Crystal Structures of Engrailed Homeodomain Mutants

IMPLICATIONS FOR STABILITY AND DYNAMICS*§

Elliott J. Stollar, Ugo Mayor‡, Simon C. Lovell, Luca Federici, Stefan M. V. Freund‡, Alan R. Fersht‡, and Ben F. Luisi§

From the Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, United Kingdom and the §Medical Research Council Centre for Protein Engineering, Hills Road, Cambridge CB2 2QH, United Kingdom

We report the crystal structures and biophysical characterization of two stabilized mutants of the Drosophila Engrailed homeodomain that have been engineered to minimize electrostatic repulsion. Four independent copies of each mutant occupy the crystal lattice, and comparison of these structures illustrates variation that can be partly ascribed to networks of correlated conformational adjustments. Central to one network is leucine 26 (Leu26), which occupies alternatively two side chain rotameric conformations (−gauche and trans) and different positions within the hydrophobic core. Similar sets of conformational substates are observed in other Engrailed structures and in another homeodomain. The pattern of structural adjustments can account for NMR relaxation data and sequence co-variation networks in the wider homeodomain family. It may also explain the dysfunction associated with a P26L mutation in the human ARX homeodomain protein. Finally, we observe a novel dipolar interaction between a conserved tryptophan and a water molecule positioned along the normal to the indole ring. This interaction may explain the distinctive fluorescent properties of the homeodomain family.

The homeodomain is a simple fold common to many different DNA-binding proteins from diverse eukaryotes. The domain is found in transcription factors that regulate a variety of genes and so control key processes ranging from early developmental decision to homeostasis and general “housekeeping” (1–3). The homeodomain fold comprises three helices, connected by a flexible loop and a β-turn, and has a small hydrophobic core that is remarkably well conserved (4). The side chains in this conserved core appear to be well packed, which is a hallmark of stable folds. However, homeodomain folds generally have poorer stabilities in comparison with other proteins of similar size (5–8). The termini of homeodomains are disordered in solution (9–11) and structures of free-and DNA-bound forms illustrate that DNA binding is accompanied by a structural condensation of the termini (12). Possibly, the homeodomain core also undergoes subtle structural changes upon DNA binding. Thus, the homeodomain would appear to mold to the surface of specific DNA by induced fit. Despite the wealth of current stereochemical and functional data for homeodomains, the basis for their comparatively low stability and its relationship, if any, to their induced fit remains to be established.

Sequence co-variance analysis of homeodomains highlights several strongly correlated residue pairs that form conserved interactions and that may affect stability and DNA binding (4). One pair identified was the surface residues 17 and 52, which forms, in the majority of cases, a salt bridge. However, in the transcriptional repressor Engrailed from Drosophila, two lysines are instead found in these positions. We have substituted lysine 52 for glutamate, to mimic the conserved Glu17-Arg52 salt bridge, and for alanine, to relieve the repulsion of the clustered positive charges. Standard equilibrium measurements demonstrate that both substitutions substantially increase the stability of the folded state.

We have also determined the crystal structures of the Engrailed homeodomain (En-HD)1 mutants K52E and K52A, under similar crystallization conditions, to 2.1-Å resolution. These structures, together with solution data, suggest that the stabilization is because of the relief of electrostatic repulsion. There are four independent copies in the asymmetric unit, and for each mutant we observe significant conformational differences between the individual molecules that can be partly attributed to correlated orientations of side chains. The positional variations in the Leu26 network can account for backbone NMR relaxation data from the wild type protein, which indicates greatest conformational exchange in the intrahelical loop containing Leu26. The interactions within the Leu26 network may account for the observed co-variation of the participating residues among the diverse homeodomain family members. Our observations may also explain the dysfunctional phenotype associated with a P26L mutation in the human ARX homeodomain, which is found in cases of rare X-linked myoclonic epilepsy (13, 14). These results provide further structural insights into the determinants of stability and functional specificity for this highly conserved domain.

* This work was supported by the Wellcome Trust, the Biotechnology and Biological Sciences Research Council, the Medical Research Council, the European Union Marie Curie program (to L. F.), and “Ikertzaileen prestakuntza” grant (to U. M.) from the government of the Basque Country. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1P7I and 1P7J for K52A and K52E, respectively) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

§ The on-line version of this article (available at http://www.jbc.org) contains animations.

‡ To whom correspondence should be addressed: Dept. of Biochemistry, University of Cambridge, 80 Tennis Court Rd., Cambridge CB2 1GA, UK. E-mail: Ben@cryst.bioc.cam.ac.uk.

Received for publication, July 23, 2003
Published, JBC Papers in Press, August 15, 2003, DOI 10.1074/jbc.M308029200

© 2003 by The American Society for Biochemistry and Molecular Biology, Inc.

This paper is available on line at http://www.jbc.org
**Experimental Procedures**

Mutagenesis and Protein Expression—Mutagenesis was carried out using the QuickChange technique (Stratagene). Purification of the wild type protein has been described previously (15), and the mutants were purified in the same fashion. Protein purity and quality were corroborated by SDS-PAGE and matrix-assisted laser desorption ionization mass spectrometry.

X-ray Crystallography—Crystals were grown at 20 °C by the vapor-diffusion method from hanging droplets by mixing 40 mg ml⁻¹ protein (in water) in a 1:1 volume ratio with 30% (w/v) polyethylene glycol 3000, 100 mM CHES, pH 9.5. The crystals appeared within 1 week and were refined in water with a Microlab 500 titrator. Differential scanning calorimetry experiments were performed in a VP-DSC instrument (MicroCal) with a cell volume of 0.52 ml, at 1°min. All equilibrium data fitted well to the standard two-state equations.

NMR Spectroscopy—NMR experiments were carried out on a DRX 600 spectrometer equipped with inverse triple resonance probes and single axis gradients. Protein samples in the range of 0.5–2 mM were prepared in 50 mM sodium acetate buffer, pH 5.7, with 100 mM NaCl and 5% D₂O. All data were processed in FELIX 98 ( Biosym Technologies, San Diego, CA).

15N T₁ and T₂ experiments incorporated spin-echo delay times of 500 μs, the relaxation times were set to 10, 20, 40, 60, 80, 120, 200, 240, 280, 320, and 360 ms. The recycle delays for the T₁ experiments were 3.0 s. Each spectrum comprised 512 × 128 complex points that were zero-filled 2-fold in both dimensions and multiplied by exponential window functions in both dimensions prior to analysis. Peak intensities were fitted to single exponential equations in KaleidaGraph 3.0 (Abelbeck Software) to determine R₁ (1/T₁) and R₂ (1/T₂) rates for individual amides. The uncertainties of the R₁ and R₂ data were estimated from non-linear least squares fitting according to the Levenberg-Marquardt algorithm. Heteronuclear 15N-H steady-state nuclear Overhauser effect (NOE) enhancements were determined as previously described (5).

Cross-correlation between 15N chemical shift anisotropy and 15N-H dipolar coupling was measured according to Bax and co-workers (24). Cross-correlation experiments were acquired at 600 MHz with 1024 × 128 complex points and 128 scans per increment. The durations for the dephasing delays were set according to 46.7, 68, and 132 ms. Resonance intensities were obtained from peak heights using three data point interpolations. The cross-correlation term, η, was calculated according to the equation I/Ip = tanh(2A η), where I is the cross-correlation according to scheme A, Ip the cross-correlation with emission of H9004/180 pulse (scheme B), η is the dephasing delays, and Δ the cross-correlation between 15N chemical shift anisotropy and 3N-H dipolar coupling (24). Data were fitted using the Kaleidagraph software. Small differences in the relaxation losses in the two schemes because of potential pulse imperfections were not corrected.

Computational—Structures were superimposed using CompareR (25) and the McLachlan superposition algorithm as implemented in the program ProFit. Root mean square deviations were calculated with ProFit, side chain dihedrals with Dang. Angles between helices were calculated by building a fixed geometry α-helix (α = −47° φ = −56°) in a standard frame of reference and superimposing this on each helix in the structure in turn. The transformation matrices required for superposition allowed calculation of the relative helix orientations. The figures were prepared using VMD (26) and Grasp (27).

**Results**

**Stability of the En-HD Fold**—The same denaturation m-value of 0.80 ± 0.05 kcal mol⁻¹ m⁻¹ was obtained from urea denaturation curves (ΔGm=Urea/ΔUrea) for the wild type and the mutants (data not shown). Apparent free energies of denaturation (ΔGm=Urea) were increased for K52A and K52E compared with the wild type by, on average, about 1.2 and 1.8 kcal mol⁻¹, respectively (Table II). The apparent free energy of denaturation was validated using MOLPROBITY (22) and PROCHECK (23).

**Equilibrium Denaturation Experiments**—Urea denaturation of wild type En-HD, K52A, and K52E (10 μM) in 50 mM sodium acetate buffer, 100 mM NaCl, pH 5.7, was monitored by fluorescence (excitation 280 nm, emission cut-off 320 nm) using an Aviv 202SF instrument equipped with a Microlab 500 titrator. Differential scanning calorimetry experiments were performed on wild-type En-HD, K52A, and K52E (in the range 300–350 μM) and were performed in a VP-DSC instrument (MicroCal) with a cell volume of 0.52 ml, at 1°min. All equilibrium data fitted well to the standard two-state equations.

**NMR Spectroscopy**—NMR experiments were carried out on a DRX 600 spectrometer equipped with inverse triple resonance probes and single axis gradients. Protein samples were solved using molecular replacement (EMPR (17) and AMORE (18)) with the wild type structure (1ENH) as the search model. To minimize phase bias, only the peptide backbone was used for initial phase calculations and additional atoms were added or removed iteratively using arpWarp (19). Initially the structure was refined using CNS (20) and at later stages with REFMAC (21) including translation-libration-screw displacement refinement. The model was built using 2Fo – Fc, maps, Fc – Fo, maps, and composite omit maps. There were two copies in the asymmetric unit of crystals of both K52A and K52E mutants, and non-crystallographic symmetry was applied in the first stages of refinement and then released later on. This procedure gave the best R-free to R-factor ratio and meaningful crystal contacts. In parallel trials, refinement with non-crystallographic symmetry applied throughout gave inferior structures as judged by real space refinement scores, R-free/R-factor ratios, and map quality. The stereochemistry

---

**Table I**

| Mutant | K52E | K52A |
|--------|------|------|
| Space group | P2₁2₁2₁ | P2₁2₁2₁ |
| Unit cell dimensions (Å) | a = 44.85 | a = 44.80 |
| | b = 51.68 | b = 51.26 |
| | c = 112.79 | c = 108.63 |
| Copies in A.U. | 4 | 4 |
| Resolution (Å) | 20.2-2.1 | 20.2-2.1 |
| Measured reflections | 159,036 | 59,319 |
| Unique reflections | 14,837 | 14,992 |
| Completeness to 2.1 Å (%) | 98.3 (90.4) | 98.3 (91.0) |
| (last shell) | 98.3 (91.0) | 98.3 (91.0) |
| R-merge (%) | 6.4 (17.6) | 7.7 (19.4) |
| Mosaicity (°) | 0.37 | 0.704 |
| Average Wilson B factor | 23.1 | 28.1 |
| Mean B value | 16.3 | 16.8 |
| R-cryst (%) | 19.8 | 20.0 |
| R-free (%) | 24.1 | 24.0 |
| Number atoms in | 1913 | 1834 |
| refinement | | |
| R.m.s. deviation of bond | 0.006 | 0.011 |
| lengths (Å) | | |
| R.m.s. deviation of bond | 0.93 | 1.29 |
| angles (deg) | | |
| Ramachandran analysis | | |
| No. of residues in favored | 194 (96.5) | 200 (99.5) |
| region (%) | 194 (96.5) | 200 (99.5) |
| No. of residues in allowed | 7 (3.5) | 1 (0.5) |
| region (%) | 0 (0) | 0 (0) |
| No. of residues in outlier | 0 (0) | 0 (0) |

---

---
ation of the wild type protein was 2.1 kcal mol\(^{-1}\), as determined by calorimetric measurements, which is small for a protein of this size (28). DNA-binding proteins tend to have lower stabilities than proteins of corresponding mass, with reported \(\Delta G_{D,N}\) ranging from 3.0 to 3.7 kcal mol\(^{-1}\) (29–31). But even so, the En-HD protein is still only marginally stable in comparison.

Crystal Structures of K52E and K52A: the Basis for Stability—Both mutants contain four independent copies in the asymmetric unit; labeled A, B, C, and D. Representative electron density is shown in Fig. 1 and a schematic representation of the fold is shown in Fig. 2A.

The multiplicity of copies has allowed us to confirm the features at the site of mutation in each case. In the wild type structure (1ENH), Lys\(^{52}\) is sandwiched between Phe\(^{20}\) and Trp\(^{48}\), and its charged amino group protrudes into solution. Lys\(^{55}\) and Lys\(^{17}\) are in close proximity with Lys\(^{52}\) (Fig. 2B), which possibly causes charge repulsion. The electrostatic surfaces of K52E and K52A show that mutation of residue 52 to a glutamate (Fig. 2C) or an alanine (Fig. 2D) reduces the amount of positive charge clustered on the surface, as expected.

For the K52E mutant, we observe that the glutamate packs against the conserved residues Trp\(^{48}\) and Phe\(^{20}\) in all four copies, and it forms (in some copies) simultaneous salt bridges to Lys\(^{55}\) and Lys\(^{17}\) (Fig. 2E). In the crystal structures of both the wild type and K52A mutant, the third helix is disordered after residues 55–56, whereas in the K52E mutant this helix can be resolved for a further 1–2 residues. This is most likely because of the anchoring of Lys\(^{55}\) through its salt bridge with Glu\(^{52}\). The positions of the residues around Glu\(^{52}\) in the K52E mutant are shown in an animation, which highlights the positional variation of the lysines around the glutamate (E52.avi; see Supplementary Materials).

For the K52A mutant, the methyl group of the alanine packs against Trp\(^{48}\) and Phe\(^{20}\) (Fig. 2F), but this interaction represents a fraction of the non-polar area of the lysoine side chain in the wild type structure or the glutamate side chain in the K52E structure. The alanine is unable to make any interactions with the surrounding lysines; these consequently adopt different positions and in some cases appear completely disordered. As this mutation is substantially stabilizing, it seems that the relief of electrostatic repulsion must outweigh energetically the loss of favorable packing interactions and is likely to be the major driving force of the stabilization in both K52E and K52A mutants.

A Conserved Indole-Water Polar Interaction—Tryptophan is often observed to form hydrogen bonds with water molecules that lie in the same plane as the indole ring.\(^5\) The ring of aromatic residues can also act as a hydrogen-bond acceptor from a group oriented normal to the center of the aromatic ring (32). In the En-HD crystal structures, we find a water molecule

\(\text{FIG. 1.} \) Stereo image of a sample electron density for K52E copy A contoured at \(1\sigma\).

\(\text{FIG. 2.} \) Structure and electrostatic surfaces of En-HD and its mutants. \(A\), schematic illustration of the homeodomain fold as represented by copy A of the mutant K52E. The side chains of residues 17, 48, 52, and 55 are shown. \(B\), electrostatic surface of wild type, K52E copy A \((C)\), and K52A copy A \((D)\). The view is from the same perspective as shown in \(A\). For clarity and for purposes of comparison (because of poor density in some cases), the side chains of residues Arg\(^{24}\) and Arg\(^{29}\) were not included and the N terminus begins at residue 7. \(E\), cartoon depiction of the site of mutation for the K52E mutant, and \(F\), the K52A mutant. Each copy is color coded, copy A (blue), copy B (red), copy C (yellow), and copy D (pink). The dashed lines indicate salt bridge links.

\(^5\) www.biochem.ucl.ac.uk/~mcdonald/atlas.
lying along the normal to the indole plane, but importantly, the water is positioned directly above the heterocyclic nitrogen rather than the center of the ring (Fig. 3A).

A water molecule is present at this position in 3 of the 4 copies of both K52A and K52E mutants, and it is oriented through a hydrogen bond from the backbone carbonyl of Trp48. In the case of K52E, the water molecule forms an additional hydrogen bond with the Glu52 side chain (Fig. 3A). The persistence of this water molecule at the same position in the wild type protein and its occurrence in the corresponding position in the MATa1 homeodomain structure (1MH3 (33)) suggests that the water-indole interaction may be important. Water molecules are also found at corresponding positions in several homeodomain-DNA complexes, where the water molecule donates a hydrogen bond to the phosphate backbone (34–38) (Fig. 3B). In fact, a data base search has found hundreds of examples of similar water-indole interactions in protein structures (results not shown).

Inspection of the hydrogen bonding patterns in the protein-DNA complexes suggests that neither the lone pair nor the protons of the water molecule are directed toward the indole nitrogen. The interaction may therefore be dipolar, which is supported by calculations that show that the interaction has favorable electrostatic and van der Waals components.6

The water-indole interaction could explain the distinctive fluorescence properties of homeodomains. With refolding, fluorescence quenching is observed for the En-HD, antennapedia, bicoid, and ubx homeodomains (39, 40) as well as in paired, msx-1, oct-1, oct-2, and pit-1 homeodomains.7

It has been argued that the aromatic residue at position 8 is largely responsible for the quenching as a consequence of its \( \pi \) system receiving a H-bond from the indole nitrogen (40). However, the fluorescence of the F8A mutant of En-HD has the same degree of quenching as the wild type (41). In the case of En-HD, Trp48 is in close proximity of the terminal amino groups of Lys17 and Lys35, which are also potential quenchers.

We have tested the fluorescence quenching of both K52A and K52E, where the positions of Lys17 and Lys35 are different (Fig. 2, E and F) and both, surprisingly, show the same magnitude as wild type (data not shown). We thus are lead to consider that the water molecule is responsible for the majority of the quenching. In fact the mutant K52L, which modeling suggests could displace this water, has 50% less quenching while, in the buffer conditions of the experiment, retains similar stability to wild type (data not shown). Based on these observations, we propose that the characteristic fluorescence quenching in homeodomains is caused by the dipolar interaction between the water and the excited transition state of the indole.

**Conformational Variation**—The wild type structure superimposes onto all copies of both mutants with a backbone root mean squared deviation (r.m.s.d.) range of 0.21–0.43 Å. The interhelical angles vary by less than 5°. Whereas the backbone positions do not vary appreciably in the K52A and K52E structures, the positions of the long side chains on the protein surface vary substantially (Fig. 4A). For many of these residues, the side chain positions correlate with alternative conformations of at least one neighboring side chain. Variations are also because of different direct crystal contacts and are indicated where relevant to the analysis. These crystal contacts are distributed over the surface of the proteins and are different for each of the copies and, even to some degree, between the corresponding copies of the two mutants (Fig. 4B). Many residues also differ in their \( \chi_1 \) side chain torsion angle for the different copies (Fig. 4C).

The r.m.s.d. for each side chain have been ranked and color coded in the refined structure of the K52E mutant, as shown in Fig. 4D. Generally, the core residues (mostly blue in Fig. 4D) change position and \( \chi_1 \) values of Leu26. This residue also appears to make the most pairwise correlated adjustments, and we will focus further on this residue in the next section.

**The Core Residue Leu26**—One of the greatest differences between copies is in the position and \( \chi_1 \) values of Leu26. This residue resides in the loop between helices 1 and 2 and is part of an important hydrogen bonding site with the backbone. The Core Residue Leu26

---

6 E. J. Stollar, J. Luis, S. Velankar, A. Golovin, M. Orozco, B. F. Luisi, manuscript in preparation.
7 U. Mayor and A. R. Fersht, unpublished results.
of the hydrophobic core of the protein. Relative to the rest of the core, the Cα of Leu26 spans about 1.8 and 1.4 Å between copies for K52E and K52A, respectively. It switches between its two most favored χ1 rotameric states (-gauche and trans) as defined by Markly et al. (42). These different conformations are large compared with the estimated coordinate error of the model of 0.2 Å.

Residues Phe49, Arg30, Arg31, Ile45, and Lys 46 create a hydrophobic enclosure for Leu26. In Fig. 5A, all the residues involved in Leu26 packing in the K52E mutant and the wild type structure (1ENH) are shown (except Ile45, for clarity). This view shows Leu26 to be distributed between the -gauche (copies B and C) and trans (copies A and D and wild type) rotameric states with a significant change in position. A similar change in position of Leu26 has 50% greater buried surface area (Fig. 6A). An animation of the substates of the Leu26 network for both K52A and K52E mutants (L26A.avi and L26E.avi, respectively) can be found in the Supplementary Materials. The animation makes more apparent the correlated positions of the residues surrounding Leu26, which we discuss further below.

**Positional Correlation of Leu26 and Neighboring Residues**—Through analysis of Leu26 positions, it became clear that the contacting residues varied in position as well. There is a clear positional correlation between Arg30 and Leu26 that can be illustrated by comparing the two extremes of this network. In the Leu26 -gauche state, Arg30 is fully extended and makes most contacts with its guanido group to surrounding polar groups in the loop between helices I and II (Fig. 6B). Furthermore, the positional change can also be observed in the superposition of the wild type (3HDD) and Q50A (1DOU) En-HDs in their DNA bound states (Fig. 5, C and D).

The van der Waals packing interactions between the residues in a Leu26 -gauche (copy B) and Leu26 trans (copy A) network for K52E are shown in Fig. 6, A and B. The packing interactions in the two states are quite different, highlighting the importance of the position of Leu26 to the network of its contacting residues. In the trans position Leu26 has 50% greater buried surface area (Fig. 6A). An animation of the substates of the Leu26 network for both K52A and K52E mutants (L26A.avi and L26E.avi, respectively) can be found in the Supplementary Materials. The animation makes more apparent the correlated positions of the residues surrounding Leu26, which we discuss further below.

**Fig. 4. Distribution of side chain orientations in the asymmetric unit.** A, r.m.s.d. for each side chain between all 4 copies. K52E is labeled in blue and K52A in red. These values were calculated using ProFit. Where residues have no density in one or more copies, no values have been calculated for that residue. B, crystal contacts, as defined as the separation between symmetry related atoms of less than 3.5 Å. The first 8 lines correspond to the 4 copies A, B, C, and D for each mutant as indicated. The black squares are contacts for K52A, the red squares are for K52E, and the blue squares are for the wild type structure (1ENH). C, χ1 angles plotted for each copy in K52E. The color coding is the same as described in the legend to Fig. 2. D, the three bands of r.m.s.d. values were plotted onto copy C of K52E. Blue residues correspond to an r.m.s.d. range of 0–0.6 Å, orange >0.6–1.2 Å, and red, >1.2–1.8 Å.
span across helices 1 and 2 and loses two hydrogen-bonding interactions to the surrounding residues. The electron density at Arg30 for the Leu26 trans conformation (copy D) also indicates that Arg30 is in equilibrium with an alternative conformation where it loses all contacts with the loop (side chain indicated in green). Consistent with this observation, the crystal structures of the wild type En-HD protein in complex with DNA (43, 44) also show that Arg30 assumes multiple conformations (Fig. 5, C and D).

There is also a positional correlation between Ile45 and Leu26. With the switch in Leu26 conformation, Ile45 alternates between its two most favorable rotameric states (Fig. 6, A and B). Last, Leu26 packs against Arg31 and Lys46 to varying degrees.

In summary, it is clear that the changes in positions for many of the residues contacting Leu26 are correlated (this is shown most strikingly in the accompanying animations, Supplementary Materials). The results for the independently refined K52A structure are highly similar and show the same trend despite several different crystal contacts between equivalent copies (Fig. 4B).

**Backbone Dynamical Properties of En-HD**—The crystal structures of the K52E and K52A mutants show significant conformational variations in the side chains of Leu26 and its surrounding residues that are consistent with the wild type crystal structures. To explore whether these correlations might occur in solution, backbone dynamics of the wild type En-HD were evaluated by NMR. 15N longitudinal relaxation rate constants ($R_1$), 15N transverse relaxation rate constants ($R_2$), and the 15N-1H NOE, cross-correlation between 15N Chemical shift anisotropy and 15N-1H dipolar interaction ($η$) were measured for the amide resonances of wild type En-HD at 5 °C (Fig. 7) and 25 °C (data not shown). Both the N and C termini (residues 0 to 7 and 56 to 59, respectively) were found to be flexible, but the rest of the protein was largely rigid, except for the loop between helices I and II, and the turn between helices II and III, which displayed increased individual mobility relative to the mean rotational
The observed boundaries for the secondary structure elements are in agreement with our crystallographic data.

Interestingly, an increase of $R_2$ was observed in the loop between helices I and II (Fig. 7B), suggesting that the amide groups in the loop undergo chemical exchange processes on the microsecond to millisecond time scale. This was further supported by the lack of correlation observed between the cross-correlation term $\gamma$ and $R_2$ values for the residues in the loop (Fig. 7C). Similar chemical exchange contributions have been observed for the MATa1 homeodomain (45).

The NMR observations thus indicate that the intrahelical loop containing Leu$^{26}$ is in conformational exchange, and the crystal structures indicate that Leu$^{26}$ may contribute to this exchange both directly and indirectly. First, Leu$^{26}$ is itself varying in position, and second it correlates with the position of Arg$^{30}$. As we discussed in the previous section, Arg$^{30}$ makes varying hydrogen bonding interactions to the loop backbone; indeed in one case the hydrogen bonds appear to be lost entirely (Fig. 5, C and D). Taken together, the crystallographic and NMR results are complementary and suggest that the multiple conformations of the helix I/helix II loop exist, predominantly, through the substrates sampled by Leu$^{26}$, Arg$^{30}$, and the network of contacting residues.

FIG. 6. Comparing the Leu$^{26}$-gauche and trans network. A, packing in copy D from K52E representing a Leu$^{26}$ trans network. B, packing in copy B from K52E representing a Leu$^{26}$-gauche network. C, Arg$^{30}$ interactions in copy D from K52E representing a Leu$^{26}$ trans network. Arg$^{30}$ is modeled in a second conformation (green). D, Arg$^{30}$ interactions in copy B from K52E representing a Leu$^{26}$-gauche network. The dots represent van der Waals contacts between atoms. The dotted lines indicate H-bonds. Molecular interactions were calculated with Probe (60) after addition of hydrogen atoms with Reduce (61) and displayed in Mage (62).

DISCUSSION

A detailed analysis of the mutant Drosophila En-HDs, K52E, and K52A show that the conformations of the protein vary significantly between independent copies of the crystallographic asymmetric unit, which is likely to reflect plastic deformation in response to the distinct lattice contacts. Matthews and others (46–48) have argued that packing interactions in protein crystals are relatively weak, so that the perturbations they cause are representative of low-energy changes in conformation that are characteristic of protein structures generally. We have cross-compared each conformation to separate direct crystal contact effects from other more interesting phenomena such as correlated side chain positions and flexibility. Therefore, with careful inspection, we have used the differing crystal environments to explore the protein structure, beyond an analysis of just a solitary molecule.

Stabilization of the En-HD Fold—The stability of the En-HD fold is unusually small for a protein of this size, and our structural and biophysical data illustrate that the density of like-charges on the surface of the En-HD protein is unfavorable. However, this cluster seems to support DNA binding, as the target site affinity of K52A and K52E are reduced 20- and 1000-fold, respectively, in comparison with the wild type En-
Thus residue 52, although not involved in any obvious direct interaction with the DNA, nonetheless contributes to the electrostatic surface of the protein. This phenomena has been noted in studies with the ribonuclease barnase, where mutation of several of the functional positively charged residues caused large decreases in activity but an increase in stability (49).

The relief of repulsion appears to be the major source of stabilization in both mutants, and the additional stability of the K52E mutant is attributable to favorable hydrophobic packing and new salt bridges. This increase in stability is similar to the mutation H52R in the VND homeodomain that makes similar additional contacts (50). Surprisingly the K52E mutation achieves nearly the same increase in stability obtained by Marshall and colleagues (51) who mutated 25 residues in the En-HD through computational design.

In total our analysis agrees with previous studies (47, 52) that relieving electrostatic repulsions serves to increase stability while not perturbing the structure. Salt bridge formation only contributes a small additional energy benefit, probably because its full benefit requires extensive cooperating bridges (53). Interestingly, for the wild type the repulsion can also be relieved by addition of NaCl. In fact the free energy of denaturation is zero in the absence of salt, whereas addition of 500 mM NaCl stabilizes the native state by 3 kcal mol$^{-1}$ (54).

**A Special Interaction between Water and Trp48**—We observe a distinctive interaction between the indole ring nitrogen of Trp48 and an off-plane water molecule. This interaction is likely to be dipolar in nature, and it is likely to be favorable, which may explain why this contact has persisted in the evolution of the homeodomain family. We also note that this water is a key contact for DNA binding: a water molecule in the corresponding position is seen to mediate a contact to the phosphate backbone in several homeodomain-DNA complexes. The interactions may also occur more widely in protein structures: a data base search has found similar indole-water interactions that appear to support certain peptide geometries in other folds. Finally, we note that the homeodomains exhibit a distinctive tryptophan fluorescence quenching (39, 40), and we propose that this is caused by the water-indole interaction.

**Leu26 Network: Implications for Induced Fit**—We have identified a network of residues that mutually pack in several different ways. The residues involved include Leu$^{26}$ and Ile$^{45}$, which are both part of the hydrophobic core, and Arg$^{30}$, which bridges helices I and II. The conformations of these three residues loosely define structural substates, and closely matching substates can be found in other En-HD structures including the DNA-bound forms (Fig. 5, A, C, and D). Interestingly, the two crystal forms of MatA1 show a 4.2-A difference in position for Leu$^{26}$ and reveal correlated positions for the salt bridging residues Lys$^{23}$/Glu$^{30}$ (Fig. 5B). One of the substates of the free MatA1 (involving residues Leu$^{26}$, Glu$^{30}$, and Lys$^{23}$ in 1MH3) closely matches the conformation seen in the MatA1/MatA2/DNA crystal structure (55).

We hypothesize that the free homeodomain proteins exist in an ensemble of substates centered on Leu$^{26}$. Backbone NMR relaxation data support the proposal for the existence of substates around the helix I/helix II loop (where Leu$^{26}$ lies), as significant conformational exchange is observed on the microsecond to millisecond time scale. This result is consistent with

---

8 M. Watson and J. Thomas, personal communication.
The position of Leu[44] might have consequences for DNA binding. In En-HD, Leu[26] directly contacts two DNA-binding residues (Lys[46] and Arg[31]) and, through these residues, has an indirect interaction with the key specificity determining residue, Gln[50] (Figs. 4 and 5). In fact, there is a clear positional interdependence between these three DNA-binding residues (see Supplemental Materials for a visual illustration), which Pabo and co-workers (57) have also described in their analysis of the En-HD Q50A-DNA structure. It thus seems likely that the conformation of Leu[26] will influence the position of the DNA-binding residues.

The substrate organization of the core and DNA-binding surface may allow En-HD to make an induced fit to target DNA. In support of this proposal, we note that other molecular association events have been attributed to correlated side chain positions and multiple side chain conformations (48). For example, in the case of the 1-Å resolution structure of calmodulin, Wilson and Brunger (58) describe multiple conformations for residues in the binding pocket for its target proteins. They hypothesize that correlated side chain disorder in calmodulin assists target-specific deformation of these pockets. Taken together, the crystallographic analysis presented here and in earlier studies support the proposal that correlated side chain networks provides a mechanism for induced fit in homeodomains.

**Sequence Analysis**—Clarke identified a dominating pattern of pairwise co-variation centered on Leu[26] and including residues 44, 46, 47, 50, and 54 (4). Based on these observations alone, it was suggested that structural changes in the co-varying residues accompany DNA binding. The network of interacting residues identified by crystallography is in accord with the co-evolving network and supports our proposal that DNA binding is accompanied by induced fit.

Using the co-varying network, homeodomains can be divided into two classes. One class has branched aliphatic residues at position 26, whereas the second contains proline at that position. Our analysis of 129 representative human homeodomain sequences (59) finds that the branched aliphatic subgroup tends to have a potential salt bridge between residues 19 and 44, 46, 47, 50, and 54 (4). Based on these observations alone, it was suggested that structural changes in the co-varying residues accompany DNA binding. The network of interacting residues identified by crystallography is in accord with the co-evolving network and supports our proposal that DNA binding is accompanied by induced fit.

The importance of position 26 and its network of interacting residues identified by crystallography is in accord with the arguments made here.
39. Subramaniam, V., Jovin, T. M., and Rivera-Pumar, R. V. (2001) J. Biol. Chem. 276, 21506–21511
40. Nanda, V., and Brand, L. (2000) Proteins 40, 112–125
41. Guydosh, N. R. (2002) A Protein-folding Study of Mutants of the Engrailed Homeodomain. M. Phil. thesis, University of Cambridge, Cambridge, United Kingdom
42. Markley, J. L., Sax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., Wright, P. E., and Wuthrich, K. (1998) Eur. J. Biochem. 256, 1–15
43. Clarke, N. D., Kissinger, C. R., Desjarlais, J., Gilliland, G. L., and Pabo, C. O. (1994) Protein Sci. 3, 1779–1787
44. Kissinger, C. R., Liu, B. S., Martin-Blanco, E., Kornberg, T. B., and Pabo, C. O. (1990) Cell 63, 579–590
45. Anderson, J. S., Forman, M. D., Modleski, S., Dahlquist, F. W., and Baxter, S. M. (2000) Biochemistry 39, 10045–10054
46. Zhang, X. J., Wozniak, J. A., and Matthews, B. W. (1995) J. Mol. Biol. 250, 527–552
47. Xiao, L., and Honig, B. (1999) J. Mol. Biol. 289, 1435–1444
48. Rejto, P. A., and Freer, S. T. (1996) Prog. Biophys. Mol. Biol. 66, 167–196
49. Meiering, E. M., Serrano, L., and Fersht, A. R. (1992) J. Mol. Biol. 225, 585–589
50. Weiler, S., Gruschat, J. M., Tsao, D. H., Yu, L., Wang, L. H., Nirenberg, M., and Ferretti, J. A. (1998) J. Biol. Chem. 273, 10994–11000
51. Marshall, S. A., Morgan, C. S., and Mayo, S. L. (2002) J. Mol. Biol. 316, 189–199
52. Delbruck, H., Mueller, U., Perl, D., Schmid, F. X., and Heinemann, U. (2001) J. Mol. Biol. 313, 359–369
53. Horovitz, A., Serrano, L., Avron, B., Bycroft, M., and Fersht, A. R. (1990) J. Mol. Biol. 216, 1031–1044
54. Mayor, U. (2003) The Complete Folding Pathway of Engrailed Homeodomain. Ph.D. thesis, University of Cambridge, Cambridge, United Kingdom
55. Li, T. -T. Stark, M. R., Johnson, A. D., and Wolberger, C. (1995) Science 270, 262–269
56. Cox, M., van Tilborg, P. J., de Laut, W., Boelens, R., van Leeuwen, H. C., van der Vliet, P. C., and Kaptein, R. (1990) J. Biol. Chem. 265, 25–32
57. Grant, R. A., Rould, M. A., Klemm, J. D., and Pabo, C. O. (2000) Biochemistry 39, 8187–8192
58. Wilson, M. A., and Brunger, A. T. (2000) J. Mol. Biol. 301, 1237–1256
59. Banerjee-Basu, S., and Baxevanis, A. D. (2001) Nucleic Acids Res. 29, 3258–3269
60. Word, J. M., Lovell, S. C., LaBean, T. H., Taylor, H. C., Zalis, M. E., Presley, B. K., Richardson, J. S., and Richardson, D. C. (1999) J. Mol. Biol. 285, 1711–1733
61. Word, J. M., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (1999) J. Mol. Biol. 285, 1735–1747
62. Richardson, D. C., and Richardson, J. S. (1994) Trends Biochem. Sci 19, 135–138