The Impact of R53C Mutation on the Three-Dimensional Structure, Stability, and DNA-Binding Properties of the Human Hesx-1 Homeodomain

Dedicated to Prof. M. Rico on the occasion of his 65th birthday

Isabel de la Mata,[b] Jose L. Garcia,[c] Carlos González,[d] Margarita Menéndez,[e] Javier Cañada,[c] Jesús Jiménez-Barbero,[c] and Juan Luis Asensio*[a]

Septo-optic dysplasia (SOD) is loosely defined as any combination of optic nerve hypoplasia, pituitary gland hypoplasia, and midline abnormalities of the brain. Recent studies have shown that this rare disease has its origin in key mutations in Hesx-1, a protein that plays a critical role in normal development of the forebrain, eyes, and other anterior structures during embryogenesis. R53C mutation in the Hesx-1 homeodomain has recently been identified in some patients with SOD. However, no detailed description of the effect of this mutation on the protein structure, stability, and function has been reported so far. The impact of R53C substitution on Hesx-1 homeodomain structure, stability, and DNA-binding properties was analyzed by using a combination of NMR spectroscopy, molecular modeling and circular dichroism experiments. Although R53C mutation has very slight effects on protein structure, it has a profound impact on homeodomain stability. We show that intramolecular disulfide formation can be easily accomplished in the mutated homeodomain, which suggests that such bond formation could take place in vivo. This modification has a key effect on the homeodomain structure and stability, mainly through its effect on helix I/helix III packing. Finally, CD titrations allowed us to establish the energy cost of the R53C substitution with respect to Hesx-1-homeodomain – DNA complex stability. Detailed structural models are provided for the wild-type human Hesx-1 homeodomain and the R53C mutant, which is associated with SOD in humans. The reported effect of R53C mutation on protein stability and DNA-binding properties, together with the significant structural perturbations induced by disulfide formation in the mutated polypeptide might explain the loss of activity of the mutant in vivo.

KEYWORDS:

circular dichroism · DNA recognition · molecular modeling · NMR spectroscopy · protein models

Introduction

Homeobox genes code for a large superfamily of regulatory proteins. They all contain a 60-residue structural motif (called the homeodomain) responsible for sequence-specific DNA binding activity. Homeodomains have been shown to play a particularly critical role in developmental patterning and differentiation and, in some cases, homeodomain-containing DNA-binding proteins are involved in oncogenesis in humans. In fact, several types of deformities and cancers have been attributed to defects in homeoproteins.[1–4]

A new homeobox gene named Hesx-1 (also known as Rpx) that is apparently involved in the determination of the pituitary gland has recently been discovered in mice.[5] Hesx-1 has been classified as a member of a new class of homeobox genes, namely the *Anf* class,[6] which encodes a homeodomain consensus sharply different from those of other classes. The main feature of these proteins that permitted all *Anf* genes to be assigned to a distinct class is the much higher degree of identity revealed when their homeodomains are compared with each other (more than 75%) than with any homeodomain of other known classes (less than 55%). Moreover, the amino acid sequences of those regions presumably involved in DNA

[a] *Dr. J. L. Asensio*
Instituto de Química Orgánica General
CSIC, Madrid 28006 (Spain)
Fax: (+34) 91-5644-853
E-mail: iqoa110@iqog.csic.es

[b] *Dr. I. de la Mata*
Departamento de Bioquímica de la Universidad Complutense, Madrid (Spain)

[c] *Prof. Dr. J. L. Garcia, Dr. J. Cañada, Prof. Dr. J. Jiménez-Barbero*
Centro de Investigaciones Biológicas
CSIC, Madrid (Spain)

[d] *Dr. C. González*
Instituto de Estructura de la Materia
CSIC, Madrid (Spain)

[e] *Dr. M. Menéndez*
Instituto Química Física Rocasolano
CSIC, Madrid (Spain)

Supporting information for this article is available on the WWW under http://www.chembiochem.org or from the author.

  1439-4227/02/03/08 $ 20.00+.50/0 ChemBioChem **2002**, 3, 726 – 740

recognition, that is, those that determine the functional specificity of the homeodomain, are conserved to a greater extent within the *Anf* class than between classes and show clear differences from analogous sequences of other homeodomains.

Recent studies have shown that Hesx-1 plays a critical role in normal development of the forebrain, eyes, and other anterior structures such as the pituitary gland during embryogenesis.[4] Thus, mutations of Hesx-1 that affect its DNA-binding ability have been associated with septo-optic dysplasia (SOD) in humans and mice.[4] Septo-optic dysplasia is extremely heterogeneous and is loosely defined as any combination of optic nerve hypoplasia, pituitary gland hypoplasia, and midline abnormalities of the brain.[7–9] R53C mutation in the Hesx-1 homeodomain has been recently identified in some patients with SOD.[4] This nonconservative substitution affects the DNA binding capacity of the protein. In fact, the importance of R53 for homeodomain – DNA interaction has been highlighted by several studies in the last few years. R53 makes contact with a phosphate group in both paired and engrailed protein – DNA complexes according to X-ray data[10, 11] and this residue is conserved in 99% of the homeodomain sequences. Furthermore, different substitutions of R53 in other homeoproteins have been described to have decreased DNA affinity that ranges from a 20-fold reduction in binding activity for R53H substitution in the *Drosophila proboscipedia* protein[12] to a complete suppression of activity for the R53G mutation in PAX3 (associated with Waardenburg syndrome in humans).[13] Clearly, these different effects may reflect not only the loss of a single arginine – phosphate interaction in the corresponding protein – DNA complexes of the mutated polypeptides, but also the impact of each substitution on the protein structure and stability. A complete understanding of the effect of mutations at the key position 53 on DNA-binding activity therefore requires an individual and complete analysis of the structural and thermodynamic effects associated with these mutations. This particularly applies to the R53C substitution as introduction of a cystein residue in the polypeptide chain could, in some cases, lead to significant structural changes through intra- or intermolecular disulfide bond formation. Here we present a detailed analysis of the impact of the single mutation R53C, associated with septo-optic dysplasia, on the structure, stability, and DNA-binding properties of the human Hesx-1 DNA binding domain. NMR spectroscopy, circular dichroism, and molecular modeling studies were used and the results obtained provide new insights into the different factors that affect the stability of the homeodomain scaffold, as well as the structural requirements for DNA recognition in this important DNA binding motif.

Results and Discussion

NMR solution structure of the wild-type Hesx-1 homeodomain

Initially, many backbone and side-chain resonances of the Hesx-1 homeodomain were assigned by using well-established protocols for NOESY, TOCSY, and DQF-COSY experiments at 600 and 800 MHz. However, heteronuclear experiments were carried out on a ^{15}N-labeled sample as well because of resonance overlap in some regions of the homonuclear spectra. 3D NOESY-HSQC experiments yielded additional information on sequential connectivities that could not be established from the analysis of 2D spectra. Sequence-specific assignments for the ^{1}H and ^{15}N resonances of the wild-type Hesx-1 DNA binding domain are shown in Table S1 in the Supporting Information. The N-terminal arm (residues 1 – 7) is disordered in solution, as indicated by the absence of nonlocal NOEs. Three helical regions that span residues 10 – 22, 28 – 38, and 42 – 56 were identified from the analysis of sequential and medium-range NOEs, as expected for a homeodomain structure. Figure 1 shows an ^{15}N HSQC spectrum of the wild-type Hesx-1 homeodomain and the secondary structure elements identified by the analysis of NOE data. Two different components can be distinguished in the C-terminal region. Residues 56 – 60, although disordered to some extent, exhibit a nonrandom structure. Weak NOEs connecting positions $i/i+3$ and $i/i+4$ were observed in this region, which suggests the existence of a low-population helical structure. In contrast, residues 60 – 67 are disordered in solution.

In a second step, a set of 607 NOEs (359 long-range) were unambiguously assigned and converted into distance constraints as described in the Experimental Section. A group of 30 structures was obtained (see Table 1) from 300 randomized conformations by using the DYANA program.[14] These 30 structures were the best in terms of the target function and were further refined by using the AMBER 5.0 package[15] with the force field described by Cornell et al.[16] Explicit water molecules were included in these calculations to prevent unrealistic interactions between disordered regions of the protein and the structured helical core. The backbone of the calculated structures is extremely well-ordered between residues 8 and 53 (rmsd = 0.62 Å), while the convergence is small between residues 1 – 8 and the C-terminal fragment (56 – 67). The structures have very small deviations from ideal geometry and reasonable nonbonded contacts. A schematic illustration of the 30 simulated annealed structures is depicted in Figure 2a. These structures fully explain all the unusual chemical shifts observed in the NMR spectra. The methyl group of L16 appears upfield-shifted to -0.54 ppm. The NMR models show that this methyl moiety is located on top of the aromatic ring of W48. In a similar way, the aromatic protons Hε3, Hζ3, and Hη2 of W48 are upfield shifted, with chemical shifts below 7.0 ppm. It can be observed (Figure S1 in the Supporting Information) that these protons are involved in a T-type W48 – F49 interaction. These three residues, L16, W48, and F49, are conserved in most homeodomain sequences. The global structure is very similar to that reported for other homeoproteins.[17] This structure contains a disordered N-terminal arm (residues 1 – 7), three helical regions, named helix I (residues 10 – 22), II (28 – 38), and III (42 – 56, also known as the recognition helix), and a disordered C-terminal region (61 – 67). Secondary and tertiary structures are stabilized by the formation of a hydrophobic core, which is the result of the packing of the three helices. Figure 2b shows the organization of those side chains located at the interface between helices I, II, and III, that is, those that define the hydrophobic core of the domain. Helix II and helix III form the helix – turn – helix motif also found in other DNA-binding proteins.[18] The loop region spanning resi-

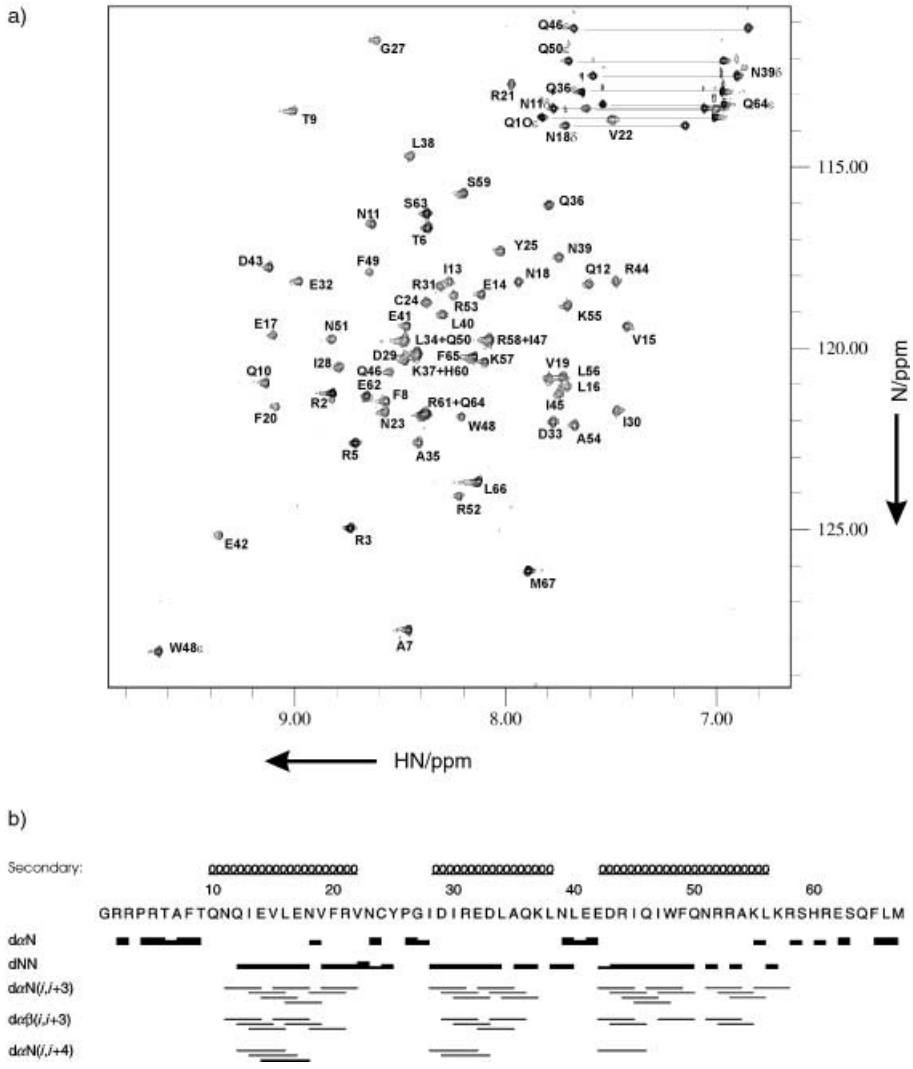

Figure 1. *a) 800 MHz N^{15} HSQC spectrum of the wild-type Hesx-1 homeodomain in sodium phosphate buffer (10 mM) at 283 K and pH 6.0. b) Secondary structure elements identified by analysis of medium- and long-range NOEs.*

dues 22 – 28 is well-ordered in solution and packs against the recognition helix (helix III). This helix – loop interaction could have a significant role in the stabilization of the homeodomain scaffold (see below). Indeed, a detailed inspection of the obtained structure shows the existence of a clear interaction between the guanidinium moiety of the R53 side chain located in the second half of helix III and the aromatic ring of Y25, located in the loop region that consists of residues 22 – 28 (see below).

Structural comparison with other homeodomains: The sequence homology of the Hexs-1 DNA binding domain with most other homeodomains is less than 55 %. This fact has led some authors to classify Hesx-1 as a member of a new class of homeobox proteins, the *Anf* family. None of these *Anf* proteins have been studied by X-ray or NMR spectroscopy methods so far. Close similarities in the overall fold of all the studied homeodomains have been reported. This is not surprising as most of the residues that pack into the core of the individual proteins are well conserved. However, it is important to bear in mind that, despite these global similarities, individual sequence variations may result in significant local differences. In fact, a comparison between solved homeodomain structures reveals that notable differences are located in two defined regions: helix III and the loop region at residues 22 – 28 (connecting helix I and helix II). Thus, the length of helix III varies from 11 residues in the *fushi-tarazu* (*ftz*),[19] vnd-NK2,[20] and Oct-2[21] homeodomains to 19 residues in *Antenappedia* (*Antp*).[22] The loop made up of residues 22 – 28 displays diverse geometries depending on its sequence. It is well known that these local differences may have a large influence on both the DNA-binding properties and the thermal stability of the domain. Thus, the reported melting temperature T_m for other homeodomains varies from 25 °C in vnd-NK2[20] or *ftz*[19] to 61 °C in mat-α2.[23] In addition, DNA affinity and specificity are sensitive to sequence variations in these two protein regions (loop 22 – 28 and helix III), which are both involved in DNA binding according to X-ray and NMR spectroscopy studies. The NMR spectroscopy structure of Hesx-1 shows that helix III is 15 residues long (residues 42 – 56). This is shorter than the helix III described for *Antp* and longer than that observed in NK-2 or *ftz*. In addition, the NMR data suggest that residues 57 – 60 adopt a loose, flexible helical structure.

Table 1. *Statistics for the Hesx-1 wild-type homeodomain and the reduced R53C mutant at different stages of structure refinement[a].*

	No. of NOEs in total (long-range)	DYANA target function	DYANA backbone rmsd	DYANA heavy atom rmsd	AMBER backbone rmsd	AMBER heavy atom rmsd	NOE constraint violation information	
			30 structures	30 structures	30 structures	30 structures	> 0.2 Å	max
Wild-type	607 (359)	0.53 ± 0.04	0.65 ± 0.25	1.52 ± 0.15	0.62 ± 0.13	0.91 ± 0.35	1	0.20
R53C mutant	703 (383)	0.34 ± 0.06	0.79 ± 0.20	1.59 ± 0.18	0.53 ± 0.15	1.11 ± 0.43	1	0.21

[a] The root mean square deviation (rmsd) values (Å) of the backbone and heavy atom superimpositions of residues 8 – 53 on the target function are given as averages (± standard deviation). The number of medium- plus long-range NOEs is shown in brackets.

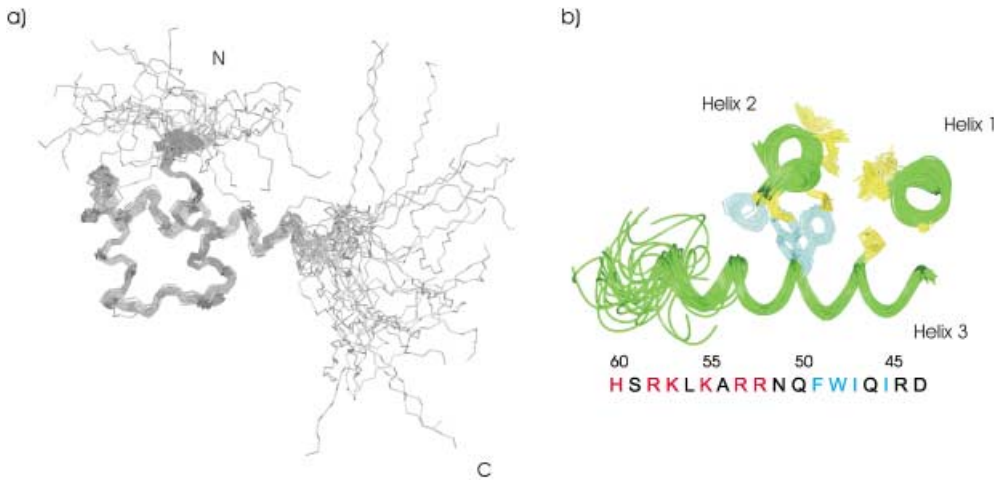

Figure 2. *a) Ensemble of 30 NMR structures of the wild-type Hesx-1 DNA binding domain. b) Hydrophobic residues involved in helix/helix packing. The residue composition of the recognition helix (helix III) is shown at the bottom; red = positively charged residues; blue = hydrophobic residues.*

Sequence homology of *Anf* proteins with homeodomains of other classes is especially low in the loop region (residues 22 – 28), which is presumably involved in DNA binding. The Hesx-1 NMR structure shows that the geometry of this part of the protein closely resembles that observed in the *Antp* paired homeodomain[10] and is clearly different from that reported for mat-α2[24] or Oct-2.[21] This is not surprising as the *Antp* polypeptide has greater homology with Hesx-1 in this protein region. A comparison of the Hexs-1 loop structure with those of other homeodomains is shown in Figure 3. A comparison of the Hexs-1 global fold with that reported for other homeoproteins is shown in Figure S2 of the Supporting Information.

Finally, the solution structure of the human Hesx-1 DNA binding domain shows that the R53 and C24 side chains (also located in the loop region, residues 22 – 28) are close in space, which strongly suggests the possibility of intramolecular residue 53 – 24 disulfide bond formation in the R53C mutant (see below). This information was taken into account for the expression and purification of the mutant. Dithiothreitol (DTT) concentration was increased from 1 to 3 mM and the pH value was lowered from 7.5 to 7.0 throughout the purification process (see the Experimental Section) in order to prevent oxidation of the protein.

NMR solution structure of the R53C mutant (reduced form): Comparison with the wild-type protein

The assignment of the ^{1}H NMR spectrum of the R53C mutant in water was based on data obtained previously for the wild-type protein. In this case, analysis of homonuclear 2D NOESY, DQF-

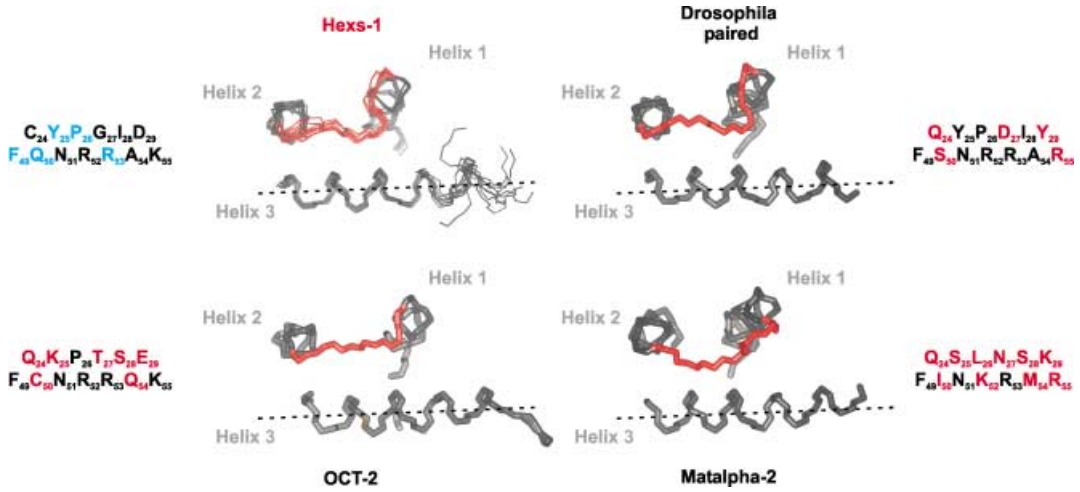

Figure 3. *Structural comparison of the loop region that spans residues 22 – 28 in Hesx-1 and drosophila paired (X-ray structure), Oct-2 (average NMR structure), and mat-α2 (X-ray structure) homeodomains. Those residues involved in contacts between the loop region and helix III are shown in cyan in the Hesx-1 sequence. Sequence substitutions in this region of the proteins with respect to the Hesx-1 sequence are shown in red. Clear similarities between the Hexs-1 and drosophila paired homeodomains can be seen.*

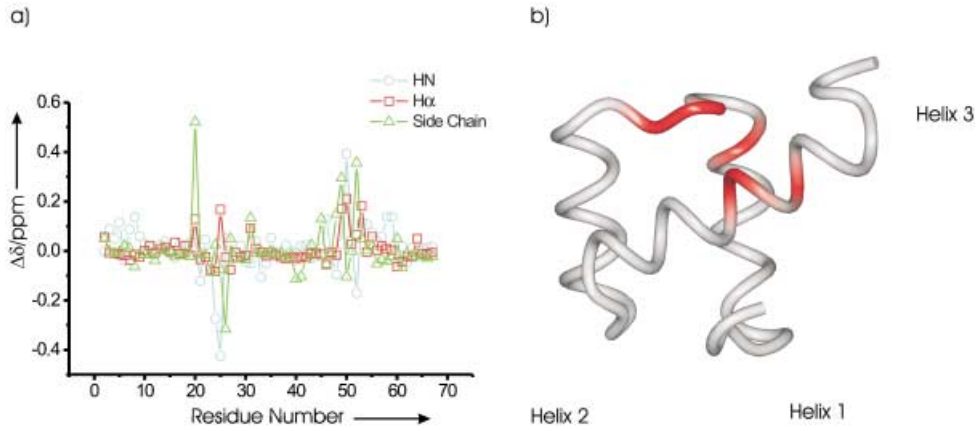

Figure 4. *a) Chemical shift differences between the wild-type and reduced mutant polypeptides for HN (cyan), Hα (red), and side-chain (green) protons. b) Ribbon representation of the wild-type 3D structure. Those residues with $\Delta\delta > 0.2$ ppm are represented in red.*

COSY, and TOCSY experiments were enough to obtain a complete description of the proton chemical shifts. An initial indication of the structural impact of R53C mutation can be easily obtained from a comparison of the proton chemical shifts of the wild-type and mutated polypeptides. This simple analysis ruled out the existence of major structural rearrangements as a result of R53C mutation. Figure 4a shows $\Delta\delta$ values for the HN, Hα, and side-chain protons of the two proteins. It can be clearly observed that three defined regions are mainly affected by the R53C substitution. These regions correspond to the second half of helix III (that is, those residues close to R53 in the polypeptide sequence), the last third of helix I, and the loop region spanning residues 22 – 28. All these regions are close in space, as indicated by the 3D structure of the wild-type Hesx-1 homeodomain (Figure 4b).

A set of 703 NOEs (more than 383 long-range) were unambiguously assigned and converted into distance constraints as described in the Experimental Section. An ensemble of 30 structures was obtained by following a protocol identical to that previously described for the wild-type protein. Figure 5 shows the experimental structure of the mutant in comparison with that of the wild-type protein. Statistics for the wild-type homeodomain and for the reduced R53C mutant at different stages of refinement are shown in Table 1. As expected, the two families of structures are basically identical. Nevertheless, a detailed inspection of spectroscopic and structural data reveals slight differences between the structures. Indeed, some differences in the NOE data were observed that affect precisely those residues located at the end of helix I and in the loop region that spans residues 22 – 28, although these differences were not clearly reflected in the final structures. The mutated polypeptide seems to exhibit a slightly higher helix population, although a subtle difference in the geometry or dynamics of the last part of helix I (18 – 22 region) could also exist. For example, the residue 18 – 21 HN/Hα_{i-3} NOE is fairly intense for the mutant. This contact, typical for helical structures, is almost absent in the wild-type protein. In addition, a higher number of NOEs and more intense effects were observed for the R53C mutant between residues 50 and 25, which are located in helix III and in the loop region of residues 22 – 28. These experimental observations suggest that slight changes do exist in this helix – loop contact with respect to the wild-type protein. As previously mentioned, the wild-type protein structure shows a clear interaction between R53 and Y25. This guanidinium – aromatic interaction probably contributes to the stability of the native homeodomain scaffold. The interaction is totally absent in the R53C mutant as a result of the shorter side-chain length of the cystein residue in the mutant and its different chemical nature. The NMR spectroscopic data also suggest the existence of slight differences in the dynamic behavior of the loop region at residues 22 – 28, with a shorter average helix – loop distance in the mutated protein.

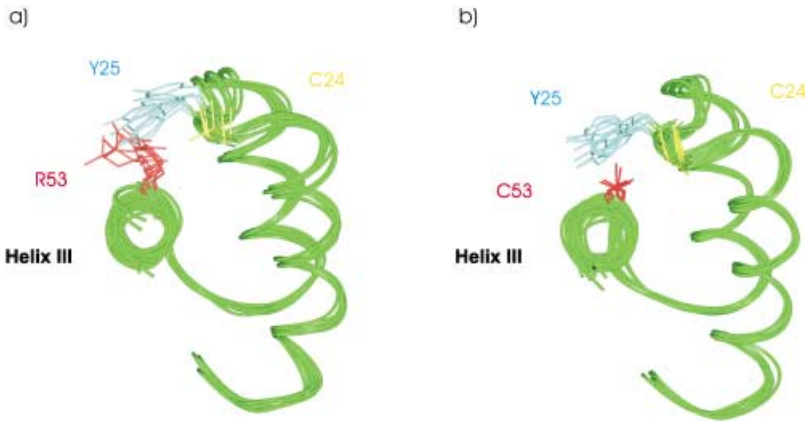

Figure 5. *Experimental NMR structures of the wild-type (a) and reduced R53C mutant (b) proteins. The residues at positions 24, 25, and 53 are shown.*

The NMR structure obtained for the reduced mutant confirms the proximity of C24 to C53 (the cystein introduced by the mutation), which suggests the possibility of intramolecular disulfide bond formation under oxidizing experimental conditions.

Structural analysis of the R53C mutant (oxidized form): NMR spectroscopy and molecular modeling studies

In order to test the hypothesis that disulfide formation occurs in the R53C mutant, the peptide was subjected to a controlled oxidation, carried out under high dilution conditions to prevent the formation of oligomers, as described in the Experimental Section. Titration of the free SH moieties with 5,5-dithio-bis-(2-nitrobenzoic acid) (DTNB), together with MALDI-TOF analysis, unambiguously confirmed the formation of one intramolecular disulfide bond in the mutated polypeptide and ruled out the existence of oligomeric species.

As a next step, several efforts were made to obtain detailed structural information for the oxidized polypeptide by NMR spectroscopy methods. A variety of experimental conditions were tested. Thus, 2D experiments were carried out at different temperatures (278 – 303 K), different ionic strengths (100 – 500 mM NaCl), and different pH values (5 – 7). At low temperature, line broadening made even backbone assignment impossible. In fact, several proton signals almost disappear at 278 – 283 K. Much better quality 2D spectra were obtained at higher temperatures (293 – 303 K). Surprisingly, NMR spectroscopy analysis of the oxidized form under these conditions clearly indicated the existence of structural heterogeneity. Careful inspection of NOESY and TOCSY spectra proved the presence of several conformations in very slow exchange on the chemical shift time scale (no exchange peaks were deduced from the analysis of ROESY experiments). This conformational heterogeneity can not be attributed to the presence of oligomeric species (see above). Reduction of the disulfide bond between residues 53 and 24 can be performed directly in the NMR spectroscopy tube by addition of deuterated DTT. After a couple of hours, a 1D spectrum identical to that of the reduced homeodomain was obtained (Figure 6a). Comparison of 2D experiments at different temperatures shows that the relative population of the different species present in solution is dependent upon temperature (Figure 6b). This structural heterogeneity seems to affect mainly those aromatic residues involved in helix I/helix III packing. Analysis of NOESY and TOCSY experiments at 293 – 303 K indicates that helix II and the first third of helix III remain unaffected by the disulfide formation between residues 24 and 53; all residues in these protein regions provide single sharp signals with chemical shifts almost identical to those observed for the reduced polypeptide and could be fully assigned. In contrast, at least three different sets of signals were identified for residues F8, F49, F20, and W48 (see Figure S3 in the Supporting Information). Some additional aromatic systems remained unassigned. Figure 6a shows the HN/aromatic region of the spectra for the oxidized and reduced forms of the mutated Hesx-1 DNA binding domain. The existence of at least three different HN signals for the W48 indole ring in the oxidized form can be observed. These three signals collapse to give one sharp signal shortly after the addition of DTT, which indicates that the observed structural heterogeneity is directly related to the presence of the C24 – C53 disulfide bond.

We were able to obtain an almost complete backbone (HN/Hα) assignment for one of the most populated conformations present in solution at 293 K. Inspection of the short – medium-range NOEs for this species also confirmed the presence of the three helical regions previously observed for the reduced mutant. Therefore, a significant population of the oxidized polypeptide maintains the secondary structure elements characteristic of the wild-type protein and the reduced R53C mutant. In conclusion, although secondary structure seems to be mainly preserved according to NMR data, formation of a disulfide bond between positions 24 and 53 seems to have a significant effect on helix I – helix III interaction.

It was not possible to get NMR-based high resolution structural data for the oxidized homeodomain. The low intensity of some signals in the most affected protein regions made it impossible to obtain a complete proton assignment of the side chains, even for the most populated conformers. In addition, those residues that could be unambiguously assigned provided an extremely low number of NOEs (apart from some weak HN_i/$H\alpha_{i-3}$ contacts in helical regions, in most cases only weak intraresidue and sequential cross-peaks could be identified). Both effects could be related to the low effective concentration of each individual species in equilibrium and/or to an increase in mobility in those regions of the polypeptide affected by disulfide formation. Finally it must be mentioned that at NMR spectoscopy concentrations (0.5 – 1 mM) the oxidized homeodomain exhibits a high tendency to precipitate at $T > 10\,^{\circ}C$. A different NMR sample had to be prepared for each single homonuclear 2D experiment.

In order to get an insight into the nature of this structural perturbation, a 5-ns, unrestrained molecular dynamics (MD) simulation in the presence of explicit solvent and counterions was computed for the oxidized mutant. MD simulations for the wild-type and reduced mutant were also performed as an independent check on the validity of the calculations for the oxidized mutant. The results obtained provided additional information on the dynamics and structural features of these two proteins and permitted a direct comparison between the simulated and NMR spectroscopy data. The experimental NMR structures were used as a starting point in the calculations for the wild-type protein and the reduced R53C mutant. The starting structure for the oxidized mutant was generated from the experimental coordinates corresponding to the reduced form (see the Experimental Section). All simulations were performed by using periodic boundary conditions and the particle mesh Ewald approach[25] to introduce long-range electrostatic effects. In all cases, protein structures were represented by using AMBER-95 force-field parameters,[16] while water was represented by means of the TIP3P model.[26] Figure 7 shows backbone rms deviations with respect to the initial (experimental) structures for the wild-type protein (blue), reduced (green), and oxidized (red) R53C mutants during the 5-ns MD simulations. After a 5-ns run, the deviation from the NMR spectroscopy experimental coordinates is below 1.3 Å for both the wild-type homeodomain and the reduced mutant. Taking into account that the resolution of the NMR structures is around 0.7 Å (for backbone superimposition), these numbers represent an excellent agreement between theoretical and experimental data. Therefore, both MD simulations and NMR structures indicate the

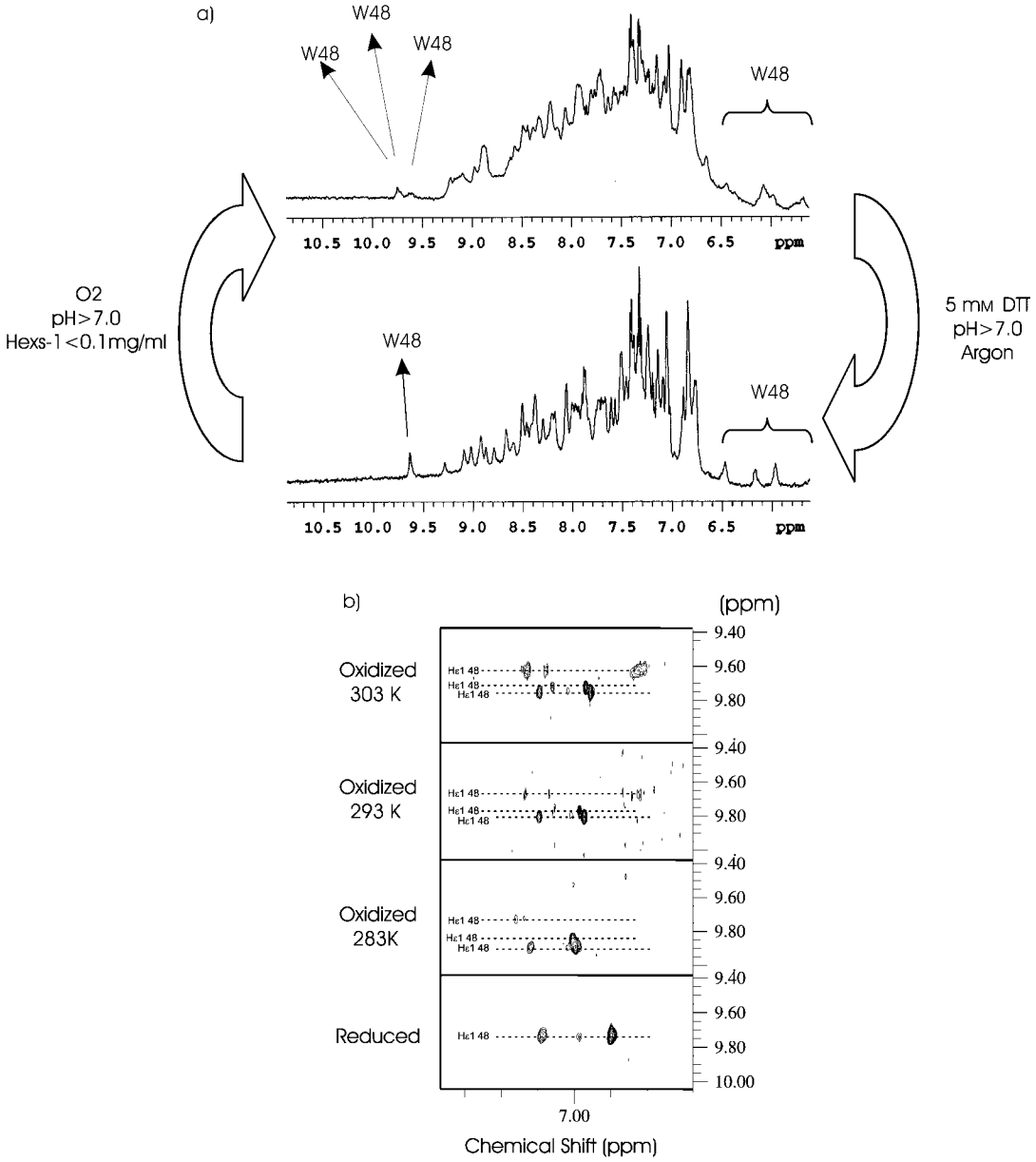

Figure 6. *a) 1D NMR spectrum corresponding to the aromatic/HN region of the reduced (bottom) and oxidized (top) R53C mutant homeodomains. The existence of at least three signals from the indole HN proton of W48 in the oxidized form can be clearly seen. b) Dependence of the populations of the different conformations on temperature: a region of a NOESY spectrum showing the intraresidue Hε1/Hδ1 and Hε1/Hζ2 contacts in the indole ring of W48 at different temperatures. The presence of at least three different indole systems is clear at temperatures of 293 K or higher.*

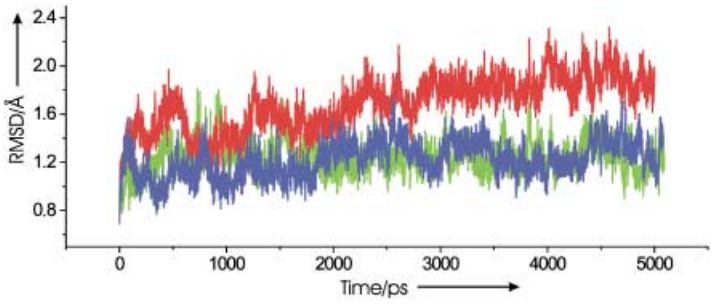

Figure 7. *Backbone (residues 8 – 53) rms deviations with respect to the initial (experimental) structures of the wild-type (blue) and reduced (green) and oxidized (red) R53C mutant proteins measured during the 5-ns MD simulations.*

existence of minimal structural differences between the wild-type protein and the reduced mutant and it is probable that the MD simulations obtained under these conditions provide a fair description of the actual situation, not only for these polypeptides but also for the oxidized mutant. It is a key point that the stacking interaction between the guanidinium moiety of R53 and the aromatic ring of Y25, previously deduced from the analysis of NMR data for the wild-type homeodomain, is maintained during the 5-ns MD run (Figure 8). In contrast, this loop – helix interaction is totally absent in the reduced mutant, as was also deduced from the NMR spectroscopy analysis. The MD simulation of the mutant indicates the existence of a high degree of flexibility at position 25. This flexibility could have

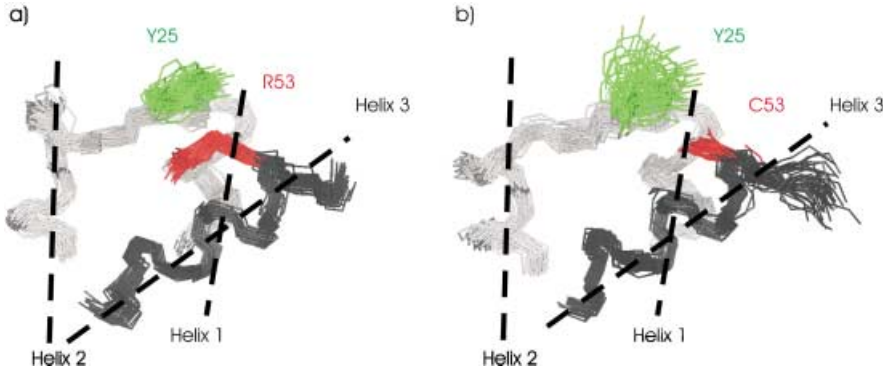

Figure 8. *Loop – helix-III interactions according to MD simulations for both the wild-type protein (a) and reduced R53C mutant (b). The stacking interaction between the R53 side chain and Y25 in the wild-type homeodomain can be seen. This contact does not exist in the mutated polypeptide. Snap-shots were taken every 50 ps for the last 2 ns of the MD trajectories.*

significant consequences for DNA recognition by the mutated polypeptide (see below). Observed variations in the distance between the sulfur atoms of the cystein residues at positions 24 and 53 during the MD trajectory for the reduced mutant confirm the possibility of intramolecular disulfide formation (Figure S4, Supporting Information).

Completely different behavior is observed for the oxidized protein. In this case, a larger rms deviation (Figure 7) of up to 1.9 Å is observed with respect to the initial structure, which indicates that the generated disulfide bond is not totally compatible with the geometry of the homeodomain scaffold. In fact, some kind of rearrangement has to take place in order to stabilize the oxidized structure. Figure S5 (see the Supporting Information) shows the rms deviations with respect to the initial structure for the hydrophobic side chains involved in the helix I/helix III packing. According to the MD simulations, most of the structural perturbations observed in the oxidized mutant affect this region of the protein (the rms values are larger than 2.5 Å for the amino acids in these regions). Indeed, the MD data strongly support the existence of significant perturbations of the helix-I – helix-III interaction as was also deduced from the qualitative NMR spectroscopy analysis. Figure 9a – c shows helix I/helix III packing during the last 2 ns of the trajectory (once complete equilibration of the system has been achieved) for the wild-type polypeptide and the reduced and oxidized mutants. It can be clearly seen that a significant change in the angle between helix I and helix III takes place for the oxidized mutant as a consequence of the residue 24 – 53 disulfide formation. This angle is around 55 – 75 degrees for both the wild-type polypeptide and the reduced mutant, while it amounts to 40 – 50 degrees for the oxidized mutant. Moreover, this change in the relative orientation of helices I and III also affects the arrangement of those hydrophobic side chains

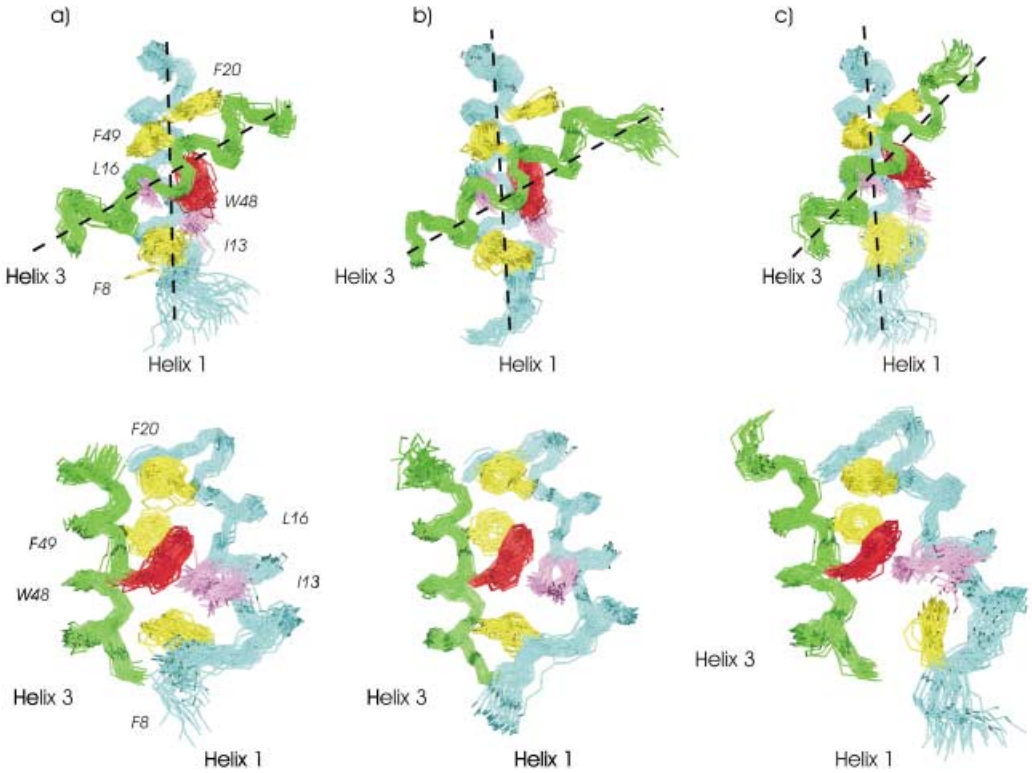

Figure 9. *Helix I/helix III packing during to the MD simulations for the wild-type polypeptide (a) and the reduced (b) and oxidized (c) mutants. The differences in helix/helix angle between the oxidized mutant (c) and both the reduced mutant (b) and wild-type species (a) is evident (top). In addition, alterations in the structure of the hydrophobic core of the oxidized mutant can be clearly observed (bottom). Snap-shots were taken every 50 ps for the last 2 ns of the trajectories (when the system was totally equilibrated).*

involved in helix/helix packing. Significant changes in side-chain – side-chain interactions (lower part of Figure 9) are evident for some residues. The relative orientation of the W48 and F49 aromatic rings located in the center of the hydrophobic core is slightly modified. This rearrangement is in agreement with the chemical shift variations observed for W48 in the NMR spectra of the most populated conformers of the oxidized mutant. In addition, the packing of the F8 aromatic side chain against the W48 indole ring present in both the wild-type homeodomain and the reduced mutant is totally lost in the oxidized form. In conclusion, both NMR spectroscopy and modeling studies indicate the existence of structural changes in the mutated Hesx-1 as a consequence of disulfide formation. Although secondary structural elements are mainly preserved, key changes in the tertiary structure take place. These changes mainly involve variations in the helix I/helix III packing that are essential to accommodate a covalent bond between positions 24 and 53. The NMR spectroscopy experiments are consistent with the existence of multiple conformers in very slow exchange. Clearly, this effect cannot be adequately reproduced by the MD simulations as much longer trajectories would be required. Nevertheless, both the NMR and the theoretical MD data point to a common explanation for this experimental observation. Deviation in helix I/helix III packing from the optimal unconstrained geometry probably prevents the most favorable interactions between hydrophobic side chains and thus generates a mixture of different nonoptimal arrangements with similar energy. Clearly, this slight but measurable change in the protein structure and dynamics also has an impact on protein stability.

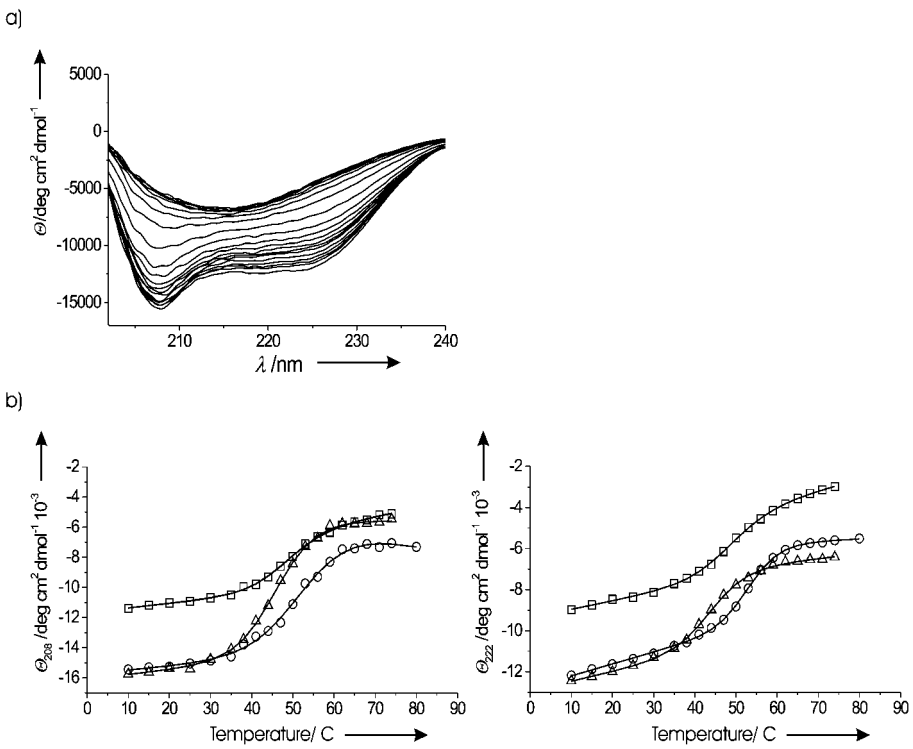

Figure 10. *a) Far-UV CD spectra of the wild-type protein at different temperatures (283 – 348 K). b) Melting of the wild-type polypeptide (circles) and both reduced (triangles) and oxidized mutants (squares) monitored by circular dicroism at 222 nm (left) and 208 nm (right). Continuous lines are fitted curves. In all cases, experiments were carried out at 50 μM protein concentration in sodium phosphate buffer (10mM) at pH 6.3.*

Influence of R53C mutation on protein stability

In order to determine the influence of R53C substitution on the stability of the Hesx-1 homeodomain scaffold, the conformational stabilities of the wild-type protein and both reduced and oxidized mutants were analyzed by circular dichroism measurements. Melting curves were monitored by recording the far-UV CD spectra at different temperatures between 283 and 348 K (Figure 10). Progress curves for ellipticity changes at 222 nm as a function of increasing temperature were also collected by using two different rates for temperature increase (see the Experimental Section and Figure S6 of the Supporting Information). The thermal transitions involving the three homeodomains were found to be partially reversible. However, the CD profiles recorded at different scan rates did not show significant variations. Therefore, kinetic control of the denaturation processes can be discarded and thermodynamic analysis of the CD curves is possible.[27, 28] All these CD data were analyzed assuming a two-state model. The fact that a good fit of the theoretical curves to the experimental data was obtained for the oxidized mutant, where NMR spectroscopy analysis proves the existence of different forms in slow exchange, probably indicates that all these different species have similar stabilities. The values obtained for T_m and ΔH are shown in Table 2. Two different conclusions can be drawn from these results. First, the substitution of R53 by a cystein residue has a very significant effect on the thermal stability of the Hesx-1 DNA binding domain, with a decrease in T_m of 10 K. This destabilization has a mainly enthalpic origin, as shown by the corresponding van't Hoff ΔH

Table 2. *Thermodynamic parameters derived from the analysis of CD melting curves for the wild-type Hesx-1 homeodomain and both the reduced and oxidized R53C mutants.[a]*

	Melting temperature [K]	ΔH [Kcal mol^{-1}]
Wild-type	328	55
Reduced mutant	318	42
Oxidized mutant	321	38

[a] Experiments were carried out in sodium phosphate buffer (10 mM) at pH 6.3 and 50 μM protein concentration and results were analyzed assuming a two-state model. The estimated errors are ±0.5 K for the melting temperatures and ±10% for the the van't Hoff enthalpies.

values (Table 2). The structural origin of this effect on Hesx-1 stability is probably related to the stacking interaction previously observed in the wild-type protein between positions Y25 and R53. As discussed above, both the NMR and MD data indicate that this interaction is totally disrupted by R53C substitution. Although R53C substitution has a minimal effect on the average structure of the homeodomain, this loop – helix interaction seems to have a very significant effect on the thermal stability, according to the experimental denaturation curves.

The second conclusion is that oxidation of the R53C mutant through intramolecular disulfide formation has a somewhat stabilizing effect ($\Delta T_m = 3$ K) with respect to the reduced form. As expected, this stabilization has an entropic origin and is probably related to the reduced conformational freedom of the denatured state. Although differences between the enthalpic terms of the two polypeptides are almost within the experimental error, the CD data suggest a more favorable ΔH value for the reduced mutant. The entropic benefit derived from disulfide formation is thus partially cancelled in the oxidized mutant by a less favorable enthalpic contribution. This result is in perfect agreement with the changes in tertiary structure previously deduced from both NMR spectroscopy and modeling studies. Disulfide formation probably disrupts the optimal helix-I – helix-III packing interactions, which results in a high structural heterogeneity and a less favorable enthalpic term for protein folding.

Studies on DNA recognition by the human Hesx-1 homeodomain: Impact of R53C mutation on protein – DNA interaction

Interaction of the wild-type homeodomain with DNA: First, the interaction between the wild-type polypeptide and the target DNA, which contained a palindromic PIII sequence (GTC**TAAT**TGA**ATTA**ACG), was analyzed at low ionic strength (10 mM sodium phosphate, pH 6.3) by far-UV CD (Figure 11 a). This DNA sequence was chosen because it has been previously reported that Hesx-1 is able to bind the sequence to give a 2:1 protein:DNA complex.[29] Comparison of circular dichroism spectra of the protein, DNA, and protein/DNA complex reveals a significant conformational change in the protein upon binding (Figure 11 a). Subtraction of the DNA component from the protein – DNA complex CD spectrum leaves a difference spectrum (Figure 11 a) with more than 20% ellipticity increase at 222 nm. The almost perfect subtraction of the DNA signal in the 240 – 320 nm range supports the assignment of the spectral variation to a conformational change that occurrs in the protein and not in the DNA component of the complex. In addition, the shape of the difference CD curve is consistent with an increase in helical structure upon complex formation. An increase in α helix content has been described for other homeodomains as a result of protein – DNA interaction.[23, 30, 31] The structural origin of this effect is probably directly related to the residue composition of

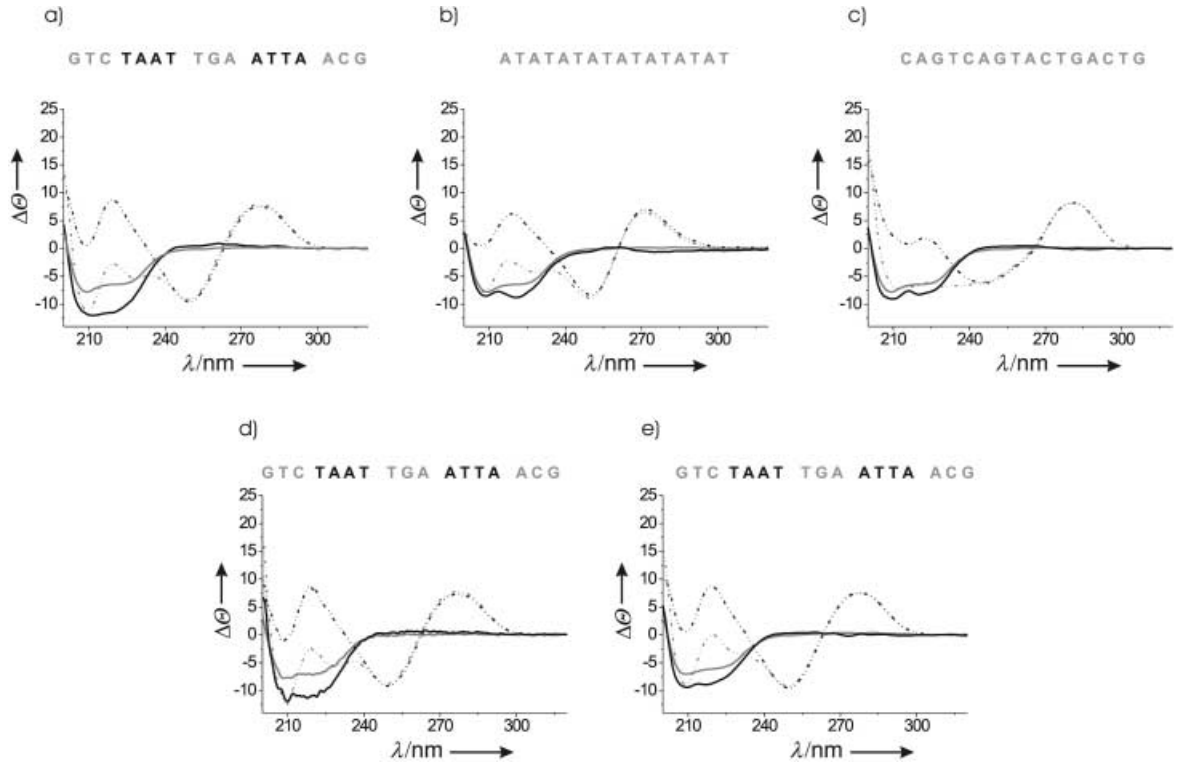

Figure 11. Binding studies by circular dichroism: a – c) Effect of DNA binding on the CD spectrum of wild-type Hesx-1 when both specific (a) and nonspecific (b – c) DNA sequences are employed. a) CD spectra of the wild-type protein (grey), DNA (black dotted), and the protein – DNA complex (grey dotted) at pH 6.3 and 303 K in sodium phosphate buffer (10 mM). Subtraction of the DNA contribution from the protein – DNA CD spectrum leaves a difference spectrum (black) with more than 20% ellipticity at 222 nm. The same experiment was carried out with the palindromic P3 DNA (a) and two nonspecific sequences (b and c). DNA sequences are shown in the upper part of the figure. d – e) Effect of binding of the palindromic P3 DNA sequence on the CD spectra of the reduced (d) and oxidized (e) mutants. In all cases, spectra were recorded at 10 μM protein concentration and 20 μM DNA concentration. Ellipticities are expressed in deg cm^2 dm^{-1} × 10^{-3}.

helix III. Two different components can be distinguished in this region of the protein according to its sequence (Figure 2b). The helical region that spans residues 41 – 51 is basically amphipathic in nature and is characterized by the existence of a large number of hydrophobic residues. This part of helix III packs against both helix I and II and defines a significant part of the protein hydrophobic core. In contrast, the helical region that spans residues 52 to 60 includes several polar, positively charged lysine and arginine residues. According to our NMR spectroscopy analysis, this segment of the protein (especially the last 5 residues), though not totally disordered in the free state, is characterized by a low helical content. It is well known that some transcription factors of the basic leucine zipper and basic helix – loop – helix families specifically recognize DNA by means of intrinsically flexible, positively charged, peptide segments that assume a helical conformation upon binding to target DNA sequences.[32] It has also been shown that 20 base pair double-stranded DNA oligonucleotides can act as templates to promote random coil to α helix transitions in short peptides that contain conveniently positioned alanine and lysine/arginine residues through the establishment of nonspecific polar arginine/lysine – phosphate interactions.[32] On this basis, it seems reasonable to assign the observed effect on the CD spectra to an increase in the helical content of the region spanning residues 52 – 60 upon DNA binding. In any case, these changes in the CD spectra constitute a convenient tool to monitor the Hesx-1 – DNA interaction. The same experiments were carried out by employing two different protein:DNA ratios with similar results, which suggests a high degree of protein saturation under the employed experimental conditions. As a control, the nonspecific interactions of the Hesx-1 homeodomain with two other DNA sequences that do not include the target P3 site were also studied. The difference CD spectrum showed much smaller variations in the protein spectra with both these DNA fragments, which suggests a decrease in binding affinity (Figure 11 b – c). Therefore, it can be concluded that the Hesx-1 homeodomain interacts with PIII DNA with high affinity (the apparent affinity constant K_a is greater than $10^6\,\text{M}^{-1}$ at low ionic strengh) and that the observed interaction is sequence specific.

Several attempts were made to obtain K_a values by monitoring protein ellipticity changes at 222 nm. However, at low ionic strength, precipitation of the sample at protein:DNA ratios larger that 4:1, probably as a result of the nonspecific aggregation of more protein molecules on the 2:1 protein:DNA complex, precluded the acquisition of reliable data. In order to minimize this effect, a set of CD titration experiments was also carried out at high ionic strength (10 mM sodium phosphate, 200 mM NaCl, 20 mM $MgCl_2$, pH 6.3). Figure 12a shows a typical titration curve obtained for the wild-type-protein – P3-DNA interaction by monitoring the changes in protein ellipticity at 222 nm upon addition of DNA (the DNA contribution was substracted for each individual point). The curves were fitted assuming a 2:1 stochiometry in the protein – DNA complex and two independent and equivalent binding sites in the oligonucleotide. The apparent affinity constant determined for the palindromic P3 sequence in this way (Table 3) was $K_a = 60000\,\text{M}^{-1}$. In order to test the above model, CD titrations were repeated with oligonucleotides that contained only one binding site on the DNA sequence (HS-DNA, Figure 12a). Unexpectedly, the obtained K_a value was larger than the apparent one measured for the complete P3 site, which strongly suggests the existence of negative cooperativity in the case of the P3 DNA. When the binding studies were performed with the palindromic sequence P5 (Table 3), the apparent binding constant observed was similar to that obtained for HS-DNA (Table 3). It can be speculated that in this case the consensus recognition sequences (TAAT) in the DNA fragment are sufficiently far apart to avoid unfavorable interactions between the two homeodomains in the bound state. Hesx-1 has been classified by some authors as a member of the paired-like class of homeobox genes.[5] It is known that the proteins included in this class bind cooperatively to palindromic DNA sequences and rely on the conserved 60-residue homeodomain to achieve cooperativity.[10, 33] In contrast, our results indicate that the interaction between two Hesx-1 homeodo-

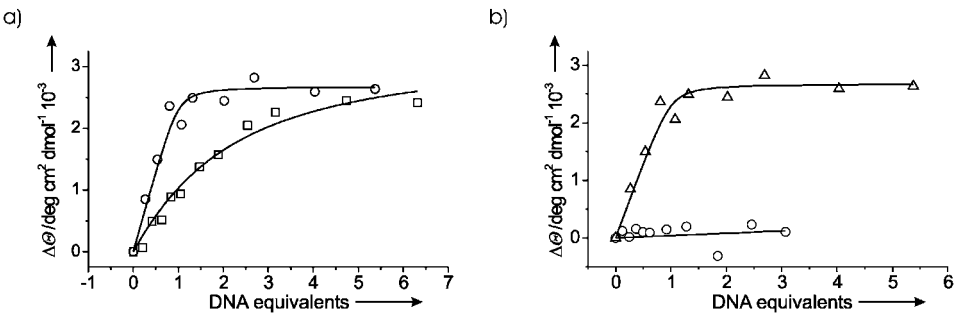

Figure 12. *CD titration studies of the proteins (20 μM) with increasing amounts of DNA in sodium phosphate buffer (10 mM), NaCl (200 mM), and $MgCl_2$ (20 mM) at 303 K and pH 6.3. The variations in protein ellipticity can be easily followed once the DNA contribution has been substracted for each point. Five different wavelengths (220 – 225 nm) were employed in every case and the obtained K_a values were averaged. a) Titration plot obtained with wild-type Hesx-1 and the palindromic P3 DNA fragment (squares) or the HS-DNA fragment (circles). b) CD titration plots corresponding to the interaction of the wild-type polypeptide (triangles) and reduced mutant (circles) with the HS-DNA duplex.*

Table 3. *Affinity constants, K_a, for the interaction of wild-type Hesx-1 and the mutants with different DNA sequences.[a]*

DNA sequence (5′ – 3′)	Wild-type	Reduced mutant	Oxidized mutant
P3: GTC**TAAT**TGA**ATTA**ACG	$6 \times 10^4 \pm 2 \times 10^4$	< 1000	< 1000
P5: GTC**TAAT**TGCGA**ATTA**ACG	$1.5 \times 10^6 \pm 5 \times 10^5$	< 1000	< 1000
HS: GTC**TAAT**TGACGCG	$2 \times 10^6 \pm 1 \times 10^6$	< 1000	< 1000

[a] Values are given in units of M^{-1}. Experiments were carried out at 303 K and pH 6.3 in sodium phosphate buffer (10 mM), NaCl (200 mM), and $MgCl_2$ (20 mM). Apparent affinity constants were estimated assuming two independent binding sites on the DNA molecule for the DNA fragments that included the palindromic P3 and P5 sequences.

mains and the palindromic P3 DNA fragment exhibits negative cooperativity, at least under the employed experimental conditions. This result supports the inclusion of Hesx-1 as a member of the new *Anf* family. Comparison of the Hesx-1 homeodomain sequence with those of other known paired-like homeodomains reveals some differences in those regions that are presumably involved in protein – protein contacts upon binding to palindromic P3 DNA fragments (Figure S7, Supporting Information). These key residues may be deduced from the analysis of the X-ray structure of a paired-class cooperative homeodomain dimers in complex with DNA.[10] The residues in positions 27, 29, and 43, usually D, F/Y, and A in most paired-like homeodomains, are replaced by G, D, and D, respectively, in Hesx-1. These nonconservative substitutions could be at the origin of the different behavior exhibited by the Hesx-1 homeodomain towards the palindromic P3 DNA sequence. However, it has been shown[4] that the whole Hesx-1 protein is able to bind DNA as a dimer and that the dimerization constant of the protein is not affected by R53C substitution. Protein regions outside the DNA binding domain are probably responsible for protein dimerization. Under these conditions, stabilizing interactions between homeodomain regions within the protein – DNA ternary complex are clearly not required for efficient binding to the palindromic P3 DNA sequence. In this context, the biological significance of the negative cooperativity effect described above remains an open question.

Effect of R53C substitution on DNA binding activity: The same set of experiments as that described above was carried out for the R53C mutant in both reduced and oxidized forms to obtain a first indication of their DNA binding activities. Figure 11 a, d – e shows the difference CD curves obtained for the wild-type protein and the mutant (in both reduced and oxidized forms) upon binding to the P3-DNA sequence. These curves clearly indicate the existence of significant DNA binding activity in the reduced mutant. Indeed, the influence of DNA binding on the CD spectra of the reduced polypeptide is similar to that previously described for the wild-type protein and larger than that observed for the oxidized form. No significant differences in DNA binding activity between the wild-type homeodomain and the reduced mutant can be deduced under these experimental conditions (low ionic strength, 10 mM sodium phosphate, pH 6.3,). This observation suggests that the association constant for the reduced mutant under these conditions must be larger than $10^5\,M^{-1}$.

Finally, in order to quantify the DNA binding activity of the mutants, CD titrations were carried out for both the oxidized and reduced forms at high ionic strength with the half-site containing DNA and the palindromic P3 and P5 sequences (Table 3). In all cases, a much lower affinity than that deduced above or that measured for the wild-type peptide was detected. The observed low affinity prevented an accurate estimation of K_a values. Nevertheless, an upper limit of $1000\,M^{-1}$ was established for both forms and all DNA sequences (Figure 12 b, Table 3). This result implies a reduction in binding activity of at least three orders of magnitude for the half-site-containing DNA sequence as result of the R53C substitution. Accordingly, the energy cost associated with R53C mutation with respect to Hesx-1 – DNA complex stability is at least 4.1 Kcal mol^{-1} at 303 K.

Although the NMR spectroscopy analysis of the reduced mutant indicated the absence of major structural changes with respect to the wild-type protein, subtle differences in the arrangement and/or dynamics of key side chains may have significant effects on the thermodynamic balance of the protein – DNA recognition process. According to the NMR structure of the wild-type protein, R53 is involved in a stacking interaction with the aromatic ring of Y25. In contrast, this loop – helix interaction is totally absent in the reduced mutant. Our results strongly suggest that this interaction has a remarkable effect on DNA recognition. Both R53 and Y25 are conserved in most homeodomain sequences and both have been shown to participate in DNA recognition through the establishment of nonspecific polar interactions with phosphate groups. Interestingly, the interaction between these two side chains is maintained in most homeodomain – DNA complexes studied so far. Figure 13 shows this contact in a paired-class (Pax) cooperative

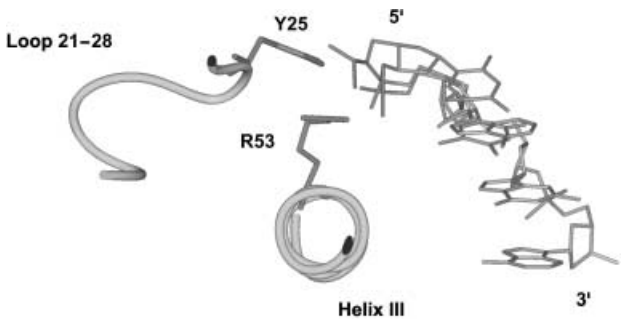

Figure 13. *Stacking interaction between Y25 and R53 according to X-ray data from a paired class cooperative homeodomain dimer complexed with DNA.[10].*

homeodomain dimer in complex with DNA.[10] This R53 – Y25 stacking probably preorganizes the side chains of the amino acids into the appropriate orientation for hydrogen bonding to DNA phosphate groups. In this sense, disruption of the R53 – Y25 interaction by the R53C mutation would imply not only the loss of a single arginine – phosphate polar contact in the protein – DNA complex, but also the loss of the proper Y25 – phosphate interaction. This effect could partially explain the large influence that R53C substitution has on protein – DNA complex stability. According to this hypothesis, R53C substitution would reduce homeodomain binding activity by significantly decreasing the contribution of nonspecific polar interactions to the global ΔG value.

Although we were not able to quantitatively measure the effect of disulfide formation on DNA recognition by the mutant, CD experiments carried out at low ionic strength suggest a decrease in DNA binding activity for the oxidized polypeptide with respect to the reduced mutant. Docking of structural models of the oxidized polypeptide to B-DNA indicates that the deduced changes in helix I/helix III packing could have a significant effect on protein – DNA complex stability, mainly through their effect on those contacts established by the protein

in the region of residues 22 – 31. According to both paired and engrailed X-ray structures of protein – DNA complexes, this protein region is involved in nonspecific polar contacts with DNA. Therefore, a further reduction in the contribution of nonspecific polar interactions to the ΔG value would be expected as a result of disulfide formation in the mutated protein.

The reported effect of R53C substitution on protein stability and DNA-binding properties, together with the significant structural perturbations induced by intramolecular disulfide bond formation in the mutated polypeptide, might help to explain the loss of activity of the mutant in vivo.

Experimental Section

Plasmid construction: The Hesx-1 wild-type and R53C mutated cDNAs cloned into a His-tagged pJBE21 expression vector were kindly provided by Dr. M. T. Dattani (London Centre for Paediatric Endocrinology and Metabolism). The cDNA fragments encoding the corresponding homeodomains were amplified by PCR by using the primers 5′-GGAATTCCATATGGGCCGAAGACCAAGAACTGC-3′ and 5′-CGGGATCCTTACATTAGAAACTGTGATTCTC-3′. The resulting fragments were digested with *NdeI* and *BamHI* endonucleases and cloned into the digested plasmid pT7 – 7. The obtained plasmids, pHESXWT and pHESXM, were verified by the DNA automatic sequencing service at Centro de Investigaciones Biológicas (CSIC).

Protein expression and purification: Plasmids pHESXWT and pHESXM were expressed in *Escherischia Coli BL21*(De3). Transformed cells were grown to an optical density (OD_{600}) value of 0.8 at 37 °C in LB medium then cultures were induced for 2.5 h with isopropyl – D-thiogalactopyranoside (0.5 mM). Cells were harvested and lysed, followed by clarification and removal of nucleic acids by polyethyleneimine precipitation. Polyethyleneimine precipitates were dialyzed against sodium phosphate buffer A (50 mM; pH 7.5), which contained DTT (1 mM), and both wild-type and mutant homeodomain peptides were purified by two chromatographic steps. The dialyzed fraction of the wild-type protein was loaded onto a BioRex A-70 (BioRad) column equilibrated in buffer A. The column was extensively washed with buffer A and then the protein was eluted with a linear gradient of NaCl (0 – 300 mM) in buffer A. The protein fractions that contained homeodomain were pooled and dialyzed against buffer A and the desalted protein solution was applied to a 5-mL Econo-Pac S cartridge (BioRad) equilibrated in buffer A. The column was extensively washed with buffer A and then the protein was again eluted with a linear gradient of NaCl (0 – 300 mM) in buffer A. The pure homeodomain fractions were pooled and desalted by dialysis against sodium phosphate buffer (2 mM; pH 6.5). The eluted fractions were checked for the homeodomain by SDS-PAGE on 17% acrylamide gel. At the end of the procedure, the purity of the protein was shown to be at least 95%. The protein concentration was determined by the Bradford method and also spectrophotometrically. The mutant peptide was expressed and purified as described for the wild-type protein, with buffer A replaced by sodium phosphate buffer (50 mM; pH 7.0) that contained DTT (3 mM). In order to obtain the oxidized mutant polypeptide, the pure protein was oxidized under very mild conditions: room temperature, sodium phosphate buffer (10 mM; pH 8.0), protein concentration below 0.1 mg ml^{-1}, under an air atmosphere. Disulfide formation was probed by MALDI-TOF mass spectrometry and titration of free SH groups with DTNB.

NMR experiments: ^1H NMR spectra were recorded in 85:15 $^1H_2O/^2H_2O$ on Bruker Avance 800, Bruker AMX-600, and Varian Unity 500 MHz spectrometers. The 2D-homonuclear experiments were performed on solutions (1.0 mM) at 278 – 303 K. In addition, 2D ^{15}N HSQC and 3D HSQC NOESY spectra were obtained for the wild-type protein by using a ^{15}N-labeled sample (1 mM) at 278 K. 2D HSQC experiments were carried out with spectral widths of 1784 Hz (^{15}N) and 10415 Hz (^1H) and 64 scans per increment. Experimental data consisted of 4096×300 complex points zero filled to 4096×1024 complex points. 3D NOESY HSQC spectra were collected with 32 scans per increment, mixing times of 100 and 150 ms, and a spectral width of 10416 Hz in both w1 and w3 and 1780 Hz in w2. A total of $192 \times 48 \times 4096$ data points were obtained. Quadrature detection in both w1 and w2 was obtained by using the time-propotional phase incrementation (TPPI-States) method.[34]

TOCSY,[35] NOESY,[36] and DQF-COSY[37] experiments were carried out in the phase-sensitive mode by using the TPPI method[37] for quadrature detection in F1. Typically, a data matrix of 512×2000 points was used to digitize a spectral width of 7800 Hz. 80 scans were used per increment with a relaxation delay of 1 s. Prior to Fourier transformation, zero filling was used in F1 to expand the data to 1000×2000 points. Baseline correction was applied in both dimensions. The corresponding shift was optimized for the different spectra. The TOCSY spectra were acquired by using the MLEV-17 sequence[35] during the 60-ms isotropic mixing period. The NOESY experiments were performed with mixing times of 100 – 200 ms.

Structure calculations: Upper limits for proton – proton distances were obtained from NOESY cross-peak intensities at three mixing times (100, 150, and 200 ms) by using the isolated spin pair approximation. Cross peaks were classified as strong, medium, and weak, which corresponded to upper limits of 2.5, 3.5, and 5.0 Å, respectively. The lower limit for proton – proton distances was set as the sum of the van der Waals radii of the protons. Distance geometry calculations were performed on a Silicon Graphics O2 computer with the program DYANA.[14] A set of 607 constraints for the wild-type homeodomain and 703 for the mutant protein were used in the final round of calculations, besides several structurally meaningless constraints.

The thirty best DYANA-program structures in terms of the target function were submitted to a simulated annealing protocol[38] that used the AMBER 5.0 package and the Cornell et al. force field.[16] Explicit TIP3P[16] water molecules and periodic boundary conditions were employed in these calculations to prevent unrealistic interactions between disordered regions of the protein and the structured helical core. After an initial restrained energy minimization (REM) with 1000 steepest-descent iterations, the structures were heated to 500 K in 10 ps and, at this temperature, 10 ps of restrained molecular dynamics (RMD) were performed. The structures were then subjected to a cooling procedure for 20 ps until a temperature of 1 K was reached. The final structures were energy minimized (REM) by using 2000 conjugate gradient iterations.

A similar protocol was used for the structure calculation of the reduced R53C mutant. In this case, 703 distance restraints were used in the final refinement.

Molecular modeling: MD simulations were carried out by using the sander module within the AMBER 5.0 package and the Cornell et al. force field.[16] The NMR spectroscopy experimental structures were employed as starting geometries for both the wild-type and reduced mutant homeodomains. The starting structure for the oxidized polypeptide was generated with the DYANA program and the experimental restraints of the reduced form. The C24 – C53 disulfide bridge was included in the DYANA calculation in the form of distance

constraints for S—S (2.0 Å $< r <$ 2.1 Å) and for Cβ—S (3.0 Å $< r <$ 3.1 Å) atomic separations. The three homeodomains were immersed in a bath of 3000 – 3200 TIP3P water molecules[16] and neutralized with Na^+ and Cl^- ions by using standard parameters for the Cornell et al. force field. These ions were placed in the most favorable locations by using coulombic potential terms with the LEAP[39] module. All simulations were performed by using periodic boundary conditions and the particle mesh Ewald approach[25] to introduce long-range electrostatic effects. The SHAKE[40] algorithm for hydrogen atoms, which allows the use of a 2-fs time step, was employed. Finally, a 9-Å cutoff was applied to Lennard – Jones interactions.

Equilibration of the system was carried out as follows: First, a short minimization with positional restraints on protein atoms was run to remove any potentially bad contact. The force constant for the positional constraints was 500 Kcal mol^{-1} Å^{-1}. We then ran a 12.5-ps molecular dynamics calculation at 300 K, with the positional restraints on the polypeptide maintained, in order to equilibrate the water box and ions. A 9 Å cut-off was used for the treatment of the electrostatic interactions in these two steps. Next, the system was equilibrated by using the mesh Ewald method, as water properties are slightly different with this treatment; that is, density, average water – water energy, and water diffusion values are slightly lower.[41] For this reason, a short MD simulation (12.5 ps) was run at 300 K, also with the Ewald approach, for long-range electrostatic effects. The system was subjected to several minimization cycles (each with 1000 steepest descent iterations) to gradually reduce positional restraints on the protein molecule from 500 kcal mol^{-1} Å^{-1} to 0. Finally, unrestrained 5-ns MD trajectories at constant pressure (1 atm) and temperature (300 K) were collected and analyzed with the Carnal program.[42]

CD experiments: Circular dichroism spectra were recorded on a JASCO J-810 spectropolarimeter fitted with a peltier temperature control accessory. Far-UV spectra were recorded in 0.1-cm path length quartz cells at protein concentrations of 0.1 – 0.3 mg mL^{-1} . The CD spectrum of the buffer alone was subtracted from the experimental spectra. Final spectra were the average of 4 – 6 runs.

DNA oligonucleotides were obtained from the oligonucleotide synthesis service at CSIC. Complementary strands were hybridized in equimolar concentrations before each binding experiment by heating to 90 °C for 10 min in sodium phosphate (10 mM)/sodium cloride (200 mM) at pH 6.3, followed by slow cooling to room temperature and dyalisis in the experimental buffer. Complete hybridization of the strands was verified by NMR spectroscopy.

Comparison of circular dichroism spectra of the protein, DNA, and protein – DNA complex species revealed a significant conformational change in the protein upon binding to DNA. Thus, difference spectra were calculated by subtraction of the spectrum of the duplex DNA alone from the spectra of the protein/DNA mixtures. The experiments were performed at a protein concentration of 10 uM and protein:DNA ratios of 1:1 to 1:2, in sodium phosphate (10 mM) at pH 6.3 and 303 K.

Thermal denaturation experiments were carried out by increasing the temperature from 283 to 348 K at two different rates (20 and 30 degrees h^{-1}) and monitoring changes in ellipticity at 222 nm. In addition, far-UV CD spectra were recorded at temperatures of 283 – 348 K with temperature increments of 3 degrees. The temperature was allowed to equilibrate for 5 min before each spectrum was acquired. The obtained denaturation curves were analyzed by fitting the experimental data to a two-state model.[43]

DNA CD titrations: Binding of the wild-type protein and both reduced and oxidized mutants to DNA was analyzed by monitoring protein ellipticity variations at 220 – 225 nm. The DNA contribution to each individual point was subtracted from the spectra of the corresponding protein/DNA mixtures. The curves were fitted assuming 2:1 stochiometry in the protein – DNA complex and noncooperativity (that is, assuming two independent and equivalent binding sites in the oligonucleotide) to give apparent K_a values for the palindromic P3 and P5 DNA fragments. In contrast, a 1:1 protein:DNA stochiometry was assumed for the half-site DNA. All experiments were carried out at a protein concentration of 15 µM in NaCl (200 mM), $MgCl_2$ (20 mM), and sodium phosphate (10 mM) at pH 6.3 and 303 K.

Financial support from DGES (grant nos. BQU-2000 – 1501-C01 and BQU2001 – 3693-C02 – 02) and from the Comunidad de Madrid (grant no. 08.5/0031.1/99) is gratefully acknowledged. We thank Dr. M. Gairi for the 800 MHz NMR measurements. Access to the 800 MHz facility of the Unitat de RMN d'alt Camp of the Universitat de Barcelona is also acknowledged.

[1] C. Cillo, A. Faiella, M. Cantile, E. Boncinelli, *Exp. Cell Res.* **1999**, *248*, 1.
[2] K. S. Koeneman, F. Yeung, L. W. Chung, *Prostate* **1999**, *39*, 246.
[3] A. S. Tucker, A. Al Khamis, P. T. Sharpe, *Dev. Dyn.* **1998**, *212*, 533.
[4] M. T. Dattani, J. P. Martinez-Barbera, P. Q. Thomas, J. M. Brickman, R. Gupta, I. L. Martensson, H. Toresson, M. Fox, J. K. Wales, P. C. Hindmarsh, S. Krauss, R. S. Beddington, I. C. Robinson, *Nat. Genet.* **1998**, *19*, 125.
[5] E. Hermesz, S. Mackem, K. A. Mahon, *Development* **1996**, *122*, 41.
[6] O. V. Kazanskaya, E. A. Severtzova, K. A. Barth, G. V. Ermakova, S. A. Lukyanov, A. O. Benyumov, M. Pannese, E. Boncinelli, S. W. Wilson, A. G. Zaraisky, *Genes* **1997**, *200*, 25.
[7] D. L. Reeves, *Bull. Johns Hopkins Hosp.* **1941**, *69*, 61.
[8] C. G. D. Brook, *Br. Med. J.* **1972**, *3*, 811.
[9] U. Roessmann, *J. Clin. Neuro-ophthalmol* **1989**, *9*, 156.
[10] D. S. Wilson, B. Guenther, C. Desplan, J. Kuriyan, *Cell* **1995**, *82*, 709.
[11] C. R. Kissinger, B. S. Liu, E. Martin-Blanco, T. B. Kornberg, C. O. Pabo, *Cell* **1990**, *63*, 579.
[12] C. Benassayaq, *Mech. Dev.* **1997**, *63*, 187.
[13] A. S. Fortin, D. A. Underhill, P. Gros, *Hum. Mol. Genet.* **1997**, *6*, 1781.
[14] P. Günter, C. Mumenthaler, K. Wüthrich, *J. Mol. Biol.* **1997**, *273*, 283.
[15] P. A. Kollman, S. J. Weiner, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, P. J. Weiner, *J. Am. Chem. Soc.* **1984**, *106*, 765.
[16] W. D. Cornell, P. C. Cieplack, I. Bayly, I. R. Gould, K. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179.
[17] C. Wolberger, *Curr. Opin. Struct. Biol.* **1996**, *6*, 62.
[18] M. R. R. Wintjens, *J. Mol. Biol.* **1996**, *262*, 294.
[19] Y. Q. Qian, K. Furukubo-Tokunaga, D. Resendez-Perez, M. Muller, W. J. Gehring, K. Wüthrich, *J. Mol. Biol.* **1994**, *238*, 333.
[20] D. H. Tsao, J. M. Gruschus, L. H. Wang, M. Nirenberg, J. A. Ferretti, *J. Mol. Biol.* **1995**, *251*, 297.
[21] M. Sivaraja, M. C. Botfield, M. Mueller, A. Jancso, M. A. Weiss, *Biochemistry* **1994**, *33*, 9845.
[22] P. Guntert, Y. Q. Qian, G. Otting, M. Muller, W. Gehring, K. Wüthrich, *J. Mol. Biol.* **1991**, *217*, 531.
[23] J. H. Carra, P. L. Privalov, *Biochemistry* **1997**, *36*, 526.
[24] T. Li, M. R. Stark, A. D. Johnson, C. Wolberger, *Science* **1995**, *270*, 262.
[25] C. S. A. T. Darden, *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155.
[26] W. L. Jorgensen, J. Chandresekhar, J. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926.
[27] J. M. Sánchez-Ruiz, *Subcell Biochem.* **1995**, *24*, 133.
[28] E. Freire, W. W. van Osdoll, O. Mayorga, J. M. Sanchez-Ruiz, *Annu. Rev. Biophys. Chem.* **1990**, *19*, 159.
[29] M. W. Sornson, W. Wu, J. S. Dasen, S. E. Flynn, D. J. Norman, S. M. O'Connell, I. Gukovsky, C. Carriere, A. K. Ryan, A. P. Miller, L. Zuo, A. S. Gleiberman, B. Andersen, W. G. Beamer, M. G. Rosenfeld, *Nature* **1996**, *384*, 327.
[30] D. H. Tsao, J. M. Gruschus, L. H. Wang, M. Nirenberg, J. A. Ferretti, *Biochemistry* **1994**, *33*, 15053.
[31] W. J. Gehring, Y. Q. Qian, M. Billeter, K. Furukubo-Tokunaga, A. F. Schier, D. Resendez-Perez, M. Affolter, G. Otting, K. Wüthrich, *Cell* **1994**, *78*, 211.

[32] N. P. Johnson, J. Lindstrom, W. A. Baase, P. H. von Hippel, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 4840.
[33] D. Wilson, G. Sheng, T. Lecuit, N. Dostatni, C. Desplan, *Genes Dev.* **1993**, *7*, 2120.
[34] D. Marion, M. Ikura, R. Tschudin, A. Bax, *J. Magn. Reson.* **1989**, *85*, 393.
[35] A. Bax, D. G. Davis, *J. Magn. Reson.* **1985**, *65*, 355.
[36] A. Kumar, R. R. Ernst, K. Wüthrich, *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1.
[37] D. Marion, K. Wüthrich, *Biochem. Biophys. Res. Commun.* **1983**, *113*, 967.
[38] R. M. Scheek, W. F. van Gunsteren, R. Kaptein, *Methods Enzymol.* **1989**, *177*, 204.
[39] C. E. A. F. Schafmeister, W. F. Ross, V. Romanovsky, University of California, San Francisco, **1995**.
[40] J. P. Rykaert, G. Ciccote, J. C. Berendsen, *J. Comp. Phys.* **1977**, *23*, 327.
[41] Essmann, *J. Chem. Phys.* **1995**, *103*, 8577.
[42] W. S. Ross, Carnal: Coordinate Analysis Program, University of California, San Francisco, Department of Pharmaceutical Chemistry.
[43] M. P. Irun, M. M. García-Mira, J. M. Sanchez-Ruiz, J. Sancho, *J. Mol. Biol.* **2001**, *306*, 877.

Received: January 28, 2002
Revised: May 27, 2002 [F 352]