Gabriela Iurcu-Mustata[1]
Daniel Van Belle[2]
René Wintjens[3]
Martine Prévost[4]
Marianne Rooman[4]

[1] *Department of Biology and Biochemistry, University of Houston, Houston, TX 77204-5513, USA*

[2] *Unit of Bioinformatics, Department of Molecular Biology, Université Libre de Bruxelles, rue des professeurs Jeener et Brachet 12, B-6041 Gosselies, Belgium*

[3] *UMR 8525 CNRS, Institut de Biologie de Lille—Institut Pasteur de Lille, 1 rue du Professeur Calmette, BP 447, 59021 Lille cédex, France*

[4] *Université Libre de Bruxelles, Ingénierie Biomoléculaire—CP 165/64, 50 Avenue F. D. Roosevelt, B-1050 Bruxelles, Belgium*

Role of Salt Bridges in Homeodomains Investigated by Structural Analyses and Molecular Dynamics Simulations

Received 6 October 2000; accepted 13 February 2001

Abstract: *Homeodomains are a class of helix–turn–helix DNA-binding protein motifs that play an important role in the control of cellular development in eukaryotes. They fold in a three α-helix structural module, where the third helix is the recognition helix that fits into the major groove of DNA. Structural analysis of the members of the homeodomain family led to the identification of interactions likely to stabilize the protein domains. Linking the helices pairwise, three salt bridges were found to be well preserved within the family. Also well conserved were two cation–π interactions between aromatic and positively charged side chains. To analyze the structural role of the salt bridges, molecular dynamics simulations (MD) were carried out on the wild-type homeodomain from the* Drosophila *paired protein (1fjl) and on three mutants, which lack one or two salt bridges and mimic natural mutations in other homeodomains. Analysis of the trajectories revealed only small structural rearrangements of the three helices in all MD simulations, thereby suggesting that the salt bridges have no essential stabilizing role at room temperature, but rather might be important for improving thermostability. The latter hypothesis is supported by a good correlation between the melting midpoint temperatures of several homeodomains and the number of salt bridges*

Correspondence to: Marianne Rooman; email: mrooman@ulb.ac.be

Biopolymers, Vol. 59, 145–159 (2001)

and cation–π interactions that connect secondary structures.
Biopolymers 59: 145–159, 2001

Keywords: *homeodomains; DNA-binding proteins; eukaryotes; molecular dynamics; salt bridges; cation–π interactions; paired protein; thermostability*

INTRODUCTION

Homeodomain proteins are transcription factors, which are present in all eukaryotes and play key roles in cellular differentiation during development. Their importance resides in the ability to block or favor the transcription of certain genes in some cell environments at certain stages of development. Their malfunction causes important genetic disorders.[1]

Homeodomains belong to the family of helix–turn–helix (HTH) DNA-binding domains. They typically consist of approximately 60 amino acids and display few amino acid sequence conservation. In spite of this, they present a remarkable similarity in structure. They all consist of three α-helices, preceded by an N-terminal arm (Figure 1). The third helix, called the recognition helix, is positioned roughly perpendicular to the two others and fits into the major groove of DNA. The second and third helices constitute the typical HTH motif, whereas the position of the first helix is specific to the homeodomain subfamily.[2] The N-terminal arm is flexible in the absence of DNA and acquires its structure upon binding. The rest of the structure is practically identical in the free and bound forms. The major structural differences reside in the shorter length of the recognition helix, in some instances, and the increased internal mobility of the second helix.[3–5] Some homeodomains have, however, rather low melting temperatures.[6–8]

The affinity of a protein for its DNA target is influenced by long-range electrostatic interactions that steer the protein into a partially correct orientation for binding.[9] Specificity of DNA binding is achieved through hydrogen-bond interactions between amino acid side chains and DNA bases,[10] and it has recently been shown to be increased by cation–π interactions between charged or partially charged amino acids and the aromatic rings of the bases.[11] The hydration of an interfacial cavity, which functions as a noncovalent extension of the DNA surface, also modulates binding specificity.[12] Specificity is finally increased through the association with other monomers or proteins that induce cooperative DNA binding.[13]

In homeodomains as in other HTH DNA-binding proteins, salt bridges between charged amino acid side chains linking the different helices have been observed. In the nonhomeodomain HTH called 434 repressor, the structural role of a buried salt bridge has been investigated by experimental determination of the structure of the wild type and of a mutant that lacks the salt bridge.[14] This mutation was shown to induce a translation of the first helix relative to the HTH motif and to decrease protein stability. The role of salt bridges was also investigated in homeodomains, but no definite conclusions were reached.[15]

In general, the effect of salt bridges on the thermodynamic stability of proteins is far from being clear. It depends on several factors such as the screening of the charges by the solvent, the cost of desolvating the charged groups to form the salt bridge, and the relative flexibility of the side chains involved in the ion pair.[16] Salt bridges that are inaccessible to the solvent always appear to provide a stabilizing contribution to the folding free energy.[17–20] However, they appear destabilizing compared to hydrophobic bridges.[21,22] For solvent-exposed side chains, there is controversy.[22–26] On the one hand, intrahelical salt bridges in short peptides obtained by de novo design seem to be stabilizing.[23] On the other hand, continuum electrostatic calculations suggest[22] that they are not, because of a larger attraction of the residues for water than for each other; salt bridges could rather help to limit the number of low free energy conformations of a protein or protein complex, due to the large penalty for burying uncompensated ionizable groups. Both views can, however, be reconciled by stating that the stabilizing role of salt bridges is context dependent. It has, for instance, experimentally been shown[26] that a surface side-chain to side-chain salt bridge in a rubredoxin variant is not stabilizing, but that a main-chain to side-chain salt bridge is, due to the lower entropic cost of a salt bridge involving the already immobilized protein backbone.

Furthermore, salt bridges have recently been suggested,[27] on the basis of free energy calculations, to make larger contributions to protein stability at high temperature than at room temperature; therefore they are expected to play a crucial role in promoting the hyperthermostability in proteins. Since then, a number of other studies, both theoretical and experimental, have been supporting this hypothesis.[28–31] In particular, it has been shown that electrostatic interactions seem more favorable in hyperthermophilic proteins than in their mesophilic homologs and that mutations

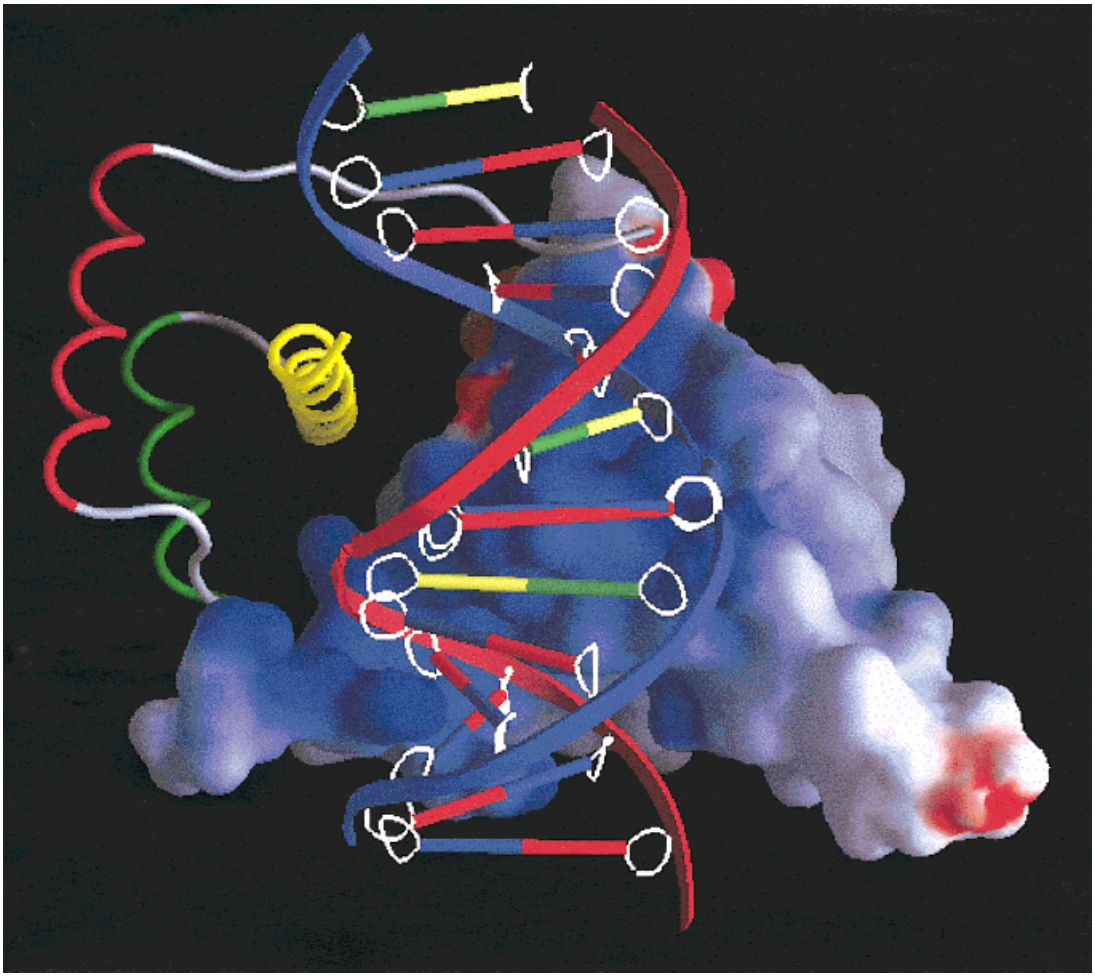

FIGURE 1 Pictorial view of the paired homeodomain dimer (1fjl.pdb). The molecular surface of the first monomer is colored according to the electrostatic potential computed with DELPHI[59] and displayed with GRASP.[60] Color codes for the electric potential are: −10 kT/e (red), 0 kT/e (white), and +25 kT/e (blue). The second monomer is displayed as a ribbon, with the first helix (H_{R-2}) colored in red, the second helix (H_{R-1}) in green, and the third helix (H_R) in yellow.

eliminating ion pairs tend to lower the melting temperature.

In this paper, we performed a structural analysis of homeodomain proteins and identified the interactions conserved within the family. These were found to include several interhelical salt bridges and cation–π interactions between positively charged and aromatic amino acid side chains. With the aim of improving our understanding of the structural or functional role of the conserved salt bridges, molecular dynamics simulations were performed on the paired homeodomain protein[32] and on three mutants that lack some of the salt bridges. The results obtained suggest the weak stabilization properties of these salt bridges at room temperature. However, comparing the melting temperatures of several homeodomain proteins leads to the idea that these salt bridges could be more stabilizing at higher temperatures.

RESULTS

Structural Description of the Paired Homeodomain Protein (PHD)

The paired homeodomain from *Drosophila melanogaster*, noted PHD, is a dimeric DNA-binding protein of the HTH type. Its structure has been solved by x-ray crystallography[32] (PDB[33,34] code: 1fjl) and is depicted in Figure 1. Each monomer contains three helices, no β-strands, and an N-terminal arm that is flexible in absence of DNA. The three helices are

FIGURE 2 Alignment of homeodomain proteins using the classification procedure described in Ref. 2. The PDB[33,34] codes and numbers are given in the leftmost column. With respect to all other family members, the structure for which the sum of the rms deviations is the lowest is given first. The rms deviations (in Å) of all other structures with respect to the first structure are indicated in the rightmost column. The helices, defined as described in Methods, are given in bold letters. The grey boxes indicate αBABBBα turns and the black boxes αGBBα turns. The salt bridges linking different secondary structure elements, calculated using the program HBPLUS,[61] are indicated with full lines above each sequence and the cation–π interactions by dashed lines under each sequence. The amino acids that break the salt bridges, chosen as mutant amino acids in PHD, are indicated by a grey circle. Note that this homeodomain family is slightly different from the one appearing in Ref. 2, because of the inclusion of two new homeodomain proteins, 1ftt and 1bw5. The structures of 1hom, 1hdd, 1fjl, 1apl, 1ftt, 1oct, 1lfb, 1bw5, and 1nk2 have been determined in Refs. 62, 63, 32, 64, 65, 66, 67, 6, and 68, respectively. For the NMR structures, we used the first set of coordinates given in the PDB files.

composed of residues 10–22, 28–38, and 42–63, respectively. The third helix is the recognition helix, referred to as H_R, that enters into the major groove of DNA. The first two helices are denoted as H_{R-2} and H_{R-1}.

These three helices are linked by short turns. The helices H_{R-1} and H_R are connected by an αGBBα turn (see Methods), corresponding to the characteristic turn of the HTH motif, observed in roughly half of the HTH DNA-binding proteins. The helices H_{R-2} and H_{R-1} are connected by an αBABBBα turn. This turn is observed in all homeodomains and nowhere else.[2]

PHD contains two interhelical salt bridges, which link the recognition helix H_R to the two other helices (Figure 2). The first bridge connects Arg52 in H_R and Glu17 in H_{R-2}. The second links Glu42, the first residue of H_R, with Arg31 in H_{R-1}. These two salt bridges are partially buried in the protein core. Arg31 and Glu42 have side-chain solvent accessibilities of 14 and 22%, respectively, and Arg52 and Glu17 of 13 and 33%.

Furthermore, PHD displays two cation–π interactions between positively charged and aromatic amino acid side chains (Figure 2). They involve two consecutive Arg residues in the recognition helix. The first cation–π interaction links Arg52 in H_R, which is already involved in a salt bridge, to Phe20 in H_{R-2}. The second one connects Arg53 with Tyr25, the residue being located in the H_{R-2}–H_{R-1} turn specific to homeodomains.

PHD in the Homeodomain Family

Homeodomains form a subfamily of HTH DNA-binding proteins. The alignment of nine representative

homeodomain structures, displaying between 13 and 52% sequence identity, is shown in Figure 2. It was obtained by classifying a set of HTH domains on the basis of the spatial arrangement of three consecutive α-helices.[2] The *Antennapedia* homeodomain (1hom) appears to be the central fragment of the family, i.e., the fragment whose root mean square (rms) deviation of heavy main chain atoms in the aligned sequence stretches, relative to all other members, is minimum. PHD has an rms deviation of 0.8 Å relative to the *Antennapedia* homeodomain.

In spite of the sequence variability within the family, several structural features are well conserved. The homeodomain-specific αBABBBα turn appears in all 9 members and the HTH-specific αGBBα turn appears in 7 of them. A network of three interhelical salt bridges is also well conserved within the family. Every member except 1bw5 displays at least one of them, and the *Antennapedia* homeodomain (1hom) contains all three. The two salt bridges present in PHD, linking H_{R-2} and H_{R-1} to H_R, are observed in four members. The third salt bridge, between residues 19 and 30 (in PHD numbering), links the first two helices and also appears in four members, but not in PHD. Note that the two residues 19 and 30 are evolutionary correlated, i.e., they are generally conserved or substituted together in the set of known homeodomain sequences.[15] We would like to stress that other (nonhomeodomain) HTH domains also contain interhelical salt bridges, but not located in the same positions within the helices.[14] This observation could indicate that salt bridges conserved within HTH subfamilies, and in particular in homeodomains, play a role in conferring the characteristic orientation of the three key helices.

Cation–π interactions between different secondary structure elements are observed in each of the homeodomains, but are not as well conserved as the salt bridges. The Arg52–Phe20 cation–π interaction of PHD appears in two other members, 1hom and 1apl. Both partners, Arg and Phe, are well conserved within the family and are only occasionally replaced by other positively charged or aromatic amino acids, respectively. Note that Arg52 is in addition involved in a salt bridge with Glu17. The other cation–π interaction of PHD, Arg53–Tyr25, is present in 1hdd and 1hom too. Tyr25 is the middle residue of the αBABBBα turn and this residue is conserved in 6 out of the 9 members. This suggests that a Tyr residue at that position is not only needed for forming the specific homeodomain turn but also the cation–π interaction, especially as Arg53, the Tyr25 partner, is conserved in all 9 family members. These cation–π interactions might thus be structurally important in the same way the interhelical salt bridges are.

The lower levels of conservation of cation–π interactions compared to salt bridges can in part be attributed to the fact that the force fields used to refine experimentally determined protein structures are not completely adapted for optimizing these interactions, and also that not all the structures are determined at a high resolution. This conclusion is supported by the fact that relaxing the geometrical criteria used for their identification leads to detect more of them, and by the high conservation of the cation–π partners among the homeodomains.

Choice of PHD Mutants

The observation that salt bridges are rather well conserved within the homeodomain family led to the idea that mutations breaking one or both characteristic salt bridge(s) might modify the typical spatial arrangement of the homeodomain helices. We chose to break each of the two PHD salt bridges in turn by performing a single amino acid mutation, and then to break both bridges by performing both mutations simultaneously. The choice of the mutations was motivated by the natural mutations occurring within the homeodomain family (Fig. 2). For breaking the H_R–H_{R-2} salt bridge, we mutated Glu17 into Phe, as suggested in the 1lfb domain (Fig. 2). Similarly, we mutated Arg31 into Leu, as in 1apl, to break the H_R–H_{R-1} bridge. We thus do not alter the recognition helix, to avoid modifying residues that are close to the protein–DNA interface.

To further evaluate the proposed mutations, we estimate the stability changes they provoke using a prediction algorithm based on database-derived potentials.[35–37] The predicted changes in folding free energy are equal to 1.3 kcal/mol for the Glu17 $\rightarrow$ Phe mutation and to 0.6 kcal/mol for the Arg31 $\rightarrow$ Leu replacement, with the convention that positive values are destabilizing mutations. The Glu17 $\rightarrow$ Phe mutation is thus predicted to be more destabilizing than Arg31 $\rightarrow$ Leu, at least if the mutant and wild type backbone structures are assumed to be almost identical.[35–37] However, these mutations probably induce structural rearrangements, especially because they are partially buried in the protein core, the side-chain solvent accessibility of the residues Glu17 and Arg31 being equal to 33 and 14%, respectively. These rearrangements are expected because the mutated and mutant amino acids have different sizes. In the case of the replacement Glu $\rightarrow$ Phe, the amino acid size increases from 186 to 222 Å^2, and for the substitution Arg $\rightarrow$ Leu it decreases from 256 to 193 Å^2. Arg $\rightarrow$

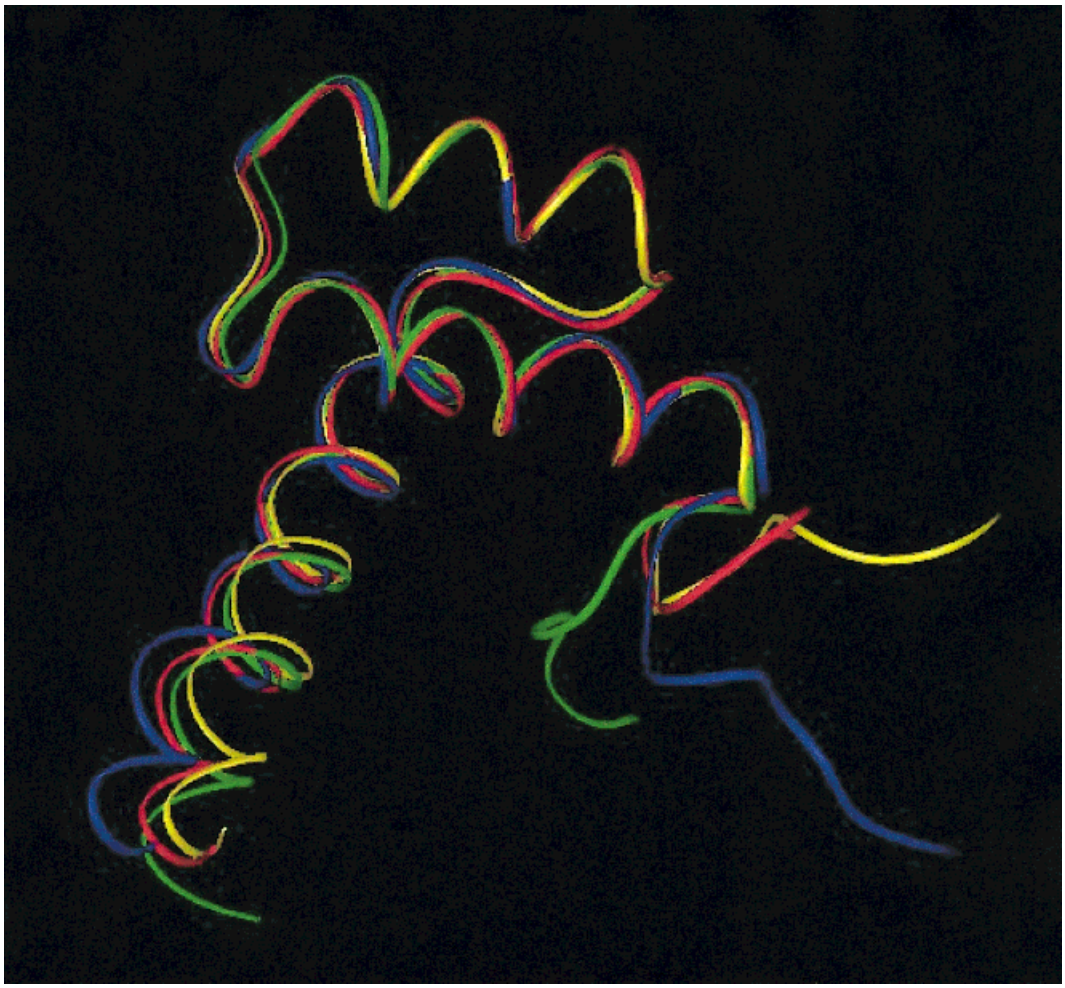

FIGURE 3 Ribbon display of the last representative structures along the 1 ns molecular dynamics simulations: in yellow, the wild-type protein, in red the Arg31 → Leu mutant, in green the Glu17 → Phe, and in blue the Arg31 → Leu/Glu17 → Phe mutant. The H_{R-2} helix is toward the right-hand side of the picture, the H_{R-1} helix is above and the H_R helix is toward the left-hand side.

Leu thus creates a small cavity, while Glu → Phe introduces some stress, which must be absorbed by local structural rearrangements.

Molecular Dynamics Simulations of the Wild-Type and Mutant Proteins

Molecular dynamics simulations were performed at room temperature on the wild-type PHD protein and on three mutants—namely Glu17 → Phe, Arg31 → Leu—and the combined Glu17 → Phe and Arg31 → Leu mutants, which lack one or both salt bridges and mimic natural mutations. The simulation of the wild-type protein started with the crystallographic coordinates of the first monomer of PHD. For the mutant proteins the starting conformations were the last conformation generated after 1.3 ns simulation of the wild-type protein, in which the Leu, Phe, or Leu/Phe side chains replaced the wild-type side chains. Details on the simulations can be found in the Methods section. Each trajectory was analyzed over 1 ns.

Structural Evolution Along the Molecular Dynamics Trajectories

Representative Structures. To simplify some of the analyses performed in this section while highlighting important features, we define 10 representative structures along each of the molecular dynamics trajectories, as described in Methods. The last representative structures of the native and three mutant proteins are shown superimposed in Figure 3. We see that the N-terminal arms adopt different conformations, which is not surprising as they are flexible in absence of

FIGURE 4 Homeodomain proteins with in addition the last representative structures along the molecular dynamics trajectory of wild-type PHD, of the two mutants Glu17 → Phe, Arg31 → Leu and of the double mutant, aligned using the classification procedure described in Ref. 2. The leftmost column contains the PDB[33,34] codes and numbers for the experimental structures; the names of the structures obtained by molecular dynamics simulations are in grey frames. The rms deviations (in Å) of each structure with respect to the first structure are indicated in the rightmost column. The helices are given in bold letters and the turn motifs in italic. The mutant amino acids are depicted by a grey ellipse.

DNA. The three helices and the intervening turns appear very similar in the four structures, with the third helix adopting a slightly different orientation. This difference is largest between the wild-type protein and the double mutant.

Homeodomain Family. To probe the global changes in the PHD mutants, we performed a structural classification of HTH domains similar to the one that led to the homeodomain family given in Figure 2, but with four additional structures in the protein set. These additional structures are the last of the ten representative structures taken along the molecular dynamics trajectory of the wild-type PHD and of the three mutant proteins. The results are given in Figure 4. The crystal structure of PHD becomes the central structure of the family. The wild-type PHD after 1 ns molecular dynamics simulation appears to be quite close to its crystal structure: the three helices and their connecting turns superimpose with it with an rms deviation of 0.5 Å. The two single site PHD mutants, Glu17 → Phe and Arg31 → Leu, both superimpose with an rms deviation of 0.6 Å with respect to the PHD crystal structure and are thus only slightly more different than the wild-type PHD structure after 1 ns simulation. The double mutant, on the contrary, appears further away from the center of the homeodomain family. It is still part of it, but differs more from the central structure PHD than the three typical homeodomains 1hdd, 1lfb, 1hom, and 1apl. It has an rms deviation of 0.9 Å with respect to the crystal PHD structure. This indicates that in the double mutant global structure rearrangements take place, in which the relative orientation of the helices changes.

Turn Motifs. Inspection of the representative structures reveals that the turn between the helices H_{R-2} and H_{R-1}, which is of the type αBABBBα in the wild-type PHD as in all homeodomains, remains of the same type along the whole trajectories of the single and double mutants. The H_{R-1}–H_R turn, which is of the type αGBBα in all HTH-type proteins with a 3-residue loop, also remains of the same type, except in two representatives of the Arg31 → Leu mutant. In the latter, the amino acid at the second position in the turn has (ϕ, ψ, ω) angles outside the considered domains but yet close to the native "B" domain. Hence the typical turn motifs of the homeodomains are not significantly affected upon mutation.

Secondary Structures. The secondary structures are not seriously affected by the mutations either. The typical hydrogen bonding pattern of the helices is one or two residues shorter near the N- or C-terminal helix end in some of the representative structures of the wild-type and mutant proteins, but the (ϕ, ψ, ω) angles always remain in the helical domain. Thus, according to our secondary structure definition, the length of the three helices remain unchanged, and their conformation only slightly deviates from the native one.

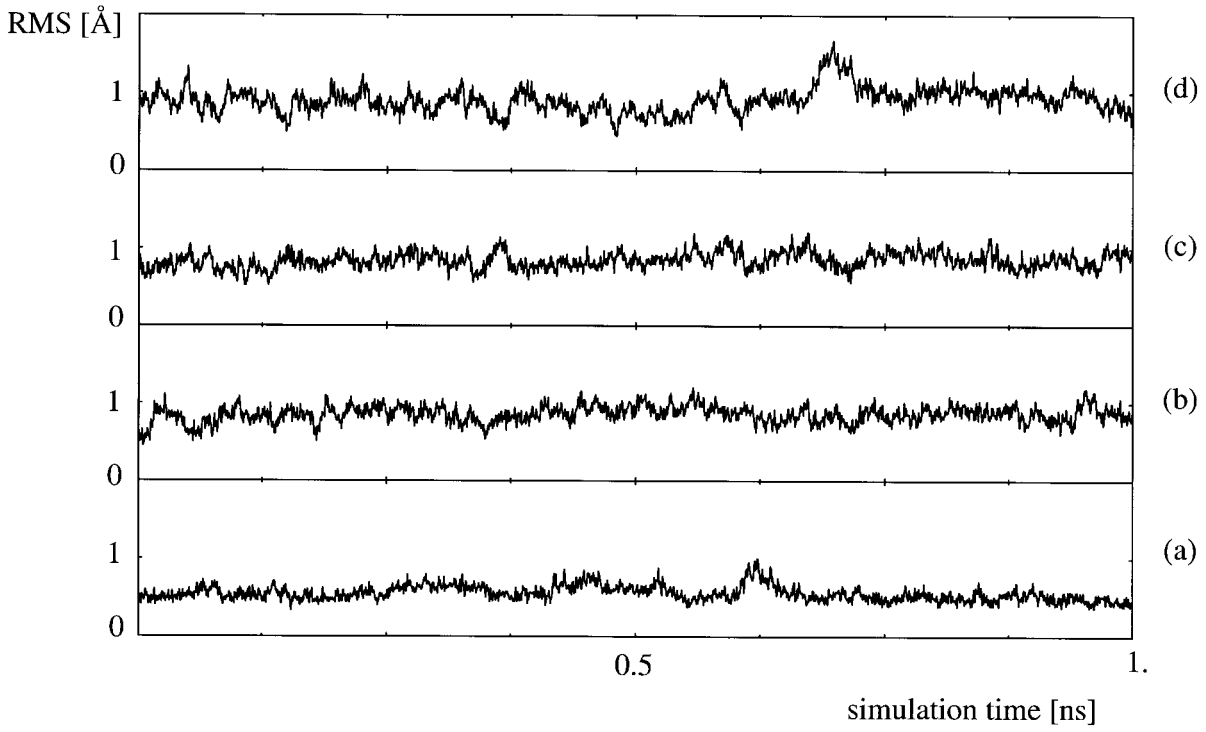

FIGURE 5 The rms deviation of the Cα atoms as a function of the simulation time for (a) the wild type and the three mutants (b) Glu17 → Phe, (c) Arg31 → Leu, and (d) Glu17 → Phe/Arg31 → Leu, computed using the BRUGEL modeling package.[69] Values correspond to snapshot structures separated by 0.2 ps along the 1 ns trajectory. The reference structure was taken as the starting conformation of the native simulation and mutant simulations respectively.

RMS Deviation. To examine the variations in the protein conformations along the 1 ns trajectories, the rms deviations with respect to a reference structure were calculated (Figure 5). The reference structure was taken as the starting conformation of the native simulation and mutant simulations respectively, which correspond to the minimized crystal structure for the former and the structure generated after 1.3 ns of the native trajectory for the latter (see Methods). The rms deviation has a value of about 0.5 Å for the native and Phe mutant and about 1 Å for the Leu and double mutants, at the beginning of the trajectory ($t = 0$). This indicates that movements have occurred during the thermalization and equilibration periods preceding the production run. There was no overall drift observed in the course of the simulations. The average rms values are 0.6 and 0.9 Å for the native and mutant trajectories, respectively, showing that the conformations generated for the mutants depart more from the starting structure than those produced in the native trajectory. The maximum rms value of about 1.5 Å was reached in the double mutant trajectory, which shows larger rms fluctuations.

Helix Orientations. The angles between the main axes of all pairs of helices were computed for all the conformations saved along the four trajectories. Their average values are shown in Table I. The magnitude of the angle between H_{R-2} and H_{R-1} is similar for all four proteins. In contrast, the angle formed by H_R with either H_{R-1} or H_{R-2} is slightly higher in the wild-type protein than in the mutants.

Salt Bridges. The formation of the two salt bridges, linking each of the first two helices with the recogni-

Table I Angle Between the Main Axes of Pairs of Helices[a]

	H_{R-2}–H_{R-1}	H_{R-2}–H_R	H_{R-1}–H_R
Wild type	26° ± 3°	70° ± 2°	85° ± 1°
Glu17 → Phe	27° ± 2°	64° ± 2°	81° ± 1°
Arg31 → Leu	27° ± 2°	64° ± 2°	81° ± 3°
Glu17 → Phe; Arg31 → Leu	27° ± 2°	66° ± 2°	82° ± 2°

[a] The axes are computed as the largest principal axis for the Cα atoms of the three helices.

Table II Percentage Presence of Salt Bridges Along the 1 ns Trajectories of the Wild Type and Mutant Proteins

Salt Bridges	Wild Type	Glu17 → Phe	Arg31 → Leu
Glu17–Arg52	100%	—	2%
Arg31–Glu42	100%	81%	—

tion helix, in the course of the 1 ns trajectories, are given in Table II. The native trajectory features a 100% persistence of both salt bridge interactions, in agreement with the crystal structure. The Glu17 → Phe mutant presents a slightly more transient Arg31–Glu42 salt bridge (81%). An almost complete absence of the Glu17–Arg52 salt bridge (2%) is observed in the Arg31 → Leu mutant trajectory. This salt bridge is also missing in the last conformation of the 1.3 ns wild-type trajectory that served as the starting structure for the modeling of the mutants. It is however formed for 20 ps in the first half of the simulation and then absent again.

Cation–π Interactions. The Tyr25–Arg53 cation–π interaction observed in the PHD crystal structure remains in 6 of the 10 representative structures for the wild-type sequence, the Arg31 → Leu mutant and the Glu17 → Phe mutant, and in 5 of them for the double mutant. The second cation–π interaction, Phe20–Arg52, is somewhat less conserved: 1, 6, 5, and 4 times among the representatives of the wild type, Arg31 → Leu Glu17 → Phe and double mutants, respectively. Interestingly, the Phe20–Arg52 cation–π interaction is replaced in one of the double mutant representatives by the cation–π Phe17–Arg52. This finding suggests that the Glu17–Arg52 salt bridge in the wild type could be replaced by a cation–π interaction Phe17–Arg52 in the mutant proteins.

To analyze this suggestion in more detail, we searched for cation–π interactions not only in the 10 representative structures, but in all 2000 structures along the molecular dynamics simulation trajectory of the Glu17 → Phe and double mutants. We find that the Phe17–Arg52 cation–π interaction replaces the salt bridge in 14 and 133 of these structures, respectively. The propensity of the double mutant to form this interaction is thus much higher than that of the single mutant. This can be taken to indicate that the replacement of the salt bridge by another favorable interaction limits the destabilization of the mutant proteins. It must, however, be noted that the fraction of structures along the simulation trajectory that present this cation–π interaction is relatively small. This could be due to the fact that the CHARMM force fields might not be well suited to account for this type of interaction[11]; this tendency could thus be much more marked than observed here.

Solvent Accessibilities. The solvent accessible surface areas and accessibilities of the side chains were computed in every conformation along the 1 ns trajectories. The average values per residue for the wild type and three mutants are depicted in Figure 6. A first observation is that Leu31 displays a smaller solvent accessibility in the Arg31 → Leu and Glu17 → Phe/Arg31 → Leu mutants (∼ 7%) than Arg31 in the wild type and Glu17 → Phe mutant (∼ 17%), which is reasonable given its hydrophobicity. The accessibility of Glu42 is higher in all mutant simulations, with the highest values in the Arg31 → Leu (56%) and Glu17 → Phe/Arg31 → Leu (52%) mutant trajectories.

The analysis of the Glu17 → Phe and Glu17 → Phe/Arg31 → Leu mutants shows that Phe17 in the mutants has a slightly higher accessibility (∼ 59%) than Glu17 in the wild type (47%). Moreover Arg52, the salt bridge partner of Glu17, also presents a higher accessibility in the mutants (∼ 29%) than in the wild type (18%). This indicates that in absence of the salt bridge, both partners tend to expose a larger fraction of their accessible surface area to the solvent. Note that the Arg31 → Leu trajectory also shows a much higher accessibility of Glu17 and to a lesser extent of its salt-bridge partner Arg52, relative to their value in the wild-type trajectory. This is due to the very weak persistence of the Glu17–Arg52 salt bridge along this trajectory.

Backbone Positional Fluctuations. The simulations were used to analyze the fluctuations that occur in the protein conformation. During the simulation the protein drifts and rotates as a whole. The average structure and fluctuations were therefore computed after each frame in the trajectory was superimposed onto a reference frame by using only rigid body rotations and translations. The positional fluctuations of the backbone atoms computed along the 1 ns trajectory are shown in Figure 7 for the four proteins. Usually the magnitude of the fluctuations permits to identify the secondary regions from the nonorganized ones, the latter featuring higher fluctuations than the former. In all four simulations, however, the fluctuations in H_{R-2} and in particular H_{R-1} display changes as large as those characterizing the loop connecting H_{R-2} and H_{R-1}. This is in agreement with the result that the two turns remain well formed during the simulations, and suggests their structural importance. The first two-thirds of H_R present less abrupt variations but increas-

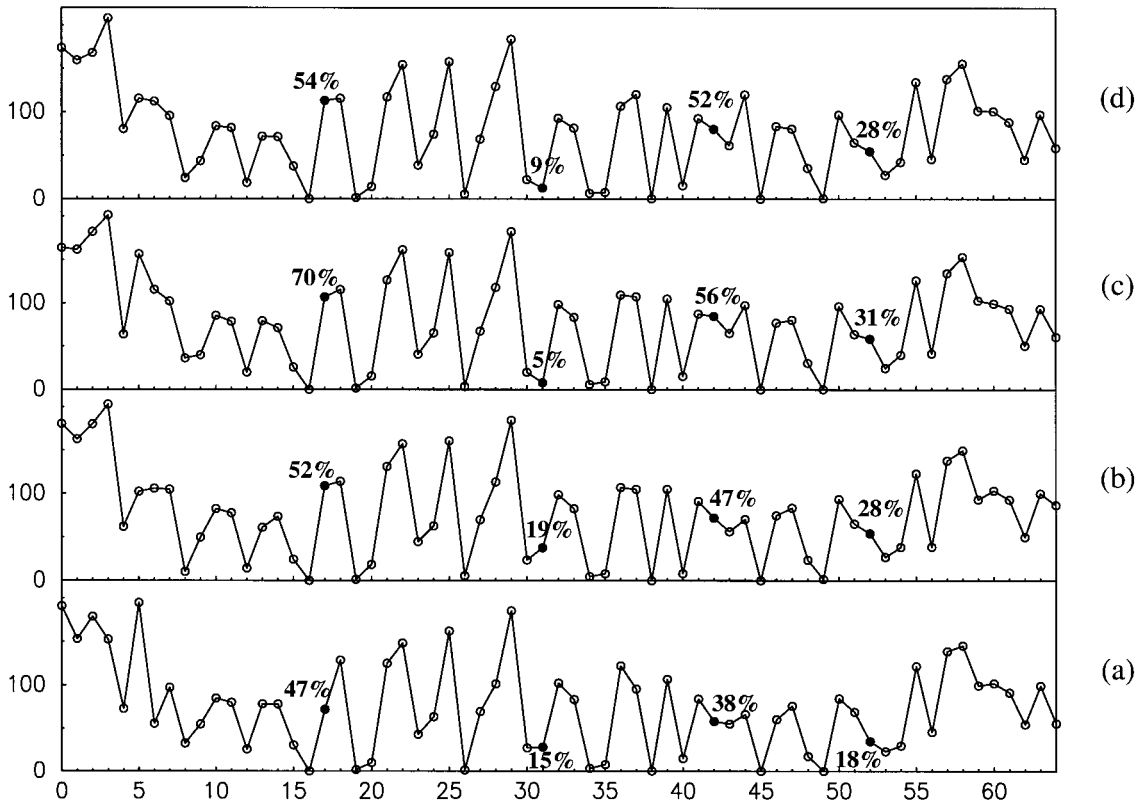

FIGURE 6 Solvent-accessible surface area (in $Å^2$) of the side chains averaged over the 1 ns trajectory as a function of the residue number for (a) the wild type and the three mutants, (b) Glu17 → Phe, (c) Arg31 → Leu, and (d) Glu17 → Phe/Arg31 → Leu. The average solvent accessibility (in %) of the side chains of the wild type and mutant amino acids and their salt bridge partners are explicitly indicated.

ing fluctuations characterize the last third part of H_R. This is related to the fact that this region acquires its structure only upon binding to DNA. The single unlikeness between the four trajectories occurs in the case of the double mutant for which the 20–29 residue portion displays larger fluctuations. Interestingly, this protein fragment is located between the positions of the mutated residues.

Side-Chain Mobility. The computed average fluctuation of the side chain torsion angles χ_1 and χ_2, which are the most constrained angles due to their nearness to the main chain provides information on the degree of side-chain flexibility. Fluctuations in χ_1 and χ_2 higher than 25° are frequently observed, but do not appear to differ significantly among the wild-type and mutant proteins (data not shown).

DISCUSSION

The structural analysis of the homeodomain family revealed the conservation of three interhelical salt bridges, linking the helices pairwise, thereby suggesting that these interactions could be important for structure or stability. Analysis of molecular dynamics simulations at room temperature on the wild-type PHD, presenting two salt bridges, and on three mutants, breaking one or both salt bridges, throws some doubt on this suggestion. Indeed, we observed only slight structural modifications in the mutant proteins along the 1 ns trajectory that can be summarized as a slight change in the orientation of the last helix H_R. The double mutant, with no salt bridges left, deviates only slightly more than the single mutants. No significant increase in the backbone or side chain fluctuations were observed either. In the Arg31 → Leu mutant Glu17 and Arg52 do not form a salt bridge in the initial structure of the simulations, and this bridge only occasionally forms again along the trajectory. In spite of this, the structure does not seem to be significantly more flexible or to deviate more from the starting structure.

These findings are in agreement with prior conclusions that partially accessible salt bridges usually give little contribution to thermodynamic stability.[16–26]

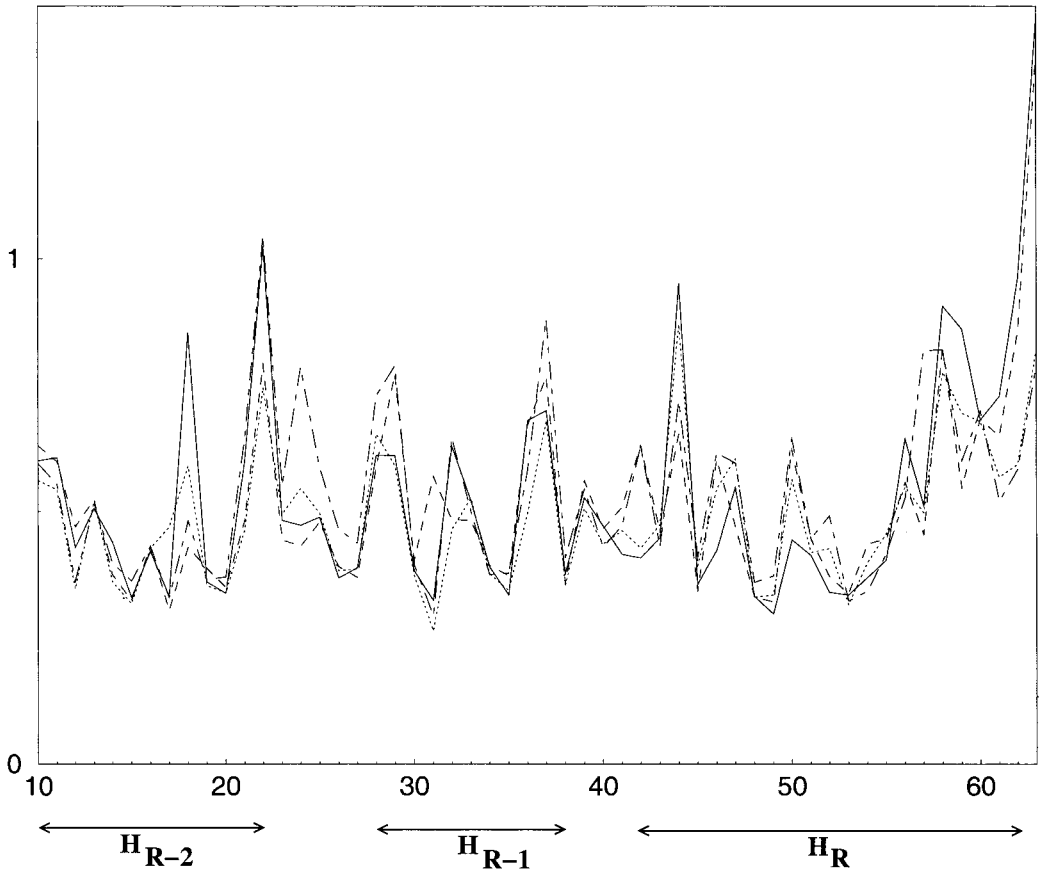

FIGURE 7 Backbone positional fluctuations of the protein structure around the average conformation computed from the 1 ns trajectory of the wild-type (solid line), Glu17 → Phe (dashed line), Arg31 → Leu (dotted line) and Glu17 → Phe/Arg31 → Leu (dot-dashed line), computed using the BRUGEL modeling package.[69] The bars at the bottom indicate the three helices of the protein.

The two salt bridges are only partially buried in the wild-type crystal structure with the solvent accessibility of the four partners varying between 13 and 33%. That residues Glu42 and Arg52 tend to increase their solvent accessibility during the simulations, when they loose their salt bridge partner, indicates that they compensate for the loss of the salt bridge interaction by an increase of their solvation.

Note, however, that the tested substitutions mimic natural mutations and are hence probably not the most disrupting ones. In particular, the Glu17–Arg52 salt bridge is occasionally replaced by another favorable interaction along the trajectories of the Glu17 → Phe mutants, i.e., a Phe17–Arg52 cation–π interaction. This could also contribute to limit the loss of stability upon mutation.

If not for stability, then why are salt bridges so well conserved in the homeodomain family? A first explanation could be that they fix the exact structural arrangement of the three helices, and that even small departures from this optimal arrangement could affect the DNA-binding affinity or specificity. Furthermore, electrostatics are very important for DNA-binding proteins. Positive charges distributed on one side of the protein and negative charges on the other side promote the correct orientation for protein–DNA interaction.[9] Besides this aspecific charge distribution, the specific localization of positive charges near the protein–DNA interface leads to the formation of cation–π interactions between positively charged amino acid side chains and DNA bases.[11] Moreover, the presence of ions in the solvent or condensed along the DNA backbone also modulates the protein–DNA electrostatics.[38] In this context, even the localization of partially buried charges, either neutralized by a salt bridge or not, might be important for protein–DNA interactions.

A different reason for the salt bridge conservation could be thermostability. Although salt bridges generally make little contribution to protein stability at room temperature, they seem to make larger contributions at higher temperature.[26–31] To investigate this

Table III Experimental Melting Midpoint Temperatures T_m for Homeodomains, Obtained in the Papers Given in the Rightmost Column

	T_m (°C)	$N_{salt\text{-}bridges}$ [f]	$N_{cation\text{-}\pi}$ [f]	References
1hom	48	3	3	32
1ftt	39–42[c]	1	2	33, 34
1nk2 *H52R T56W*[a]	33	2	0–2[g]	8
1nk2 *H52R*[a]	29	2	0–1[g]	8
1ftz [1hom][b]	27–32[d]	3	1	32
1nk2 *T56W*[a]	26	1	0–1[g]	8
1nk2	25	1	0	8
1hdp [1oct][b]	25	0	1	7
1bw5	22–30[e]	1	0	6

[a] Single and double mutants of 1nk2.

[b] 1ftz and 1hdp present a high degree of sequence identity with 1hom and 1oct, respectively, and are therefore not among the homeodomain members of Figure 2.

[c] T_m = 39°C in Ref. 40 and 42°C in Ref. 41.

[d] T_m = 32°C at pH 6.1 and 27°C at pH 4.8.

[e] The two temperatures are indicative of a two-stage melting behavior.

[f] Number of salt bridges or cation–π interactions between different secondary structure elements.

[g] As we do not have access to the mutant structures, we are unable to determine if the cation–π interactions, rendered possible by the mutations, are really formed. We do not have this uncertainty with the additional salt bridge, because it is described in Ref. 8.

possibility, we listed all melting midpoint temperatures T_m of homeodomains described in the literature and searched for a correlation between these values and the number of salt bridge and cation–π interactions (Table III). Strikingly, the *Antennapedia* homeodomain (1hom), with all three typical homeodomain salt bridges plus three cation–π interactions, has the highest T_m of all homeodomains: 48°C.[39] In contrast, the isl-1 homeodomain (1bw5) has the lowest T_m value (22°C[6]), with only one salt bridge and no cation–π interaction. Furthermore, the vnd/nk-2 homeodomain (1nk2) has a low T_m value of 25°C[8] and only one salt bridge, but the mutant His52 → Arg, which allows the additional salt bridge Glu17–Arg52 and perhaps also the cation–π interaction Phe20–Arg52, has a higher T_m of 29°C.[8] The other mutant, Thr56 → Trp, has also a slightly higher T_m (26°C[8]). Comparing its sequence to that of *Antennapedia* homeodomain (1hom) suggests that the introduction of a Trp could allow a cation–π interaction with Arg24 (Figure 2). Finally, the double vnd/nk-2 mutant, which could contain both interactions, has a still higher T_m of 33°C.[8]

These results suggest that the number of salt bridge and cation–π interactions between different secondary structure elements can play an important role in determining thermostability. In previous works,[6,8] the thermostability was related to the lengthening of the third helix, H_R, via appropriate amino acids at positions 52 and 56. This is actually in accordance with our findings, given that these positions are crucial for cation–π and salt bridge interactions. Further analyses are necessary to verify and specify these propositions.

METHODS

Molecular Dynamics Simulations

Molecular dynamics simulations were performed using the CHARMM program.[42] The crystallographic coordinates of monomer A of 1fjl, found in the Protein Data Bank,[33,34] were taken as the starting structure for the simulations of the PHD wild-type protein. This monomer was chosen because it is complete (residues A0–A64). For the mutant proteins, their starting conformation was the last conformation generated after 1.3 ns of MD simulation of the wild-type protein. The wild-type amino acid side chains were replaced in the mutant simulations by Leu, Phe, or Leu/Phe side chains. The protein and solvent interacted via the CHARMM 22 force field. All protein atoms were explicitly represented[43] and the water was represented by the TIP3P model.[44] Bonds connecting hydrogens were constrained using the SHAKE algorithm,[45] which permitted the use of an integration time step of 1 fs. The simulation consisted of a system defined by 1116 protein atoms, 13 crystallographic water molecules, and 3547 generated water molecules in a periodic volume with dimensions 52.8 × 45.5 × 49.7 Å.

To neutralize the system, seven water molecules were replaced by chloride ions. The Cl^- counterions replaced water molecule in locations corresponding to a positive electrostatic potential computed for the monomer. Nonbonded interactions were smoothly truncated at 8.5 Å with a shifting function for the electrostatic interactions and a

switching function for van der Waals interactions, the latter being applied between 7.5 and 8.5 Å. The truncation scheme applied to the calculation of electrostatic interactions was calibrated against the Ewald summation method in simulations of pure liquid water and was shown to perform well with respect to both structural and thermodynamic properties.[46] Moreover, previous simulations indicated that it properly describes the balance of protein–protein and protein–solvent interactions.[47] The observation that the conformations sampled in the wild-type protein trajectory depart little from the crystal structure, further supports the validity of our truncation scheme. However, recent studies have suggested that the Ewald summation method is a better protocol, at least for nucleic acids.[48] It has nevertheless also been shown[49–50] to present a number of artifacts, in particular when solute cavities are large compared to the unit cell, a feature characterizing our system. In our opinion, it is hard to figure out if the use of longer cut-off distances combined with the minimum image convention would improve (or spoil) the behavior of the system, as better results are not necessarily achieved by applying longer cutoffs.[51]

The simulations were carried out with the following protocol. The protein was fixed and the water molecules, chloride ions, and side chains of the mutated residues were subjected to 200 steps of steepest descent energy minimization. The water was then heated and equilibrated for 10 ps. In a next step the solvent was fixed, the constraints on the protein were removed and the protein was subjected to 5000 steps of dynamics. Then all constraints were removed and the whole system was brought up to a final temperature of 298 K for 15 ps and equilibrated for 20 ps. The subsequent production phase was performed in the microcanonical ensemble for 1 ns for the mutants and 1.3 ns for the wild type. The temperature calculated from the average kinetic energy measured during the production phase amounts to 298 K for the wild-type and three mutant trajectories.

Analysis of Generated Structures

Representative Structures Along the Molecular Dynamics Trajectory. To allow easy analysis of the structural modifications that take place during the molecular dynamics simulation, we identified 10 representative structures along the trajectory. We therefore divide the trajectory into 10 subsets, each containing an equal number of consecutively generated structures. In each subset, we compute for all the pairs of structures it contains the rms deviation of superimposed N, Cα, C, and O backbone atoms, using U3BEST.[52] The representative structure of the subset is the structure for which the average rms deviation with respect to all other structures of the subset is minimum.

Local Structure Definitions. We consider 5 domains of backbone dihedral angles (φ, ψ, ω), noted A, B, G, E, and O.[53,54] Four of them, A, B, G, and E, correspond to *trans* conformations ($\omega \approx 180°$): A groups helical structures, B corresponds to extended conformations, G and E have negative φ angles, mirror-symmetrical to A and B, respectively. The last domain, O, corresponds to *cis* conformations ($\omega \approx 0°$).

Turn motifs are denoted by the (φ, ψ, ω) angle domains of the residues in the turn, flanked by α or β according to whether the flanking secondary structure elements are helices or β-strands, with the conventions of Refs. 2, 54, and 55.

The definition of secondary structures uses the DSSP[56] assignments, which are essentially based on H-bonding patterns, and lengthens the helices as long as the flanking residues are in the helical (φ, ψ, ω) domain A. This definition is borrowed from Ref. 54 and is consistent with the above definition of turn motifs.

Cation–π Interactions. The cation–π interactions between aromatic and charged or partially charged amino acid side chains are defined geometrically by the distance and angle criteria described in Ref. 11. The distance criterion requires that the atom carrying the positive charge is located at 4.5 Å at most from one of the atoms of the aromatic cycle. The angle criterion ensures that it is situated above that cycle, and more precisely inside a cylinder having as basis a disk containing the cycle and of a radius equal to twice the cycle radius, and of height 4.5 Å.

Solvent Accessibility. The solvent-accessible surface area of an amino acid side chain in its parent structure is computed using the program SURVOL.[57] Its solvent accessibility is computed as the percentage of its solvent-accessible surface area relative to the solvent accessible surface area that it would have when included in an extended Gly–X–Gly tripeptide.

Automatic Structural Classification of Protein Domains. We use the structural classification algorithm of a set of protein domains described in Refs. 2 and 58. The classification is performed on the spatial arrangement of α-helices, irrespective of the length and conformation of the intervening loops and the possible presence of β-structures. The similarity of the arrangements is estimated by a structural alignment procedure that uses the rms deviation of superimposed N, Cα, C, and O backbone atoms as a similarity measure.

GIM acknowledges support from the Reyers Fund during the first stages of this work. MP and MR are research associate and senior research associate, respectively, of the Belgian National Fund for Scientific Research (FNRS).

REFERENCES

1. Lewis, E. B. Nature 1978, 276, 565–570.
2. Wintjens, R.; Rooman, M. J Mol Biol 1996, 262, 294–313.

3. Tsao, D. H.; Gruschus J. M.; Wang, L. H.; Nirenberg, M.; Ferretti, J. A. Biochemistry 1994, 33, 15053–15060.
4. Cox, M.; van Tilborg, P. J.; de Laat, W.; Boelens, R.; van Leeuwen, H. C.; van der Vliet, P. C.; Kaptein R. J Biomol NMR 1995, 6, 23–32.
5. Carra, J. H.; Privalov, P. L. Biochemistry 1997, 36, 526–533.
6. Ippel, H.; Larsson, G.; Behravan, G.; Zdunek, J.; Lundqvist, M.; Schleucher, J.; Lycksell, P. O.; Wijmenga, S. J Mol Biol 1999, 288, 689–703.
7. Sivaraja, M.; Botfield, M. C.; Mueller, M.; Jancso, A.; Weiss, M. A. Biochemistry 1994, 33, 9845–9855.
8. Weiler, S.; Gruschus, J. M.; Tsao, D. H.; Yu, L.; Wang, L. H.; Nirenberg, M.; Ferretti, J. A. J Biol Chem 1998, 273, 10994–11000.
9. Fogolari, F.; Elcock, A. H.; Esposito, G.; Viglino, P.; Briggs, J. M.; McCammon, J. A. J Mol Biol 1997, 267, 368–381.
10. Seeman, N. C.; Rosenberg, J. M.; Rich, A. Proc Natl Acad Sci USA 1976, 73, 804–808.
11. Wintjens, R.; Liévin, J.; Rooman, M.; Buisine, E. J Mol Biol 2000, 302, 393–408.
12. Labeots, L. A.; Weiss, M. A. J Mol Biol 1997, 269, 113–128.
13. Chariot, A.; Gielen, J.; Merville, M. P.; Bours, V. Bioessays 1999, 21, 267–270.
14. Pervushin, K.; Billeter, M.; Siegal, G.; Wuthrich, K. J Mol Biol 1996, 264, 1002–1012.
15. Clarke, N. D. Protein Sci 1995, 4, 2269–2278.
16. Yang, A.; Honig, B. Curr Opin Struct Biol 1992, 2, 40–45.
17. Fersht, A. R. J Mol Biol 1972, 64, 497–509.
18. Anderson, D. E.; Becktel, W. J.; Dahlquist, F. W. Biochemistry 1990, 29, 2403–2408.
19. Marqusee, S.; Sauer, R. T. Protein Sci 1994, 3, 2217–2225.
20. Albeck, S.; Unger, R.; Schreiber, G. J Mol Biol 2000, 298, 503–520.
21. Waldburger, C. D.; Schildbach, J. F.; Sauer, R. T. Nature Struct Biol 1995, 2, 122–128.
22. Hendsch, Z. S.; Tidor, B. Protein Sci 1994, 3, 211–216.
23. Marqusee, S.; Baldwin, R. L. Proc Natl Acad Sci USA 1987, 84, 8898–8902.
24. Sun, D. P.; Sauer, U.; Nicholson, H.; Matthews, B. W. Biochemistry 1991, 30, 7142–7153.
25. Serrano, L.; Horovitz, A.; Avron, B.; Bycroft, M.; Fersht, A. R. Biochemistry 1990, 29, 9343–9352.
26. Strop, P.; Mayo, S. L. Biochemistry 2000, 39, 1251–1255.
27. Elcock, A. H. J Mol Biol 1998, 284, 489–502.
28. Kumar, S.; Ma, B.; Tsai, C.-J.; Nussinov, R. Proteins 2000, 38, 368–383.
29. Xiao, L.; Honig, B. J Mol Biol 1999, 289, 1435–1444.
30. Frankenberg, N.; Welker, C.; Jaenicke, R. FEBS Lett 1999, 454, 299–302.
31. Merz, A.; Knochel, T.; Jansonius, J. N.; Kirchner, K. J Mol Biol 1999, 288, 753–763.
32. Wilson, D. S.; Guenther, B.; Desplan, C.; Kuriyan, J. Cell 1995, 82, 709–719.
33. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanovich, T.; Tasumi, M. J Mol Biol 1977, 112, 535–542.
34. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. Nucleic Acids Res 2000, 28, 235–242.
35. Gilis, D.; Rooman, M. J Mol Biol 1996, 257, 1112–1126.
36. Gilis, D.; Rooman, M. J Mol Biol 1997, 272, 276–290.
37. Gilis, D.; Rooman, M. Theor Chem Acc 1999, 101, 46–50.
38. Marky, N. L.; Manning, G. S. J Am Chem Soc 2000, 122, 6057–6066.
39. Qian, Y. Q.; Furukubo-Tokunaga, K.; Resendez-Perez, D.; Muller, M.; Gehring, W. J.; Wuthrich, K. J Mol Biol 1994, 238, 333–345.
40. Tell, G.; Acquaviva, R.; Formisano, S.; Fogolari, F.; Pucillo, C.; Damante, G. Int J Biochem Cell Biol 1999, 31, 1339–1353.
41. Damante, G.; Tell, G.; Leonardi, A.; Fogolari, F.; Bortolotti, N.; Dilauro, R.; Formisano, S. FEBS Lett 1994, 354, 293–296.
42. Brooks, B. R.; Bruccoleri, R. E.; States, B. D. O. D. J.; Swaminathan, S.; Karplus, M. J Comput Chem 1982, 4, 187–217.
43. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. J Phys Chem 1998, 102, 3586–3616.
44. Jorgensen, W. L.; Chandresekhar, J.; Madura, J.; Impey, R.; Klein, M. J Chem Phys 1983, 79, 926–935.
45. Ryckaert, J. P.; Cicotti, G.; Berendsen, H. J. C. J Comp Phys 1977, 23, 327–341.
46. Prévost, M.; Van Belle, D.; Lippens, G.; Wodak, S. Mol Phys 1990, 71, 587–603.
47. Prévost, M. Fold Des 1998, 3, 345–351.
48. Sagui, C.; Darden, T. A. Annu Rev Biophys Biomol Struct 1999, 28, 155–179.
49. Hunenberger, P. H.; McCammon, J. A. J Chem Phys 1999, 110, 1856–1872.
50. Hunenberger, P. H.; McCammon, J. A. Biophys Chem 1999, 78, 69–88.
51. Schreiber, H..; Steinhauser, O. Biochemistry 1992, 31, 5856–5860.
52. Kabsch, W. Acta Crystall Sect A 1978, 34, 827–828.
53. Rooman, M. J.; Kocher, J.-P. A.; Wodak, S. J. J Mol Biol 1991, 221, 961–979.
54. Wintjens, R.; Rooman, M.; Wodak, S. J Mol Biol 1996, 255, 235–253.

55. Wintjens, R.; Rooman, M.; Wodak, S. Israel J Chem 1994, 34, 257–269.
56. Kabsch, W.; Sander, C. Biopolymers 1983, 22, 2577–2637.
57. Alard, P. Ph.D. thesis, Université Libre de Bruxelles, 1991.
58. Boutonnet, N.; Kajava, A.; Rooman, M. Proteins 1998, 30, 193–212.
59. Honig, B.; Nicholls, A. Science 1995, 268, 1144–1149.
60. Nicholls, A.; Sharp, K.; Honig, B. Proteins 1991, 11, 281–296.
61. McDonald, I. K.; Thornton, J. M. J Mol Biol 1994, 238, 777–793.
62. Billeter, M.; Qian, Y. Q.; Otting, G.; Muller, M.; Gehring, W.; Wuthrich, K. J Mol Biol 1993, 234, 1084–1093.
63. Kissinger, C. R.; Liu, B.; Martin-Blanco, E.; Kornberg, T. B.; Pabo, C. O. Cell 1990, 63, 579–590.
64. Wolberger, C.; Vershon, A. K.; Liu, B.; Johnson, A. D.; Pabo, C. O. Cell 1991, 67, 517–528.
65. Viglino, P.; Fogolari, F.; Formisano, S.; Bortolotti, N.; Damante, G.; Di Lauro, R.; Esposito, G. FEBS Lett 1993, 336, 397–402.
66. Klemm, J. D.; Rould, M. A.; Aurora, R.; Herr, W.; Pabo, C. O. Cell 1994, 77, 21–32.
67. Ceska, T. A.; Lamers, M.; Monaci, P.; Nicosia, A.; Cortese, R.; Suck, D. EMBO J 1993, 12, 1805–1810.
68. Gruschus, J. M.; Tsao, D. H.; Wang, L. H.; Nirenberg, M.; Ferretti, J. A. Biochemistry 1997, 36, 5372–5380.
69. Delhaise, P.; Van Belle, D.; Bardiaux, M.; Alard, P.; Hamers, P.; Van Cutsem, E.; Wodak, S. J. J Mol Graph 1985, 3, 116–119.