 357-359.

Feng, J., Johnson, R.C. and Dickerson, R. E. (1994). Hin recombinase bound to DNA: the origin of specificity in major and minor groove interactions. Science 263, 348-355.

Finney, M. (1990). The homeodomain of the transcription factor LF-B1 has a 21 amino acid loop between helix 2 and helix 3. Cell 60, 5-6.

Fredericks, W. J., Galili, N., Mukhopadhyay, S., Rovera, G., Bennicelli, J., Barr, F. G. and Rauscher, F. J. 3rd. (1995). The PAX3-FKHR fusion protein created by the t(2;13) translocation in alveolar rhabdomyosarcoma is a more potent transcriptional activator than PAX3. Mol. Cell. Biol. 15, 1522-1535.

Galili, N., Davis, R. J., Fredericks, W. J., Mukhopadhyay, S., Rauscher, F. J. III, Emanuel, B. S., Rovera, G., and Barr, F. G. (1993). Fusion of a fork head domain gene to *PAX3* in the solid tumour alveolar rhabdomyosarcoma. Nature Genet. 5, 230-235. [*erratum,* (1994). *ibid.* 6, 214.]

Glaser, T., Lane, J. and Housman, D. (1990). A mouse model for aniridia-Wilms tumor syndrome. Science 250, 823-827.

Glaser, T., Walton, D. S. and Maas, R. L. (1992). Genomic structure, Evolutionary conservation and aniridia mutations in the human *PAX* gene. Nature Genet. 2, 232-238.

Glaser, T., Jepeal, L., Edwards, J. G., Young, S. R., Favor, J. and Maas, R. L. (1994). *PAX6* gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. Nature Genet. 7, 463-471.

Glaser, T., Walton, D. S., Cai, J., Epstein, J. A., Jepeal, L. and Maas, R. L. (1995). Molecular genetics of Ocular Disease. 51-82. (Wiley-Liss, Inc.)

Goodman, R. M., Lewithal, I., Solomon, A. and Klein, D. (1982). Upper limb involvement in the Klein-Waardenburg syndrome. Am. J. Med. Genet. 11, 425-433.

Goulding, M. D., Lumsden, A. and Gruss, P. (1993). Signals from notochord and floor plate regulate the region-specific expression of two Pax genes in the developing spinal cord. Development 117, 1001-1016.

Grainger, R. M. (1992). Embryonic lens induction: shedding light on vertebrate tissue determination. Trends Genet. 8, 349-355.

Gruss, P. and Walther, C. (1992). *Pax* in development. Cell 69, 719-722.

Halder, G., Callaerts, P. and Gehring, W. J. (1995). Induction of ectopic eyes by targeted expression of the *eyeless* genes in Drosophila. Science 267, 1788-1795.

Hanson, I., Seawright, A., Hardman, K., Hodgson, S., et al. (1993). *PAX6* mutations in aniridia. Hum. Mol. Genet. 2, 915-920.

Hanson, I., Fletcher, J. M., Jordan, T., Brown, A., Taylor, D., Adams, R. J., Punnett, H. H., and van Heyningen, V. (1994). Mutations at the *PAX6* locus are found in heterogeneous anterior segment malformations including Peter's anomaly. Nature Genet. 6, 168-173.

Harrison, S. C. (1991). A structural taxonomy of DNA-binding domains. Nature 353, 715-719.

Hastie, N. D. (1993). Wilm's tumor gene and function. Curr. Opin. Genet. Dev. 3, 408-413.

Henderson, R. and Moffat, J. K. (1971). The difference fourier technique in protein crystallography: errors and their treatment. Acta. Cryst. B27, 1414-1420.

Hill, R. E., Favor, J., Hogan, B. L. M., Ton, C. C. T., Saunder, G. F., Hanson, I. M., Prosser, J., Jordan, T., Hastie, N. D., and van Heyningen, V. (1991). Mouse *small eye* results from mutations in a paired-like homeobox-containing gene. Nature 343, 522-525. [*erratum,* (1992). *ibid.* 355, 750.]

Hill, R. E. and Hanson, I. M. (1992). Molecular genetics of the *Pax* gene family. Curr. Opin. Cell Biol. 4, 967-972.

Hodel, A., Kim, S.-H. and Brünger, A. T. (1992). Model bias in macromolecular crystal structures. Acta Crystallogr. A48, 851-858.

Hogan, B. L. M., Hirst, E. M. A., Horsburgh, G and Hetherington, C. M. (1988). Small eye (Sey): a mouse model for the genetic analysis of craniofacial abnormalities. Development 103(Suppl.), 115-119.

Holst, B. D., Goomer, R. S., Wood, I. C., Edelman, G. M. and Jones, F. S. (1994). Binding and activation of the neural adhesion molecule by Pax-8. J. Biol. Chem. 269, 22245-22252.

Hoth, C. F., Milunsky, A., Lipsky, N., Sheffer, R., Clarren, S. K., and Baldwin, C. T. (1993). Mutations in the paired domain of human *PAX3* gene cause Klein-Waardenburg Syndrome (WS-III) as well as Waardenburg Syndrome type-1. Am. J. Hum. Genet. 52, 455-462.

Jordan, S. R., Whitcombe, T. V., Berg, J. M. and Pabo, C. O. (1985). Systematic variation in DNA length yields highly ordered repressor-operator cocrystals. Science 230, 1383-1385.

Jordan, S. R. and Pabo, C. O. (1988). Structure of the lambda complex at 2.5 Å resolution: details of the repressor-operator interactions. Science 242, 893-899.

Jordan, T., Hanson, I., et al. (1992). The human *PAX6* gene is mutated in two patients with aniridia. Nature Genet. 1, 328-332.

Keller, S. A. et al. (1994). Kidney and retinal defects (Krd), a transgene induced mutation with a deletion of mouse chromosome 19 that includes the *Pax2* locus. Genomics 23, 309-320.

Kioussi, C. and Gruss, P. (1994). Differential induction of *PAX* gene expression in the presence of NGF and BDNF in primary cerebellar cultures. J. Cell. Biol. 125, 417-425.

Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B. and Pabo, C. O. (1990). Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. Cell 63, 579-590.

Klemm, J. D., Rould, M. A., Aurora, R., Herr, W. and Pabo, C. O. (1994). Crystal structure of the Oct-1 POU domain bound to an Octamer site: DNA recognition with tethered DNA-binding modules. Cell 77, 21-32.

Kozmik, Z., Wang, S., Dorfler, P., Adams, B. and Busslinger, M. (1992). The promoter of the *CD19* gene is a target for the B-cell-specific transcription factor BSAP. Mol. Cell. Biol. 12, 2662-2672.

Krauss, S., Johansen, T., Korzh, V., Moens, U., Ericson, J. U. and Fjose, A. (1991). Zebrafish *pax[zf-a]*: A paired box-containing gene expressed in the neural tube. EMBO J. 10, 3609-3619.

Lai, E., Prezioso, V. R., Smith, E., Litvin, O., Costa, R. H. and Darnell, J. E. (1990). HNF-3A, a hepatocyte-enriched transcription factor of novel structure is regulated transcriptionally. Genes Dev. 4, 1427-1436.

Lalwani, A. K., Brister, J. R., Fex, J., Grundfast, K. M., Ploplis, B., San Agustin, T. B. and Wilcox, E. R. (1995). Further elucidation of the genomic structure of *PAX3*, and identification of two different point mutations within the *PAX3* homeobox that cause Waardenburg syndrome type 1 in two families. Am. J. Hum. Genet. 56, 75-83.

Lavery, R. and Sklenar, H. (1988). The definition of generalized helicoidal parameters and an axis of curvature for irregular nucleic acids. J. Biomol. Struct. Dyn. 6, 63-91.

Liao, F., Giannini, S. L. and Birshtein, B. K. (1992). A nuclear DNA-binding protein expressed during early stages of B-cell differentiation interacts with diverse segments within and 3' of the IgH chain gene cluster. J. Immunol. 148, 2909-2917.

Liu, B., Kissinger, C. R. and Pabo, C. O. (1990). Crystallization and preliminary X-ray diffraction studies of the engrailed homeodomain/DNA complex. Biochem. Biophy. Res. Communication 171, 257-259.

Martin, P., Carriere, C., Dozier, C., et al. (1992). Characterization of a paired box- and homeobox- containing quail gene (*Pax-QNR*) expressed in the neuroretina. Oncogene 7, 1721-1728.

Matsuo, T et al. (1993). A mutation in the *Pax*-6 gene in rat small eye is associated with impaired migration of midbrain crest cells. Nature Genet. 3, 299-304.

Maulbecker, C. C. and Gruss, P. (1993). The oncogenic potential of *Pax* genes. EMBO J. 6, 2361-2367.

McConnell, S. K. (1991). The generation of neuronal diversity in the central nervous system. Ann. Rev. Neurosci. 14, 269-300.

Morrisey, D., Askew, D., Roj, L. and Weir, M. (1991). Functional dissection of the paired segmentation gene in Drosophila embryos. Genes Dev. 5, 1684-1696.

Nornes, K., Dressler, G. R., Knapik, E. W., Deutsch, U.and Gruss, P. (1990). Spatially and temporally restricted expression of *Pax2* during murine neurogenesis. Development 109, 797-809.

Otting, G., Qian, Y.Q., Billeter, M., Müller, M., Affolter, M., Gehring, W.J., and Wüthrich, K. (1990). Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. EMBO J. 9, 3085-3092.

Pabo, C. O. and Sauer, R. T. (1992). Transcription factors: structural families and principles of DNA recognition. Annu. Rev. Biochem. 61, 1053-1095.

Pilz, A. J., Povey, s., Gruss, P. and Abbott, C. M. (1993). Mapping of the human homologs of the murine paired-box-containing genes. Mammalian genome. 4, 78-82.

Plaza, S., Dozier, C., Turque, N.and Saule, S. (1995). Quail *Pax-6* (*Pax-QNR*) mRNAs are expressed from two promoters used differentially during retina development and neuronal differentiation. Mol. Cell. Biol. 15, 3344-3353.

Poleev, A., Fickenscher, H., Munlos, S., et al. (1992). *Pax8*, a human paired box gene: isolation and expression in developing thyroid, kidney and Wilm's tumor. Development 116, 611-623.

Puelles, L. and Rubinstein, J. (1993). Expression patterns of homeobox and other putatitive regulatory genes in the embryonic

mouse forebrain suggests a neuromeric organization. Trends Neurosci. 16, 472-479.

Qian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J., and Wüthrich, K. (1989). The structure of the *Antennapedia* homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. Cell 59, 573-580.

Qian, Y. Q., Furukubo-Tokunaga, K., Müller, M., Resendez-Perez, D., Gehring, W. J., and Wüthrich, K. (1994). Nuclear magnetic resonance solution structure of the *fushi tarazu* homeodomain from *Drosophila* and comparison with the *Antennapedia* homeodomain. J. Mol. Biol. 238, 333-345.

Quiring, R., Walldorf, U., Kloter, U. and Gehring, W. J. (1994). Homology of the *eyeless* gene of Drosophila to the *Small eye* gene in mice and *Airidia* in human. Science 265, 785-789.

Ravishanker, G., Swaminathan, S., Beveridge, D. L., Lavery, R. and Sklenar, H. (1989). Conformational and helicoidal analysis of 30 PS of molecular dynamics on the d(CGCGAATTCGCG) double helix: "Curves," Dials and Windows. J. Biomol. Struct. Dyn. 6, 669-699.

Read, A. P. (1995) *Pax* genes - Paired feet in three camps. Nature Genet. 9, 333-334.

Richardson, J., Cvekl, A., Wistow, G. (1995). Pax-6 is essential for lens-specific expression of *ζ-crystallin*. Proc. Natl. Acad. Sci. USA 92, 4676-4680.

Rothman, P., Li, S. C., Gorham, B., Glimcher, L., Alt, F. and Boothby, M. (1991). Identification of a conserved lipopolysaccharide-plus-interleukin-4-responsive element located at the promoter of the germ line ε transcipt. Mol. Cell. Biol. 11, 5551-5561.

Rould, M. A., Perona, J. J., Söll, D. and Steitz, T. A. (1989). Structure of *E. coli.* glutaminyl-tRNA synthetase complexed with $tRNA^{Gln}$ and ATP at 2.8 Å resolution. Science 246, 1135-1142.

Rould, M. A., Perona, J. J. and Steitz, T. A. (1992). Improving multiple isomorphous replacement phasing by heavy-atom refinement using solvent-flattened phases. Acta Crystallogr. A48, 751-756.

Ryan, G., Steele-Perkins, V., Morris, J. F., Rauscher, F. J.3rd and Dressler, G. R. (1995). Repression of *Pax-2* by WT1 during normal kidney development. Development 121, 867-875.

Salvini-Plawen, L. and Mayr, E. (1977). Evol. Biol. 10, 207.

Sanyanusin, P., Schimmenti, L. A., et al. and Eccles, M. R. (1995). Mutation of the *PAX2* gene in a family with optic nerve colobomas, renal anomalies and vesicoureteral reflux. Nature Genet. 9, 358-364.

Schultz, S. C., Shields, G. C. and Steitz, T. A. (1990). Crystallization of E coli catabolite gene activator protein with its DNA binding site: the use of modular DNA. J. Mol. Biol. 213, 159-166.

Schultz, S. C., Shields, G. C. and Steitz, T. A. (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90°. Science 253, 1001-1007.

Seeman, N. C., Rosenberg, J. M., and Rich, A. (1976). Sequence specific recognition of double helical nucleic acids by proteins. Proc. Nat. Acad. Sci. USA 73, 804-808.

Shapiro, D. N., Sublett, J. E., Downing, J. R. and Naeve, C. W. (1993). Fusion of *PAX3* to a member of the forkhead family of transcription factors in human alveolar rhabdomyosarcoma. Cancer Res. 53, 5108-5112.

Singh, M. and Birshtein, B. K. (1993). NF-HB (BSAP) is a repressor of the murine immunoglobulin heavy-chain 3'α enhancer at early stages of B-cell differentiation. Mol. Cell. Biol. 13, 3611-3622.

Spolar, R. S. and Record, M. T. Jr. (1994). Coupling of local folding to site-specific binding of protein to DNA. Science 263, 777-784.

Stapleton, P., Weith, A., Urbanek, P., Kozmik, Z. and Busslinger, M. (1993). Chromosomal localization of 7 *pax* genes and cloning of a novel family member, *pax-9*. Nature Genet. 3, 292-298.

Steigemann, W. (1974). Thesis, TU Muenchen.

Strachan, T. and Read, A. P. (1994). *PAX* genes. Curr. Opin. Genet. Dev. 4, 427-438.

Stuart, E. T., Kioussi, C. and Gruss, P. (1994). Mammalian *Pax* genes. Annu. Rev. Genet. 28, 219-236.

Sousa, R. and Lafer, E. M. (1990). The use of glycerol in crystallization of T7 RNA polymerase: implications for the use of cosolvents in crystallizing flexible proteins. Methods: a companion to Method in Enzymology 1, 50-56.

Stoykova, A. S. and Gruss, P. (1994). Roles of *Pax* genes in developing and adult brain as suggested by expression patterns. J. Neurosci. 14, 1395-1412.

Tassabehji, M., Read, A. P., Newton, V. E., Harris, R., Balling, R., Gruss, P., and Strachan, T. (1992). Waardenburg's syndrome patients have mutations in the human homologue of the *Pax-3* paired box gene. Nature 355, 635-636.

Tassabehji, M., Read, A. P., Newton, V. E., Patton, M., Gruss, P., Harris, R., and Strachan, T. (1993). Mutations in the *PAX3* gene causing Waardenburg Syndrome type 1 and type 2. Nature Genet. 3, 26-30.

Tassabehji, M., Newton, V. E., Leverton, K., Turnbull, K., Seemanova, E., Kunze, J., Sperling, K., Strachan, T. and Read, A. P. (1994). *PAX3* gene structure and mutations: close analogies between Waardenburg syndrome and the Splotch mouse. Hum. Mol. Genet. 7, 1069-1074.

Terwilliger, T. C., Kim, S. C. and Eisenberg, D. (1987). Generalized method of determining heavy-atom positions using the difference patterson function. Acta Crystallogr. 43, 1-5.

Ton, C. C. T., Hirvonen, H., Miwa, H., Weil, M. M., Monaghan, P., Jordan, T., van Heyningen, V., Hastie, N. D., Meijers-Heijboer, H., Drechsler, M., Royer-Pokora, B., Collins, F., Swaroop, A., Strong, L. C., and Saunders, G. F. (1991). Positional cloning and characterization of a paired box- and homeobox-containing gene from the aniridia region. Cell 67, 1059-1074.

Treisman, J., Harris, E. and Desplan, C. (1991). The paired box encodes a second DNA-binding domain in the paired homeodomain protein. Genes Dev. 5, 594-604.

Tsukamoto, K., Nakamura, Y.and Niikawa, N. (1994). Isolation of two isoforms of the *PAX3* gene transcripts and their tissue-specific alternative expression in human adult tissues. Hum. Genet. 93, 270-274.

van der Meer-de Jong, R., Dickinson, M. E., Woychik, R. P., Stubbs, L., Hetherington, C. and Hogan, B. L. M. (1990). Location of the gene involved the Small eye mutation on mouse chromosome 2 suggests homology with human aniridia 2 (AN2). Genomics 7, 270-275.

Vogan, K. J., Epstein, D. J., Trasler, D. G. and Gros, P. (1993). The *splotch-delayed (sp^{d})* mouse mutant carries a point mutation within the paired box of the *pax3* gene. Genomics 17, 364-369.

Wallin, J., Mizutani, Y., Imai, K., Miyashita, N., Moriwaki, K., Taniguchi, M., Koseki, H., and Balling, R. (1993). A new *pax* gene, *pax-9*, maps to mouse chromosome-12. Mamm. Genome 4, 354-358.

Walther, C., Guenet, J. L., Simon, D., Deutsch, U., Jostes, B., Goulding, M. D., Plachov, D., Balling, R., and Gruss, P. (1991a). *Pax* - a murine multigene family of paired box-containing genes. Genomics 11, 424-434.

Walther, C., Gruss, P. (1991b). *Pax-6*, a murine paired-box gene, is expressed in the developing CNS. Development 113, 1435-1449.

Wang, B. C. (1985). Resolution of phase ambiguity in macromolecular crystallography. Meth. Enzymol. 115, 90-112.

Ward, T. A., Nebel, A., Reeve, A. E. and Eccles, M. R. (1994). Alternative messenger RNA forms and open reading frames within an additional conserved region of the human *PAX2* gene. Cell Growth Differ. 5, 1015-1021.

Waters, S.J., Saikh, K. U. and Stavnezer, J. (1989). A B-cell-specific nuclear protein that binds to DNA site 5' to immunoglobulin Sα tandem repeats is regulated during differentiation. Mol Cell. Biol. 9, 5594-5601.

Williams, M. and Maizels, N. (1991). LR1, a lipopolysaccharide-responsive factor with binding sites in the immunoglobulin switch regions and heavy-chain enhancer. Genes Dev. 5, 2353-2361.

Wilson, D., Sheng, G., Lecuit, T., Dostatni, N. and Desplan, C. (1993). Cooperative dimerization of Paired class homeodomains on DNA. Genes Dev. 7, 2120-2134.

Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., and Pabo, C. O. (1991). Crystal structure of a MATα2 homeodomain - operator

complex suggests a general model for homeodomain - DNA interactions. Cell 67, 517-528.

Xu, W., Rould, M. A., Jun, S., Desplan, C., and Pabo, C. O. (1995). Crystal structure of a paired domain-DNA complex at 2.5 Å resolution reveals structural basis for *Pax* developmental mutations. Cell 80, 639-650.

Zannini, M., Francis-Lang, H., Plachov, D. and Dilauro, R. (1992). Pax-8, a paired domain-containing protein, binds to a sequence overlapping the recognition site of a homeodomain and activates transcription from 2 thyroid-specific promoters. Mol. Cell. Biol. 12, 4230-4241.

Zlotogora, J., Lerer, I., Bar-David, S., Ergaz, Z. and Abeliovich, D. (1995). Homozygosity for Waardenburg syndrome. Am. J. Hum. Genet. 56, 1173-1178.

Zwollo, P. and Desiderio, S. (1994). Specific recognition of the blk promoter by the B-lymphoid transcription factor B-cell-specific activator protein. J. Biol. Chem. 269, 15310-15317.