Structure, Vol. 10, 1541–1549, November, 2002, ©2002 Elsevier Science Ltd. All rights reserved. PII S0969-2126(02)00883-3

Structure of the DNA Binding Region of Prospero Reveals a Novel Homeo-Prospero Domain

Jodi M. Ryter,[1] Chris Q. Doe,[2] and Brian W. Matthews[1,3]
[1]Institute of Molecular Biology
Howard Hughes Medical Institute and
Department of Physics
[2]Institute of Neuroscience and
Institute of Molecular Biology
Howard Hughes Medical Institute and
Department of Biology
1229 University of Oregon
Eugene, Oregon 97403

Summary

The Prospero transcription factor promotes neural differentiation in *Drosophila*, and its activity is tightly regulated by modulating its subcellular localization. Prospero is exported from the nucleus of neural precursors but imported into the nucleus of daughter cells, which is necessary for their proper differentiation. Prospero has a highly divergent putative homeodomain adjacent to a conserved Prospero domain; both are required for sequence-specific DNA binding. Here we show that the structure of these two regions consists of a single structural unit (a homeo-prospero domain), in which the Prospero domain region is in position to contribute to DNA binding and also to mask a defined nuclear export signal that is within the putative homeodomain region. We propose that the homeo-prospero domain coordinately regulates Prospero nuclear localization and DNA binding specificity.

Introduction

The *Drosophila* central nervous system develops from a population of neural precursor cells (neuroblasts) that divide asymmetrically into two distinct daughter cells, one of which (the ganglion mother cell; GMC) divides again and terminally differentiates into a pair of neurons. An important regulator of neural differentiation is *prospero*, which encodes a 1403 amino acid transcription factor that is required in GMCs to activate neuronal differentiation genes [1], repress neuroblast-specific genes [2], and repress cell cycle genes [3]. Prospero is detected in neuroblasts, but at interphase it is cytoplasmic and thus incapable of regulating gene expression. During neuroblast mitosis, Prospero protein is asymmetrically localized to the cell membrane, resulting in all Prospero protein being deposited into the newborn GMC [4–6]. Following neuroblast cytokinesis, Prospero protein is released from the GMC membrane and is translocated into the GMC nucleus, where it can regulate gene expression. Mutants that abolish or block nuclear localization of Prospero in the GMC result in defective GMC differentiation [7], whereas mutants that result in nuclear localization of Prospero in neuroblasts show misexpression of GMC-specific genes [8]. Thus, the proper subcellular localization of Prospero (cytoplasmic in neuroblasts; nuclear in GMCs) is essential for normal CNS development.

The protein domains of Prospero involved in its subcellular localization and transcriptional regulation functions have been defined. The C-terminal 160 amino acids consist of a highly divergent putative homeodomain (HD; residues 1241–1303) [9, 10] and an associated "Prospero domain" (PD; residues 1304–1403) that are essential for sequence-specific DNA binding and the transcriptional activation function of Prospero [11]. In addition, a nuclear export signal has been defined within the putative homeodomain, which in turn can be masked by the Prospero domain [12]. Finally, a *Drosophila*-specific central domain is necessary and sufficient for asymmetric cortical localization of Prospero in neuroblasts [4], due to its ability to bind the membrane-associated adaptor protein Miranda [7, 13–17].

Prospero orthologs (i.e., members of the Prospero/Prox1 family) have been identified in many animals, such as *D. virilis*, *C. elegans*, *X. laevis*, zebrafish, chicken, mouse, and human (Figure 1) [18–24]. The only common feature of this family of proteins is the C-terminal homeodomain/Prospero domain motif. This conserved C-terminal region is required for nearly all the known functions of Prospero, including regulation of nuclear/cytoplasmic localization, sequence-specific DNA binding, and transcriptional activation. The strong conservation of the linked homeodomain/Prospero domain region, together with its constellation of interrelated functions (nuclear import/nuclear export/DNA binding), made it an attractive target for structural analysis. Atypical homeodomain structures have been solved previously [25, 26], and threading analysis predicts that the putative Prospero homeodomain is capable of assuming a typical homeodomain structure [27], but these analyses were done with isolated homeodomains. Moreover, the putative Prospero homeodomain is even more divergent than the structurally characterized atypical homeodomains, and it was not known whether it would assume such a structure either in isolation or in combination with the Prospero domain.

Here we report the 2.05 Å X-ray structure of the predicted Prospero homeodomain together with the Prospero domain. The structure reveals that the overall region forms a single structural unit that here we name a "homeo-prospero domain." It provides a structural explanation for the role of the Prospero domain region in masking the nuclear export signal that is present within the homeodomain region, and for the role of the Prospero domain region in modulating the DNA binding specificity of the putative homeodomain.

[3]Correspondence: brian@uoxray.uoregon.edu

Key words: prospero; homeodomain; DNA binding domain; neural differentiation; *Drosophila*

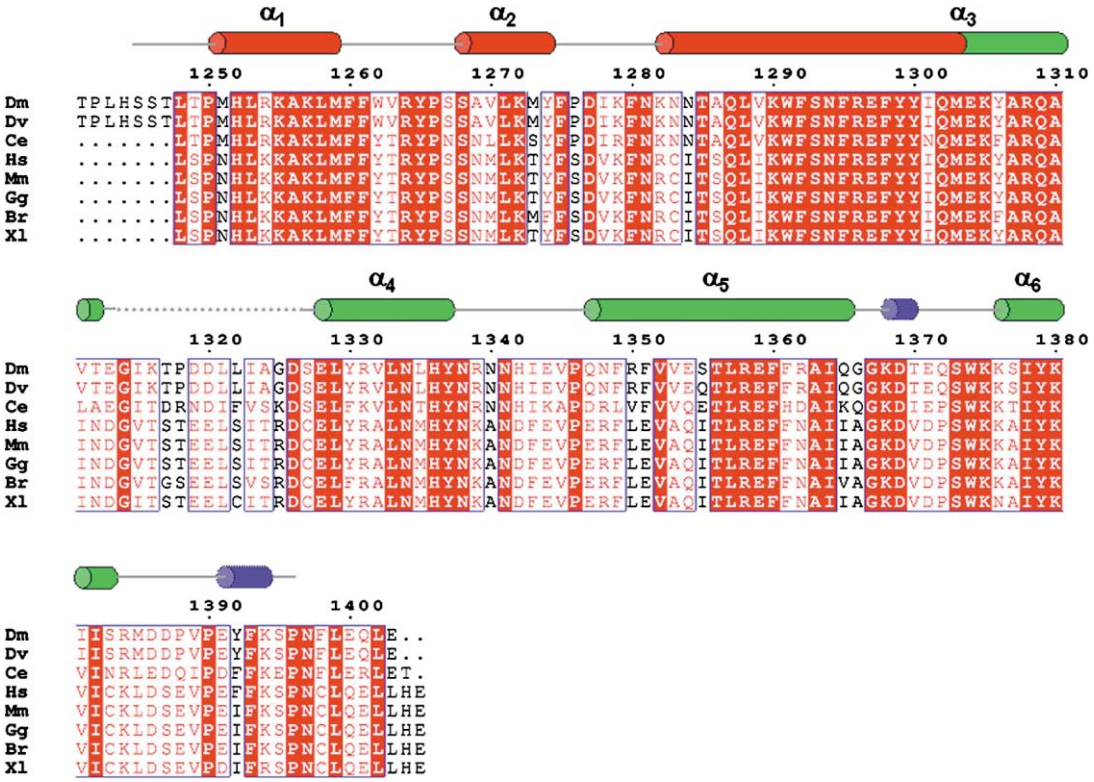

Figure 1. Sequence Comparison of the C-Terminal Residues of Prospero with Other Prospero/Prox1 Proteins Containing Putative Homeodomains and Prospero Domains

The sequences include ~160 amino acids from *Drosophila melanogaster* (Dm, residues 1241–1403), *Drosophila virilis* (Dv, residues 1394–1556), *Caenorhabditis elegans* (Ce, residues 430–586), *Xenopus laevis* (Xl, residues 583–740), zebrafish (Br, residues 582–739), chicken (Gg, residues 579–736), mouse (Mm, residues 580–737), and human (Hs, residues 579–736; numbering for *D. melanogaster* shown). Secondary structure elements (α helices, cylinders; random coil regions, solid lines; disordered residues, dotted lines) of the homeodomain and Prospero domains are colored red and green, respectively, with 3_{10} helices shown in blue. Sequence identity is illustrated as white single letter amino acid code on red. Sequence similarity is illustrated as red single letter amino acid code on white (90%). The alignment was generated using ESPript [45].

Results and Discussion

Structure Determination and Refinement

The C-terminal 163 amino acids of *Drosophila* Prospero (residues 1241–1403) were overexpressed, purified, and crystallized (Experimental Procedures; Figure 1). The X-ray structure was determined at 2.05 Å resolution utilizing MAD phasing techniques with SeMet. The experimental phases yielded interpretable electron density from which an initial model was built (Figure 2). Iterative refinement and model building resulted in the final model. Crystallographic refinement converged with a final R_{cryst}/R_{free} of 21.6%/25.6% (Experimental Procedures; Table 1).

Overview of the Homeo-Prospero Domain Structure

The overall structure of the homeo-prospero domain (HPD) is illustrated in Figures 3A and 3B. Despite classification as highly divergent based on sequence comparisons [9, 10], residues 1241–1303 were predicted to be capable of assuming a canonical homeodomain structure [27]. In fact, our structural analysis shows that this region does assume an overall fold very similar to homeodomains but with one striking difference. In every homeodomain structure determined to date, the so-called DNA recognition helix is either at the extreme C terminus of the protein or leads into a flexible linker of variable length that connects to another essentially independent domain. In Prospero, the recognition helix (α_3) connects the putative homeodomain and the Prospero domain as a single structural unit. The region corresponding to the Prospero domain can be described as a four-helix bundle (α_3–α_6; Figure 3A). Residues 1314–1326 are disordered and residues 1368–1370 and 1391–1394 form short 3_{10} helices. While interactions between the homeo- and Prospero domain regions occur primarily within the hydrophobic core, the C-terminal residues of the Prospero domain region make additional contacts with the homeodomain region by extending into a cleft between helices α_1 and α_2.

Structure of the Homeodomain Region

A structural comparison of the homeodomain region (HD) alone using Dali [28] revealed that, while possessing a highly divergent class sequence, the structure assumes a fold most similar to the *Drosophila* Engrailed homeodomain (Q50K mutant, PDB ID code 2HDD, 51 C^{α} atoms with rmsd 2.5 Å) [29]. Based on the results of the threading analysis mentioned above, and this Dali

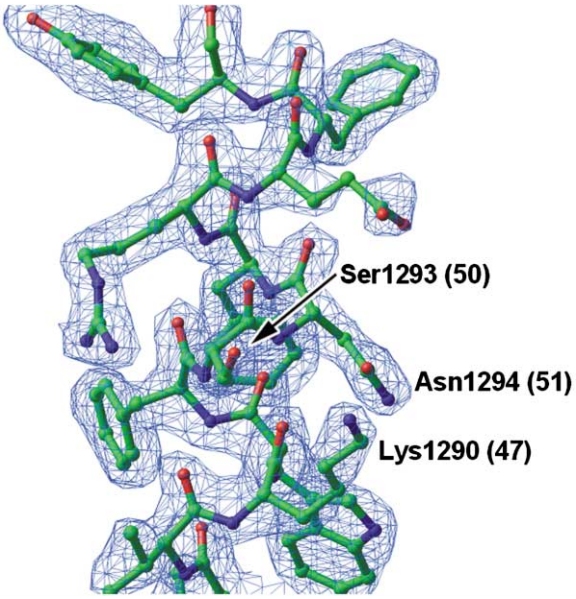

Figure 2. MAD-Phased, Experimental Electron Density Map

The map is shown contoured at 1σ surrounding the DNA recognition helix (α_3) of the HPD structure from Prospero. Key residues are labeled with numbering for *D. melanogaster* (homeodomain consensus sequence numbering shown in parentheses [30]).

comparison, a detailed residue-by-residue analysis was possible. A majority of the core residues seen in the homeodomain region are either invariant or conserved hydrophobic with respect to the Engrailed homeodomain sequence or homeodomain consensus sequence [30], indicating that this region possesses a core consistent with a standard homeodomain hydrophobic core. Variations occur primarily at the loop between α_1 and α_2 and in the middle of α_3. In these regions, residues that would be on the surface of a typical homeodomain are replaced by bulky hydrophobic groups that interact with the hydrophobic core of the Prospero domain. Within the loop, Phe1261 and Trp1262, along with the invariant Phe1260, create a very hydrophobic turn that protrudes into this core region. In the middle of α_3, Phe1298, Tyr1299, Tyr1300, and Met1303 contribute to the hydrophobic core of the Prospero domain four-helix bundle.

Structure of the Prospero Domain Region

The Prospero domain region (PD) is composed of a four-helix bundle (α_3–α_6). Consecutive helices are antiparallel and all helices interact via ridges in grooves (Figures 3A and 3B). The up-down-up-down topology is the simplest found among four-helix bundles and in this case is right turning [31]. The central residues of this bundle, representing the α-carbons in closest proximity, are Ala1307 of α_3 and Phe1361 of α_5. The $\sim$50° crossing angle between the helix axes is also the most common angle found in globular proteins. The core consists of hydrophobic residues throughout that are either invariant or highly conserved throughout the Prospero/Prox1 protein family (Figure 1). As noted above, a number of residues within the homeodomain region also contribute to this hydrophobic core.

The Interface between the Homeodomain and Prospero Domain Regions

The homeodomain and the Prospero domain regions do not just touch each other; rather, they interact quite

Table 1. Statistics of the Crystallographic Analysis

Intensity data processing Data set	λ (Å)	d_{min} (Å)	No. of observed relections	No. of unique reflections	Completeness (%)	$\langle I\rangle/\langle\sigma(I)\rangle$	R_{sym} (%)
(λ1) anomalous peak	0.9790	2.05	83,696	11,161	99.9 (100.0)	38.7 (6.7)	6.1 (23.3)
(λ2) inflection point	0.9793	2.05	72,405	11,172	99.9 (100.0)	34.7 (5.6)	6.3 (24.6)
(λ3) high-energy remote	0.9611	2.05	75,135	11,221	99.9 (99.7)	34.1 (4.7)	6.1 (25.5)

Phasing statistics		Refinement statistics	
Overall figure of merit (2.05 Å)	SHARP 0.708 (0.420) SOLOMON 0.919 (0.865)	Resolution (Å)	500–2.05
		R_{cryst} (%)	21.6
Phasing power$_{acentric}$ (iso/ano)	(λ1) 4.94/3.00	R_{free} (%)	25.6
	(λ2) 4.89/2.98		
	(λ3) —/2.75	Δ_{bonds} (Å)	0.006
		Δ_{angles} (°)	0.99
R_{cullis}	(λ1) 0.44		
	(λ2) 0.33	Average B (Å^3)	32.1
		Ramachandran analysis (%)	96.1, 3.9, 0.0, 0.0

$R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma I$, where I is observed intensity, and $\langle I\rangle$ is average intensity obtained from multiple observations of symmetry-related reflections.

$\langle I\rangle/\langle\sigma(I)\rangle$, where $\langle I\rangle$ is average I, and $\langle\sigma(I)\rangle$ is average error of I.

The values for the highest resolution shells are given in parentheses (2.12–2.05 Å intensity data, 2.19–2.05 Å phasing statistics).

Phasing power $= |F_{Hcalc}|/|F_{PH} - |F_P + F_{Hcalc}||$.

R_{cullis}(centric) $= \Sigma|F_{PH} - |F_P + F_{Hcalc}||/\Sigma|F_{PH} - F_P|$.

$R_{cryst} = \Sigma|F_{obs} - F_{calc}|/\Sigma|F_{obs}|$. R_{free} is R_{cryst}, calculated using 10% of the data, chosen randomly, and omitted from refinement.

Δ_{bonds} and Δ_{angles} give the average departure from ideal values of the bond lengths and angles, respectively.

Values of the Ramachandran analysis indicate the percentage of residues exhibiting main chain dihedral angles found in the most favored, allowed, generously allowed, and disallowed regions, respectively (PROCHECK [49]).

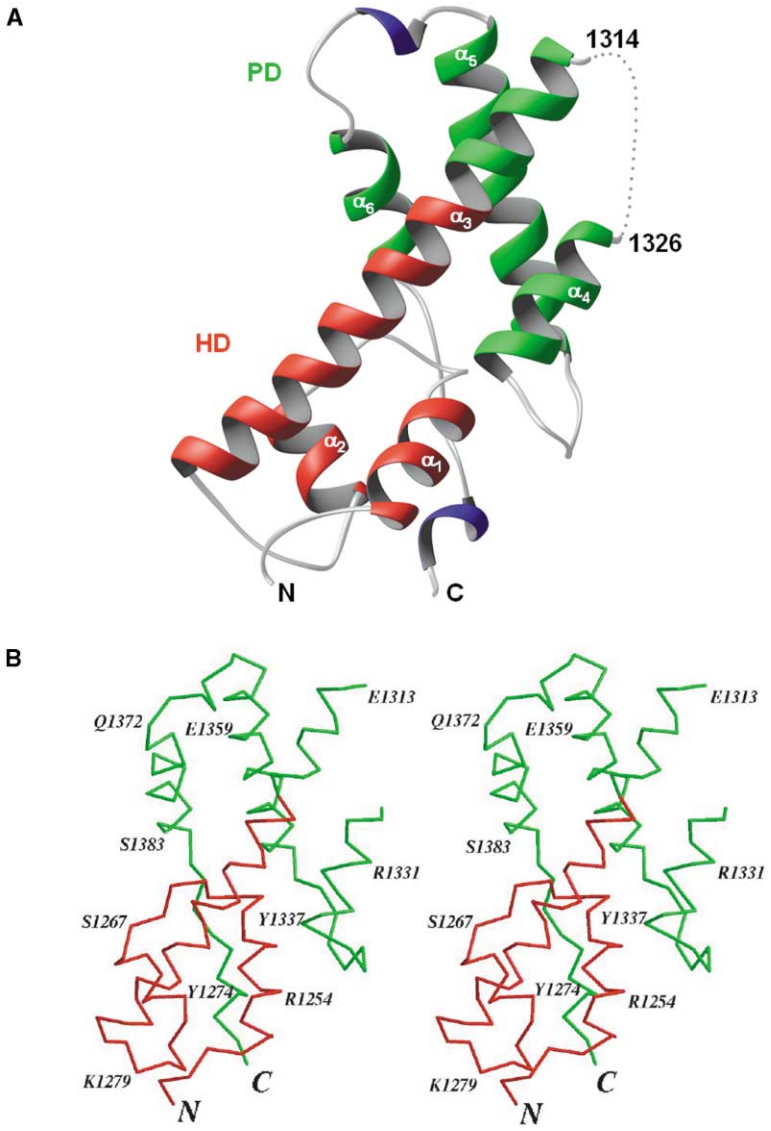

Figure 3. Overall Fold of the HPD Structure

(A) Prospero residues 1245–1396 are shown. The HD region is shown in red, and the PD region in green. 3_{10} helices are shown in blue. Residues 1314–1326 are disordered and indicated by the dotted line. The α helices within each domain are numbered consecutively on the basis of the primary sequence. Identical color coding is used in all illustrations. This and other similar illustrations were generated using RIBBONS [46].

(B) Stereo view of the α-carbon backbone of Prospero, residues 1245–1396. The HD region and PD region are colored as in (A). This illustration was generated using MOLSCRIPT [47].

extensively. If the two regions are considered as separate entities, then 22% (1499 Å^2) of the surface area of the homeodomain region contributes to the interface with the Prospero domain region and, vice versa, 35% of the surface of the Prospero domain region contributes to the same interface. These fractions are significantly higher than the average value of 15% that is observed for the surface area involved in the interfaces of typical protein dimers [32]. It confirms that the homeodomain and Prospero domain regions combine to form a distinct structural unit.

There are extensive hydrophobic core contacts that occur between the HD and PD regions. A hydrophobic cleft on the HD is formed between Trp1262 and Tyr1274 of the strand just C-terminal to α_1 and α_2, respectively. SeMet1259 forms the floor of this cleft, with Val1263 and Val1270 flanking one end. Seated in the cleft are Val1389, Pro1390, and Phe1393 of the PD region. This cleft is flanked at the opposite end by a hydrogen-bonding interaction between His1252 and Asp1277. Phe1275, of the strand C-terminal to α_2, positions Lys1255 of α_1 in an orientation optimal for a number of hydrogen-bonding interactions. These include direct interactions with backbone carbonyls of residues 1274 and 1393, thereby creating contacts between HD α_1 and α_2, as well as the PD region.

Dali analysis reveals that there are interesting similarities between the HPD structure and that of the MATa1/MATα2 homeodomain heterodimer bound to DNA [33]. In the HPD structure, the C terminus of the PD region (residues 1388–1396) extends into a cleft between α_1 (residues 1250–1259) and α_2 (residues 1268–1274) of the HD region (Figure 4, top). This rather lengthy segment includes a 3_{10} helix (residues 1391–1394) and contacts the HD region on the face opposite its DNA binding surface. This interaction between HD and PD is stabilized by hydrophobic interactions, as well as a number of hydrogen bonds. In the homeodomain heterodimer complex between MATa1 and MATα2 (Figure 4, bottom), the C terminus of MATα2 also extends away from the body of the domain and similarly makes hydrophobic and hydrogen bond interactions within a groove formed between α_1 and α_2 of MATa1. This binding groove on the homeodomain MATa1 corresponds to that within

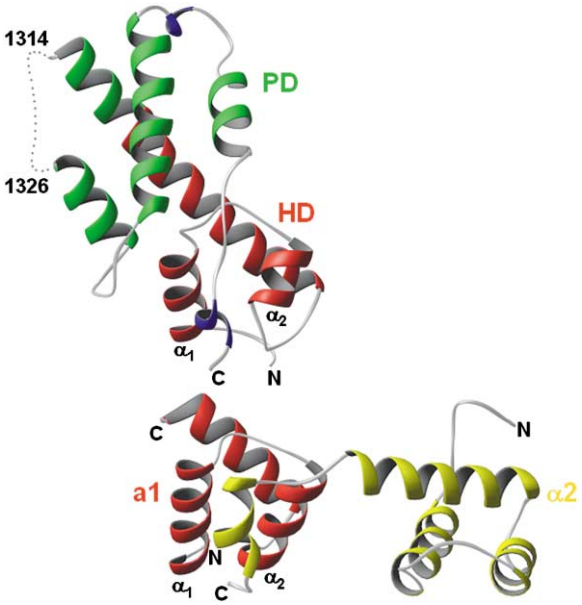

Figure 4. Comparison of the Prospero HPD and *S. cerevisiae* MATa1/MATα2 Homeodomain Heterodimer

Note the similar fashion in which the C-terminal tails of PD/MATα2 pack into a hydrophobic cleft formed between α_1 and α_2 of HD/MATa1 (see text). MATa1 and MATα2 are illustrated in red and yellow, respectively. The HPD is illustrated as described in Figure 3.

the HD region of Prospero. In addition, the polypeptide chain that binds within the respective grooves in both cases adopts an α-helical or 3_{10}-helical conformation (Figure 4).

The MATa1/MATα2 homeodomain complex is required for cell type-specific transcriptional repression in yeast [34]. The Prospero HPD is also required for regulating transcription [11], but may have an additional function in controlling the subcellular localization of Prospero. Demidenko et al. [12] utilized molecular dissection of the HPD with expression of chimeric proteins in mammalian and insect cultured cells to show that residues 1248–1261, encompassing α_1, contain a nuclear export signal (NES). In the absence of any portion of the PD region, Prospero is exported from the nucleus via the Exportin pathway and subsequently found in the cytoplasm, while presence of this region blocks nuclear export and allows Prospero to accumulate in the nucleus. The HPD structure shown in Figure 4 strongly suggests that it is the extreme C terminus of the PD region that sterically prevents access to the nuclear export signal.

Interestingly, previous analysis of Prospero during asymmetric cell division showed its phosphorylation state to correlate with subcellular localization [35]. Cytoplasmic Prospero is highly phosphorylated compared to nuclear Prospero, which raises the possibility that phosphorylated HPD may have a more open conformation in which the NES is exposed.

Protein-DNA Binding Model

In the Dali structural comparison mentioned above, the structure of the HD region is most similar to the Engrailed homeodomain (see above) [29]. It is therefore possible to use the structural superposition matrix determined by Dali to model the HPD structure in complex with DNA. The result is shown in Figure 5A. Three regions of potential protein-DNA contact are suggested. First, the DNA recognition helix α_3 appears to contact the major groove of the DNA; second, the N-terminal arm appears to contact the minor groove of the DNA; and third, unexpectedly, the PD region appears to contact the DNA backbone via residues within or close to the N terminus of helix α_6. The Dali structural alignment of the Prospero HD region with the Engrailed homeodomain concurs completely with the threading analysis [27]. Figure 5B illustrates these alignments at the sequence level. While the convention is to describe homeodomain structures using the homeodomain consensus sequence numbering scheme [30], here we maintain the native numbering of the HD region due to its intimate association with the Prospero domain.

Figure 6A illustrates the key residues of the Engrailed recognition helix known to make contacts in the DNA major groove. Engrailed homeodomain residues Asn51 (Pros HD Asn1294) and Arg53 (Pros HD Arg1296) are invariant and presumably contact the DNA in a similar fashion. In the Engrailed structure, Ile47 (Pros HD Lys1290) and Lys50 (Pros HD Ser1293) also contact the bases in the major groove to provide binding site discrimination. Pros HD Glu1297 appears poised in the major groove available for possible water-mediated contacts. Additionally, there are a number of residues that are not only invariant between the Engrailed homeodomain and the Prospero HD region, but are also positioned to make similar contacts with the DNA phosphate backbone as seen in the Engrailed homeodomain structure. In particular, Engrailed homeodomain Arg53 (Pros HD Arg1296) and Tyr25 (Pros HD Tyr1265) are predicted to interact directly with one strand of the DNA backbone in a similar fashion, while Trp48 (HD Trp1291) could interact with the opposite strand.

In the Engrailed homeodomain structure, the N-terminal arm (residues 5–9) fits into the DNA minor groove, supplementing the contacts made by the recognition helix. In the Prospero HPD-DNA model, the N-terminal arm has a somewhat different alignment (Figure 6A) but could easily move so as to contact the DNA. A sequence comparison reveals predominantly conservative differences between the Engrailed and the Prospero sequences in the N-terminal arm, except at Engrailed homeodomain Arg5 (Pros HD Ser1245; Figure 5B), a position shown to be important for binding site discrimination [36].

This structural comparison also suggests that the Prospero HD region may contact DNA as well (Figure 6B). Lys1380, -1376, and -1375, found within or close to the N terminus of the α_6 helix, are all poised to potentially interact with the DNA phosphate backbone. The basic nature of this region is further manifest in the negative electrostatic potential of the molecular surface (Figure 6C). Tyr1379 and Ser1373 could possibly interact with the DNA backbone either directly or via hydrogen bond interactions. All these residues are conserved to a high degree in the Prospero/Prox class homeobox proteins (Figure 1). Of course, final confirmation of the HPD-DNA interactions discussed above must await the structure determination of the HPD structure in complex with its DNA binding site [11].

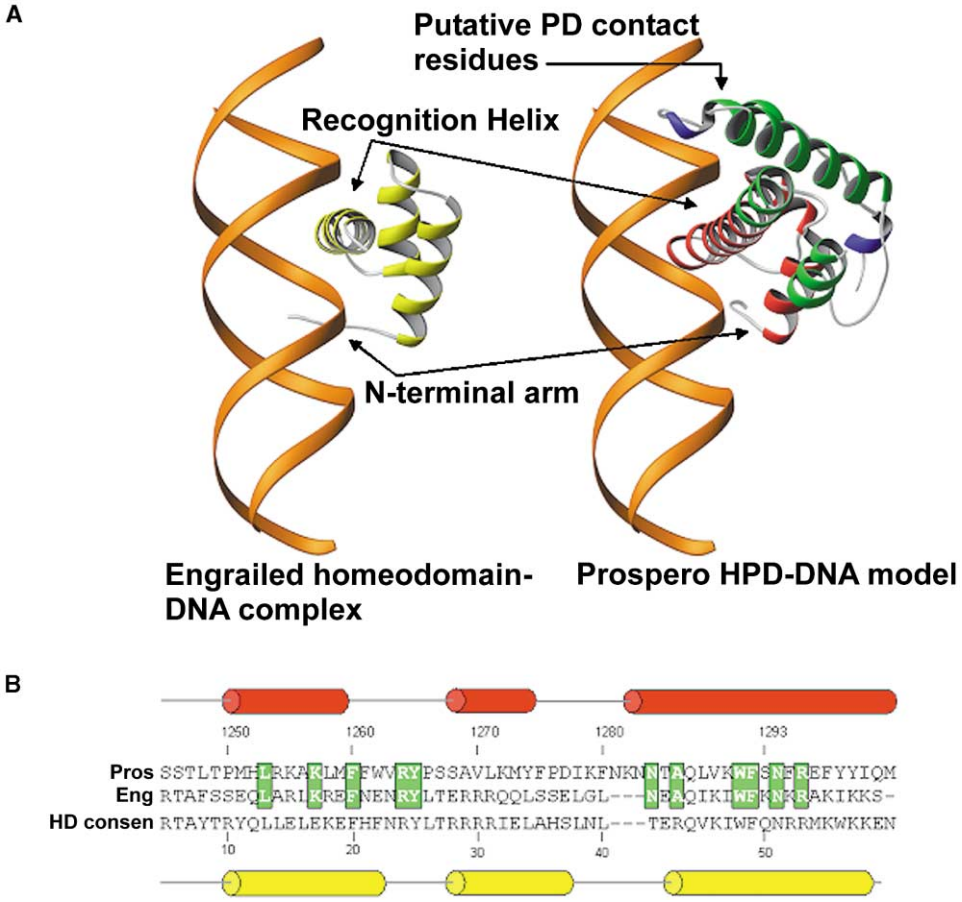

Figure 5. Model for the HPD-DNA Complex Based on the Structural Comparison with the Engrailed Homeodomain-DNA Complex and Sequence Alignment

(A) Side-by-side comparison of the Engrailed homeodomain-DNA complex and the Prospero HPD-DNA model. The DNA backbone is illustrated in orange, the Engrailed homeodomain in yellow, and the HPD as in Figure 3. A disordered region from residues 1314–1326 is not delineated. (B) Sequence alignment of the Prospero HD region (Pros), Engrailed homeodomain (Eng), and homeodomain consensus sequence (HD consen; based on 346 homeodomain sequences [30]). Secondary structure of the Prospero HD region (red) and Engrailed homeodomain (yellow) are shown above and below the sequences with their respective numbering. Invariant residues are shown in white on a green background.

Biological Implications

The Prospero transcription factor promotes neural differentiation in *Drosophila*, and its activity is tightly regulated by modulating its subcellular localization. Prospero is exported from the nucleus of neural precursors, but imported into the nucleus of daughter cells, which is necessary for their proper differentiation. Based on sequence similarity, the structure of Prospero was thought to contain, toward its carboxyl terminus, a homeodomain plus a Prospero domain, both of which are required for sequence-specific DNA binding. Here we determine the structure of this region of the protein and show that it does include the expected homeodomain. Instead of forming an essentially independent unit, however, the Prospero domain is shown to join together with the homeodomain to form a larger structural unit that we name a "homeo-prospero domain." Model building suggests that this larger structural unit serves in part to align the Prospero domain region on the DNA target. Also, the Prospero domain region is positioned in such a way that it is able to mask a defined nuclear export signal that is within the homeodomain region.

Experimental Procedures

Protein Overexpression and Purification

Escherichia coli BL21(DE3) Gold cells (Stratagene) were transformed with an expression vector encoding *Drosophila melanogaster* Prospero (C-terminal residues 1241–1403) fused to a noncleavable C-terminal hexahistidine tag. The expression vector was constructed by amplifying the Prospero gene with PCR mixtures containing Deep Vent DNA Polymerase (New England Biolabs) in a TECHNE thermocycler. The flanking restriction sites, NcoI and XhoI, were added with the following primers: 5′ end primer, 5′-CAT GCC ATG GGC ACT CCT TTG CAC TCT TCT ACA TTG ACA CCG AT-3′; 3′ end primer, 5′-CCG CTC GAG TTC CAG CTG CTC TAA AAA ATT GGG CGA CTT GA-3′. The PCR product was ligated into the Novagen pET-28b expression vector utilizing a Zero Blunt PCR cloning kit (Invitrogen) as an intermediary step to improve cloning efficiency, and the sequence was verified. Expression of the protein from pET-28b resulted in the addition of the C-terminal noncleavable hexahistidine tag LEHHHHHH.

Overexpression was performed using 4 L fermentation cultures. Cells were grown in LB-kanamycin broth (50 mg/L) at 37°C, 750

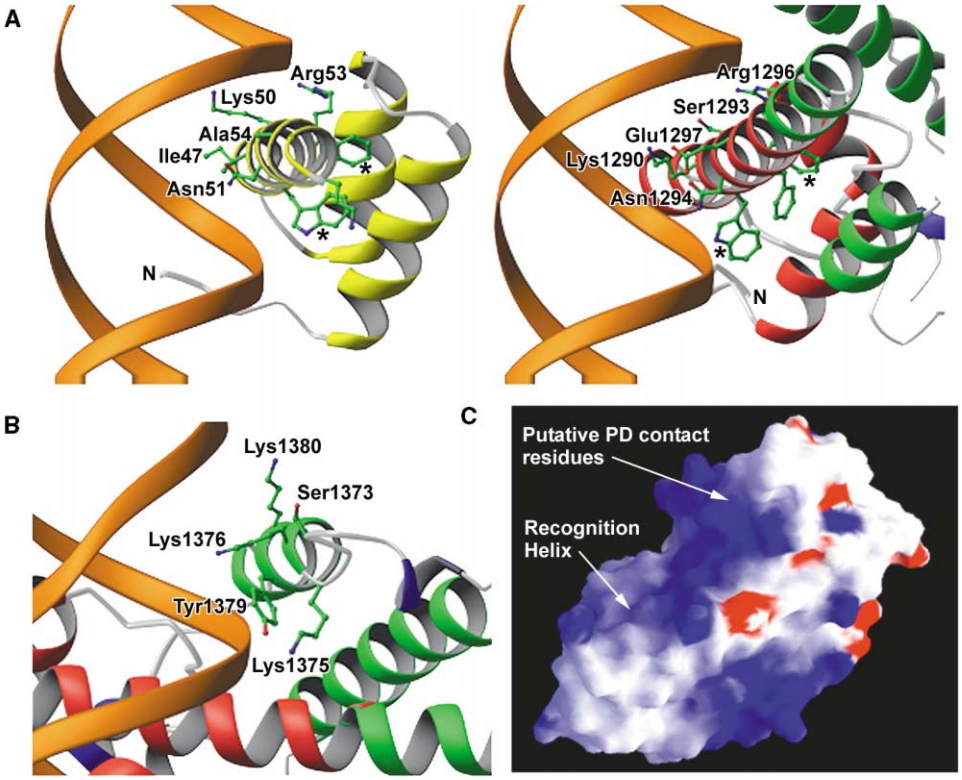

Figure 6. Detailed View of the HPD-DNA Complex Model

(A) A comparison of the homeodomain recognition helices of Engrailed (left) and Prospero HPD (right) as described in Figure 5A. In addition to the residues labeled, invariant Engrailed residues Trp48 (Pros HD Trp1291) and Phe49 (Pros HD Phe1292) are also indicated by asterisks. The N-terminal arms from each homeodomain are also visible in this view.

(B) A detailed view of the potential HPD contacts with the DNA backbone via the residues at the N terminus of the α_6 helix.

(C) GRASP [48] representation of the surface electrostatic potential of the HPD. Red and blue represent negative and positive electrostatic potentials, respectively.

rpm, and 10–15 L of air per minute to OD_{600} 0.9. The temperature was then reduced to 25°C and allowed to equilibrate for 15 min. Protein expression was induced by the addition of isopropyl β-D-thiogalactoside (IPTG) to 0.5 mM and continuing growth for 3 hr at 25°C and 250 rpm. The SeMet derivative was expressed with the same strains as described [37].

The cells were suspended in 50 ml of +T/G buffer (50 mM HEPES [pH 7.9], 10% glycerol, 0.1% Triton X-100, 0.5 M NaCl, 20 mM imidazole with 40 μg/ml DNase, 1 mM $MgCl_2$, 1 Complete/EDTA-free [Boehringer-Mannheim] inhibitor tablet), and lysed by passage through a French press. The lysate was cleared by centrifugation at 35,000 × g for 30 min. The supernatant was loaded onto an Ni-NTA agarose column (Qiagen, 10 ml) equilibrated with +T/G buffer (no DNase, $MgCl_2$, or inhibitor tablet) at 4°C. The column was washed with equilibration buffer followed by −T/G buffer (50 mM HEPES [pH 7.9], 0.5 M NaCl, 20 mM imidazole). The protein was eluted with 100 mM imidazole in −T/G buffer. The fractions corresponding to Prospero (residues 1241–1403), as determined by SDS-PAGE analysis, were pooled, 5 mM EDTA was added, and dialyzed against 4 L of 15 mM HEPES (pH 7.9), 150 mM NaCl overnight at 4°C. This was followed by an additional 4 hr dialysis in fresh buffer to ensure equilibration. The dialysate was cleared of minor precipitation by centrifugation at 35,000 × g for 30 min. Prospero (residues 1241–1403) was found to be 99.9% pure, as determined by SDS-PAGE. Glycerol was added to a final concentration of 10% and allowed to incubate on ice for 1 hr. Subsequently, the protein was concentrated using an Amicon Centriprep (10 KDa MWCO), spin filtered (#8160 Costar, Corning) to remove minor precipitation, and concentrated to the ~8 mg/ml range using an Amicon Centricon (10 KDa MWCO; ~6 mg/ml for SeMet derivative). Protein concentrations were determined by absorption at 280 nm using the theoretical extinction coefficient of Prospero (residues 1241–1403), 28,590 M^{-1} cm^{-1}, calculated using EXPASY ProtParam (http://us.expasy.org/tools/protparam.html). Typical yields were ~13 mg/4 L culture. The SeMet derivative was purified in an identical manner with typical yields of ~6 mg/4 L for SeMet derivatives and were concentrated to ~6 mg/ml. Matrix-assisted laser desorption ionization time of flight (MALDI-TOF) mass spectrometry was utilized to confirm SeMet incorporation.

Crystallization and Data Collection

Crystals of the SeMet derivative were obtained by hanging drop vapor diffusion. Equal volumes of protein (5.8 mg/ml, 15 mM HEPES [pH 7.9], 150 mM NaCl, 10% glycerol) and well solution (21% PEG 4000, 100 mM Tris [pH 8.5]) were mixed in the presence of 0.01 mM n-tetradecyl β-D-maltoside at room temperature. Crystals appeared overnight with rod-like morphology, growing to a typical length of 0.18 mm within a week (space group $P2_1$; unit cell dimensions a = 35.3 Å, b = 49.9 Å, c = 51.5 Å, β = 97.4°; one molecule/asymmetric unit). Cryogenic crystal preservation was achieved by the presence of 10% glycerol. The 2.05 Å multiwavelength data set (Table 1) was collected at 100 K at the Advanced Light Source (Berkeley Lab) on beamline 5.0.2. Data were merged and scaled using DENZO and SCALEPACK [38].

Structure Determination

The structure of Prospero (residues 1241–1403) was solved by MAD phasing using SeMet [39]. Four of five potential selenium sites were identified using SOLVE [40]. Two additional half-sites were located and MAD phases were refined using SHARP [41]. These initial phases were further improved by solvent flattening with SOLOMON

[42]. The experimental electron density for residues 1245–1396 was readily traceable despite absent density for the disordered loop residues 1314–1326. The model was built using O [43]. A final model (1351 atoms plus 83 oxygens from H_2O) was obtained after iterative rounds of refinement using CNS [44] and manual rebuilding.

Acknowledgments

We thank Dr. Gerry McDermott and Ms. Jane Tanamachi at the Advanced Light Source for assistance with data collection. We also thank Todd Lowther, Mike Quillin, Martin Sagermann, and Gerard Ostheimer for help and advice, and Ms. Jean Parker for editorial assistance. This work was supported in part by a National Institutes of Health National Research Service Award (to J.M.R., F32 GM20676-02) and support to B.W.M.

Received: May 23, 2002
Revised: August 8, 2002
Accepted: August 14, 2002

References

1. Doe, C.Q., Chu-LaGraff, Q., Wright, D.M., and Scott, M.P. (1991). The *prospero* gene specifies cell fate in the *Drosophila* central nervous system. Cell *65*, 451–465.
2. Vaessin, H., Grell, E., Wolff, E., Bier, E., Jan, L.Y., and Jan, Y.N. (1991). *prospero* is expressed in neuronal precursors and encodes a nuclear protein that is involved in the control of axonal outgrowth in *Drosophila*. Cell *67*, 941–953.
3. Li, L., and Vaessin, H. (2000). Pan-neural Prospero terminates cell proliferation during *Drosophila* neurogenesis. Genes Dev. *14*, 147–151.
4. Hirata, J., Nakaghoshi, H., Nabeshima, Y., and Matsuzaki, F. (1995). Asymmetric segregation of the homeodomain protein Prospero during *Drosophila* development. Nature *377*, 627–630.
5. Knoblich, J.A., Jan, L.Y., and Jan, Y.N. (1995). Asymmetric segregation of Numb and Prospero during cell division. Nature *377*, 624–627.
6. Spana, E., and Doe, C.Q. (1995). The prospero transcription factor is asymmetrically localized to the cell cortex during neuroblast mitosis in *Drosophila*. Development *121*, 3187–3195.
7. Ikeshima-Kataoka, H., Skeath, J.B., Nabeshima, Y., Doe, C.Q., and Matsuzaki, F. (1997). Miranda directs Prospero to a daughter cell during *Drosophila* asymmetric divisions. Nature *390*, 625–629.
8. Freeman, M.R., and Doe, C.Q. (2001). Asymmetric Prospero localization is required to generate mixed neuronal/glial lineages in the *Drosophila* CNS. Development *128*, 4103–4112.
9. Chu-Lagraff, Q., Wright, D.M., McNeil, L.K., and Doe, C.Q. (1991). The prospero gene encodes a divergent homeodomain protein that controls neuronal identity in *Drosophila*. Dev. Suppl. *2*, 79–85.
10. Matsuzaki, F., Koizumi, K., Hama, C., Yoshioka, T., and Nabeshima, Y. (1992). Cloning of the *Drosophila prospero* gene and its expression in ganglion mother cells. Biochem. Biophys. Res. Commun. *182*, 1326–1332.
11. Hassan, B., Li, L., Bremer, K.A., Chang, W., Pinsonneault, J., and Vaessin, H. (1997). Prospero is a panneural transcription factor that modulates homeodomain protein activity. Proc. Natl. Acad. Sci. USA *94*, 10991–10996.
12. Demidenko, Z., Badenhorst, P., Jones, T., Bi, X., and Mortin, M.A. (2001). Regulated nuclear export of the homeodomain transcription factor Prospero. Development *128*, 1359–1367.
13. Shen, C.P., Jan, L.Y., and Jan, Y.N. (1997). Miranda is required for the asymmetric localization of Prospero during mitosis in *Drosophila*. Cell *90*, 449–458.
14. Fuerstenberg, S., Peng, C.Y., Alvarez-Ortiz, P., Hor, T., and Doe, C.Q. (1998). Identification of Miranda protein domains regulating asymmetric cortical localization, cargo binding, and cortical release. Mol. Cell. Neurosci. *12*, 325–339.
15. Matsuzaki, F., Ohshiro, T., Ikeshima-Kataoka, H., and Izumi, H. (1998). Miranda localizes staufen and prospero asymmetrically in mitotic neuroblasts and epithelial cells in early *Drosophila* embryogenesis. Development *125*, 4089–4098.
16. Schuldt, A.J., Adams, J.H., Davidson, C.M., Micklem, D.R., Haseloff, J., Johnston, D.S., and Brand, A.H. (1998). Miranda mediates asymmetric protein and RNA localization in the developing nervous system. Genes Dev. *12*, 1847–1857.
17. Shen, C.P., Knoblich, J.A., Chan, Y.M., Jiang, M.M., Jan, L.Y., and Jan, Y.N. (1998). Miranda as a multidomain adapter linking apically localized Inscuteable and basally localized Staufen and Prospero during asymmetric cell division in *Drosophila*. Genes Dev. *12*, 1837–1846.
18. Oliver, G., Sosa-Pineda, B., Geisendorf, S., Spana, E.P., Doe, C.Q., and Gruss, P. (1993). Prox 1, a prospero-related homeobox gene expressed during mouse development. Mech. Dev. *44*, 3–16.
19. Burglin, T.R. (1994). A *Caenorhabditis elegans* prospero homologue defines a novel domain. Trends Biochem. Sci. *19*, 70–71.
20. Tomarev, S.I., Sundin, O., Banerjee-Basu, S., Duncan, M.K., Yang, J.M., and Piatigorsky, J. (1996). Chicken homeobox gene Prox 1 related to *Drosophila* prospero is expressed in the developing lens and retina. Dev. Dyn. *206*, 354–367.
21. Zinovieva, R.D., Duncan, M.K., Johnson, T.R., Torres, R., Polymeropoulos, M.H., and Tomarev, S.I. (1996). Structure and chromosomal localization of the human homeobox gene Prox 1. Genomics *35*, 517–522.
22. Glasgow, E., and Tomarev, S.I. (1998). Restricted expression of the homeobox gene prox 1 in developing zebrafish. Mech. Dev. *76*, 175–178.
23. Mizuno, N., Mochii, M., Yamamoto, T.S., Takahashi, T.C., Eguchi, G., and Okada, T.S. (1999). Pax-6 and Prox 1 expression during lens regeneration from Cynops iris and *Xenopus* cornea: evidence for a genetic program common to embryonic lens development. Differentiation *65*, 141–149.
24. Xu, C., Kauffmann, R.C., Zhang, J., Kladny, S., and Carthew, R.W. (2000). Overlapping activators and repressors delimit transcriptional response to receptor tyrosine kinase signals in the *Drosophila* eye. Cell *103*, 87–97.
25. 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.
26. Ceska, T.A., Lamers, M., Monaci, P., Nicosia, A., Cortese, R., and Suck, D. (1993). The X-ray structure of an atypical homeodomain present in the rat liver transcription factor LFB1/HNF1 and implications for DNA binding. EMBO J. *12*, 1805–1810.
27. Banerjee-Basu, S., Landsman, D., and Baxevanis, A.D. (1999). Threading analysis of prospero-type homeodomains. Silico Biol. *1*, 163–173.
28. Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. J. Mol. Biol. *233*, 123–138.
29. Tucker-Kellogg, L., Rould, M.A., Chambers, K.A., Ades, S.E., Sauer, R.T., and Pabo, C.O. (1997). Engrailed (Gln50→Lys) homeodomain-DNA complex at 1.9 Å resolution: structural basis for enhanced affinity and altered specificity. Structure *5*, 1047–1054.
30. Burglin, T.R. (1994). A comprehensive classification of homeobox genes. In Guidebook to the Homeobox Genes, D. Duboule, ed. (New York: Oxford University Press), pp. 27–71.
31. Kamtekar, S., and Hecht, M.H. (1995). Protein motifs. 7. The four-helix bundle: what determines a fold? FASEB J. *9*, 1013–1022.
32. Miller, S., Lesk, A.M., Janin, J., and Chothia, C. (1987). The accessible surface area and stability of oligomeric proteins. Nature *328*, 834–836.
33. Li, T., Jin, Y., Vershon, A.K., and Wolberger, C. (1998). Crystal structure of the MATa1/MATα2 homeodomain heterodimer in complex with DNA containing an A-tract. Nucleic Acids Res. *26*, 5707–5718.
34. Li, T., Stark, M.R., Johnson, A.D., and Wolberger, C. (1995). Crystal structure of the MATa1/MATα2 homeodomain heterodimer bound to DNA. Science *270*, 262–269.
35. Srinivasan, S., Peng, C.-Y., Nair, S., Skeath, J.B., Spana, E.P., and Doe, C.Q. (1998). Biochemical analysis of Prospero protein during asymmetric cell division: cortical Prospero is highly

phosphorylated relative to nuclear Prospero. Dev. Biol. *204*, 478–487.
36. Kissinger, C.R., Liu, B.S., 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.
37. Gassner, N.C. (1998). Structural, thermodynamic, and kinetic analysis of T4 lysozyme variants with methionine cores. PhD thesis, University of Oregon, Eugene, Oregon.
38. Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.
39. Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. Science *254*, 51–58.
40. Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D *55*, 849–861.
41. de La Fortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. Methods Enzymol. *276*, 472–494.
42. Abrahams, J.P., and Leslie, A.G. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Crystallogr. D *52*, 30–42.
43. Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for binding protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.
44. Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D *54*, 905–921.
45. Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. (1999). ESPript: multiple sequence alignments in PostScript. Bioinformatics *15*, 305–308.
46. Carson, M. (1997). Ribbons. Methods Enzymol. *277*, 493–505.
47. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.
48. Nicholls, A., Sharp, K.A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins *11*, 281–296.
49. Laskowski, R.A., MacArthur, M.W., and Thornton, J.M. (1998). Validation of protein models derived from experiment. Curr. Opin. Struct. Biol. *8*, 631–639.

Accession Numbers

Coordinates and structure factors for the homeo-prospero domain have been deposited in the Protein Data Bank under ID code 1MIJ.