

# Role of Conserved Salt Bridges in Homeodomain Stability and DNA Binding\*<sup>§</sup>

Received for publication, April 23, 2009, and in revised form, June 25, 2009. Published, JBC Papers in Press, June 26, 2009, DOI 10.1074/jbc.M109.012054

Mario Torrado<sup>†1</sup>, Julia Revuelta<sup>‡</sup>, Carlos Gonzalez<sup>§</sup>, Francisco Corzana<sup>¶2</sup>, Agatha Bastida<sup>‡</sup>, and Juan Luis Asensio<sup>‡3</sup>

From the <sup>‡</sup>Departamento de Química Orgánica Biológica, Instituto de Química Orgánica General, Consejo Superior de Investigaciones Científicas, 28006 Madrid, <sup>§</sup>Grupo de RMN, Instituto de Química-Física "Rocasolano," Consejo Superior de Investigaciones Científicas, 28006 Madrid, and the <sup>¶</sup>Departamento de Química, Universidad de la Rioja, Unidad Asociada Consejo Superior de Investigaciones Científicas, E-26006 Logroño, Spain

The sequence information available for homeodomains reveals that salt bridges connecting pairs 19/30, 31/42, and 17/52 are frequent, whereas aliphatic residues at these sites are rare and mainly restricted to proteins from homeotherms. We have analyzed the influence of salt and hydrophobic bridges at these sites on the stability and DNA binding properties of human Hesx-1 homeodomain. Regarding the protein stability, our analysis shows that hydrophobic side chains are clearly preferred at positions 19/30 and 31/42. This stabilizing influence results from the more favorable packing of the aliphatic side chains with the protein core, as illustrated by the three-dimensional solution structure of a thermostable variant, herein reported. In contrast only polar side chains seem to be tolerated at positions 17/52. Interestingly, despite the significant influence of pairs 19/30 and 31/42 on the stability of the homeodomain, their effect on DNA binding ranges from modest to negligible. The observed lack of correlation between binding strength and conformational stability in the analyzed variants suggests that salt/hydrophobic bridges at these specific positions might have been employed by evolution to independently modulate both properties.

Homeodomain proteins are transcription factors present in all eukaryotes and play key roles in cellular differentiation during development (1, 2). The homeodomain encodes a 60-residue DNA binding domain composed of a disordered N-terminal arm, three helical segments (herein referred as I, II, and III) connected by short loops, and a disordered C-terminal region. The secondary and tertiary structures are stabilized by the formation of a well defined hydrophobic core, which results from the packing of the three helices (3, 4).

\* This work was supported by Spanish "Plan Nacional" (Ministerio de Ciencia y Tecnología) Research Grants CTQ2007-67403/BQU and CTQ2004-04994/BQU.

The atomic coordinates and structure factors (code 2k40) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

<sup>§</sup> The on-line version of this article (available at <http://www.jbc.org/>) contains supplemental Tables S1 and S2 and Figs. S1–S13.

<sup>1</sup> Recipient of a Ministerio de Educación y Ciencia Formación de Profesorado Universitario fellowship.

<sup>2</sup> Recipient of a Ramón y Cajal contract.

<sup>3</sup> To whom correspondence should be addressed: Dept. de Química Orgánica Biológica, Instituto de Química Orgánica General, Consejo Superior de Investigaciones Científicas, Juan de la Cierva 3, 28006 Madrid, Spain. Tel.: 34-91-561-8806 (Ext. 329); Fax: 34-91-564-4853; E-mail: iqoa110@iqog.csic.es.

Presently, the largest searchable collection of information for the homeodomain protein family is the Homeodomain Resource Data Base (5–7). It contains around 1056 full-length protein sequences isolated from 112 different species. Inspection of these data shows that helices I/II, I/III, and II/III can be connected by three salt bridges involving pairs 19/30, 31/42, and 17/52, respectively, which in addition can exhibit different polarities. In contrast, hydrophobic bridges at these sites are present only in a minor fraction of sequences (and mainly restricted to homeodomains from warm-blooded animals). This predominance of salt *versus* hydrophobic bridges suggests a role for the former in homeodomain function and/or stability.

Here we analyze the relative effect of salt and hydrophobic bridges on the homeodomain stability and DNA binding properties, employing the human Hesx-1 DNA binding domain as model system. Our approach involves the extensive use of site-directed mutagenesis experiments together with CD, NMR, and isothermal titration microcalorimetry (ITC)<sup>4</sup> measurements. Thus, as a first step, the hydrophobic pair Val<sup>19</sup>/Ile<sup>30</sup>, present in the wild-type polypeptide, was replaced by salt bridges of different polarities, either isolated or networked. A previous statistical analysis of the information stored in the homeodomain resource data base (8, 9) revealed an intriguing correlation between the nature of pair 19/30 and the loop residue 26 (proline or a branched aliphatic amino acid in all cases). Considering this, both salt bridges and aliphatic residues connecting 19/30 were also considered in the context of either proline or leucine in 26. Finally, conserved salt bridges 31/42 and 17/52 were substituted by hydrophobic pairs observed in other natural sequences.

The work reported here implied the production of 33 homeodomain variants, including single, double, triple, and quadruple mutations. All of them were subjected to careful thermal and chemical denaturation experiments, employing circular dichroism. This analysis allowed the identification of a thermophilic variant of the Hesx-1 homeodomain. Its three-dimensional solution structure was determined by NMR methods to provide a detailed structural framework for elucidating the origin of the enhanced thermal stability. In addition, the pK<sub>a</sub> values for carboxyl groups in some selected variants were measured.

<sup>4</sup> The abbreviations used are: ITC, isothermal titration microcalorimetry; NOESY, nuclear Overhauser effect spectroscopy; DTT, dithiothreitol; DQF-COSY, double quantum-filtered correlation spectroscopy; PDB, Protein Data Bank; GdmCl, guanidinium chloride; TOCSY, total correlation spectroscopy.

## Salt Versus Hydrophobic Bridges in Stability and DNA Binding

ured by NMR spectroscopy to further test the role of conserved salt bridges on the homeodomain stability. As a final step, the DNA binding properties of selected mutants, together with their temperature dependence, were analyzed employing microcalorimetry. The results obtained provide new insights into the sequence, structure, stability, and function relationships within this important family of DNA-binding proteins.

### EXPERIMENTAL PROCEDURES

**Site-directed Mutagenesis of *Hesx-1***—Mutations were incorporated into the wild-type *Hesx-1* homeobox located on plasmid pT7-7 using the QuikChange mutagenesis kit (Stratagene, La Jolla, CA) and the appropriate mutagenic primers. The mutant plasmids were transformed into  $\text{CaCl}_2$ -competent *Escherichia coli* DH5 $\alpha$  cells. The sequences of all constructs were confirmed by DNA sequencing.

**Protein Expression and Purification**—Transformed *E. coli* BL21(DE3) cells were grown to  $A_{600} = 0.6$  at 37 °C in LB medium. Then cultures were induced with 0.5 mM isopropyl 1-thio- $\beta$ -D-galactopyranoside. Induction times and temperatures were optimized for every mutant being typically 13–24 h at 22–28 °C. Cells were harvested and lysed followed by clarification and removal of nucleic acids by polyethyleneimine precipitation. The nucleic acid-free extracts were dialyzed against 50 mM sodium phosphate buffer, pH 7.5, 2 mM  $\beta$ -mercaptoethanol (Buffer A), and then loaded onto Bio-Rex A-70 (Bio-Rad) column equilibrated in the same buffer. The column was extensively washed with Buffer A, and then the protein was eluted with a linear gradient of 0–300 mM NaCl. As the next step, ammonium sulfate precipitation was employed to further purify and concentrate the homeodomain variants. Ammonium sulfate was increased stepwise from 45% saturation to 100%, and the precipitated protein was separated by centrifugation (20 min at 12,000  $\times$  g) at every stage. The precipitated protein obtained in the last step (100% saturation of ammonium sulfate) was collected, resuspended in 5 ml of 20 mM sodium phosphate buffer, 100 mM NaCl, 2 mM  $\beta$ -mercaptoethanol, pH 7.0 (Buffer B), and loaded onto FPLC Superdex-75 HiLoad 26/60 (Amersham Biosciences) column previously equilibrated in the same buffer. The protein was then eluted employing a flow rate of 2.5 ml/min. The pure homeodomain fractions were pooled and desalted by dialysis against 2 mM sodium phosphate, 5 mM NaCl, 2 mM  $\beta$ -mercaptoethanol, pH 6.0, and then lyophilized. At the end of the procedure, the purity of the proteins was shown to be greater than 95% by SDS-PAGE. The protein concentrations were determined spectrophotometrically.

**CD Spectroscopy**—Circular dichroism experiments were recorded in a Jasco J-810 spectropolarimeter, fitted with a Peltier temperature control accessory.

GdmCl and urea denaturation data were obtained at 20 °C from samples containing 18  $\mu\text{M}$  protein in 10 mM phosphate, 100 mM NaCl, 0.5 mM DTT, pH 6, employing an autotitrator and 1-cm path length quartz cells provided with a magnetic stirrer. Urea and GdmCl stock solutions were also prepared in the same buffer. Chemical denaturations were followed by monitoring the CD ellipticity at 222 nm. Data were collected every 0.2 M from 0 to 9.4 M urea or from 0 to 8.5 M GdmCl. The

initial and final denaturant concentrations were confirmed by refractometry.  $\Delta G$  (the free energy of folding) and  $m$  values (the dependence of the free energy of folding on the denaturant concentration) were determined from the chemical denaturation data assuming a two-state transition and using the linear extrapolation method (10). Nonlinear regression calculations were performed employing Origin 5.0. The errors in the  $\Delta G$  values derived from the fitting procedure were in all cases smaller than 0.1 kcal/mol.

Thermal unfolding transitions were measured employing 1-cm path length quartz cells from samples containing 3  $\mu\text{M}$  protein in 10 mM phosphate, 0.5 mM DTT, pH 6, and five different NaCl concentrations (0, 0.1, 0.4, 1.0, and 2.0 M). The transitions were monitored by the decrease of the CD signal at 222 nm using a 2-nm bandwidth. Heating rates were 20 °C/h. Transitions were evaluated using a nonlinear least squares fit assuming a two-state model with sloping pre- and post-transitional base lines.  $\Delta G$  values at a common temperature (55 or 77 °C depending on the particular set of protein variants, see Table 1) were derived from the thermal denaturation data. In this procedure, a constant value of  $-0.7$  kcal/(mol·K) was used for the change in heat capacity upon folding ( $\Delta C_p$ ). It represents the average of all the values found in the analyses of the individual thermal transitions. In addition, it is consistent with the denaturant  $m$  values measured for the different protein variants and also with the decrease in the accessible surface area estimated for the homeodomain upon folding (11). The fact that  $m$  values exhibit very low variations for the different mutants (in all cases it is in the 0.8–1 kcal/(mol·M) range for denaturation experiments with urea and in the 1.4–1.7 kcal/(mol·M) range for those with guanidinium chloride) supports the employment of a common  $\Delta C_p$  for all of them (11). Comparable  $\Delta C_p$  values have been described for the Engrailed ( $-0.7$  kcal/(mol·K)) (9) and VndK2 ( $-0.52$  kcal/(mol·K)) (12) homeodomains. In addition, it has to be considered that most variants showed denaturation midpoints relatively close to 55 or 77 °C (depending on the set of mutants), and therefore large extrapolations were not necessary for determining  $\Delta G$  at these temperatures.

The standard errors for  $T_m$  and  $\Delta H$  determined from the analysis of the individual melting profiles were typically not higher than 0.1 °C and 1 kcal/mol, respectively. The accuracy in the  $\Delta G$  values at 55 or 77 °C is obviously correlated with the  $T_m$ . To estimate the errors in the folding free energies, their values were also derived considering a variation of 2 kcal/mol in  $\Delta H$  ( $\Delta H \pm 2$  kcal/mol), 0.2 °C in  $T_m$  ( $T_m \pm 0.2$  °C), and 0.2 kcal/(mol K) in  $\Delta C_p$  ( $\Delta C_p \pm 0.2$  kcal/(mol K)). In every case, the maximum difference between the calculated free energies was taken as the error (see supplemental Fig. S1).

Finally, the contributions to stability still present in 2 M NaCl were assumed to reflect primarily packing and hydrophobic interactions (herein referred as  $\Delta G_{2\text{ M NaCl}}$ ), whereas those that could be screened by adding 2 M NaCl were attributed to coulombic interactions ( $\Delta G_{\text{coulomb}}$ ). They were calculated from the difference  $\Delta G_{\text{coulomb}} = \Delta G_{0\text{ M NaCl}} - \Delta G_{2\text{ M NaCl}}$  and the corresponding errors estimated from  $E_{\Delta G_{\text{coulomb}}} = ((E_{\Delta G_{0\text{ M NaCl}}})^2 + (E_{\Delta G_{2\text{ M NaCl}}})^2)^{1/2}$ .

**NMR Structure Determination of the Thermostable Variant R31L/E42L**— $^1\text{H}$  NMR spectra were recorded in 85:15  $^1\text{H}_2\text{O}$ : $^2\text{H}_2\text{O}$  and in  $^2\text{H}_2\text{O}$  on Bruker Avance 800, Bruker Avance 600, and Varian Unity 500-MHz spectrometers. NMR samples included 400–800  $\mu\text{M}$  protein concentration in 100 mM NaCl, 10 mM sodium phosphate buffer, and 2 mM DTT, pH 6.0. Protein assignments were obtained using a set of two-dimensional NOESY (13), TOCSY (14) and DQF-COSY (15) experiments. The TOCSY, NOESY, and DQF-COSY experiments were carried out in the phase-sensitive mode using the TPPI method (15) for quadrature detection in the indirect dimension. Typically, a data matrix of 512\*2048 points was used to digitize a spectral width of 8000–6000 Hz. 80 scans were used per increment with a relaxation delay of 1 s. Prior to Fourier transformation, zero filling was used in the indirect dimension to expand the data to 1K\*2K. Base-line correction was applied in both dimensions. The TOCSY spectra were acquired using 60 ms of isotropic mixing period. The NOESY experiments were performed with mixing times from 50 to 200 ms.

Upper limits for proton-proton distances were obtained from NOESY cross-peak intensities at three mixing times, 50, 150, and 200 ms. Cross-peaks were classified as strong, medium, and weak corresponding to upper limits of 2.5, 3.5, and 5.0 Å. 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 using the program DYANA (16). A set of 703 constraints was used in the final round of calculations.

The 30 best DYANA structures in terms of target function were submitted to a simulated annealing protocol (17) with the AMBER 5.0 package and the parameters described by Kollman and co-workers (18). To prevent nonrealistic interactions between disordered regions of the protein and the structured helical core, explicit TIP3P (18) water molecules and periodic boundary conditions were employed in these calculations (19).

**$pK_a$  Measurements**—NMR experiments were performed at 5 °C on a Bruker Avance 800 spectrometer in 85:15  $^1\text{H}_2\text{O}$ : $^2\text{H}_2\text{O}$ . NMR samples were prepared in 100 mM NaCl, 10 mM sodium phosphate, and 2 mM DTT at 400–600  $\mu\text{M}$  protein concentrations. Resonances for all the glutamate and aspartate side chains were assigned employing a set of two-dimensional NOESY (13), TOCSY (14), and DQF-COSY (15) experiments. Changes in chemical shift for the side-chain  $\text{H}\gamma/\text{H}\beta$  protons in glutamates and  $\text{H}\beta$  protons in aspartates were monitored as a function of pH. The  $pK$  values were obtained from nonlinear least squares fitting employing the Henderson-Hasselbach equation with Hill coefficients set to 1.

**DNA Binding Experiments**—Binding studies were performed at 15, 25, and 35 °C, in 150 mM NaCl, 20 mM phosphate, 5 mM  $\text{MgCl}_2$ , pH 6.0, using a VP-ITC titration calorimeter (MicroCal, LLC) with a reaction cell volume of 1.467 ml. Both the protein and duplex DNA (5'-GTCTAATTGACGCG-3' and its complementary 5'-CGCGTCAATTAGAC-3') solutions were dialyzed against the same buffer prior to ITC experiments to ensure chemical equilibration. Typically, 4.0–14.0  $\mu\text{M}$  duplex DNA in the reaction cell was titrated with a 100–200  $\mu\text{M}$  solution of the different Hesx-1 variants contained in a 300- $\mu\text{l}$  syringe. At least 30 consecutive injections of 5–10  $\mu\text{l}$  were

applied at 5-min intervals, whereas the DNA solution was stirred at a constant speed of 300 rpm. Dilution heats of protein into DNA solutions (which agreed with those obtained by injections of proteins into the same volume of buffer) were subtracted from measured heats of binding. Titration curves were analyzed with Origin, provided with the instrument by MicroCal LLC, using a one-site binding model to fit the curves. For every single protein variant, thermodynamic parameters were derived from two independent experiments and averaged.

## RESULTS

**Occurrence of Salt Bridges within the Homeodomain Family**—As a first step, the relative occurrence of both salt bridges and hydrophobic pairs at positions 19/30, 31/42, and 17/52 was determined from the information stored in the Homeodomain Resource Data Base (1056 nonredundant full-length and domain sequences). For our analysis, a data set of 754 proteins, including the canonical 60-residue homeodomain sequence with no insertions or deletions, was considered. The results obtained are represented schematically in Fig. 1, *a–c*. It can be observed that positions 19/30 are connected by salt bridge interactions of both polarities in 52.5% of the cases (396 sequences) with a clear preference for those combinations with a positively charged residue in 30 (Fig. 1*a*). In contrast, aliphatic residues at these sites are relatively rare (2.5%).

In some cases, residues 19/30 can establish a cooperative network with positions 15/37, as that observed in the x-ray structure of the Engrailed homeodomain (9, 20). An equivalent pattern of interactions involving pairs 19/30 and 15/37 would be feasible in only 2.7% of the sequences stored in the Homeodomain Resource Data Bank (see Fig. 1*b*). In fact, an alternative network involving the triad 19/30/33 (typically Glu<sup>19</sup>/Arg<sup>30</sup>/Glu<sup>33</sup>) seems to be much more frequent within this family of proteins.

From a statistical analysis of homeodomain sequences, Clarke (8) identified a dominating pattern of pairwise co-variation centered on residue 26. Using the co-varying network, homeodomains were divided into two classes. One class has branched aliphatic residues at position 26, whereas the second contains proline. Previous analysis of a reduced data set of representative human homeodomain sequences (9) found that the branched aliphatic subgroup usually has a salt bridge connecting residues 19 and 30 (92% of cases). Strikingly, none of the Pro<sup>26</sup> subgroup, with more than 60 members, had this potential interaction. Based on this observation, it was suggested that salt bridge 19/30 in presence of Leu<sup>26</sup> could contribute to the conformational variability of the loop between helices I and II. This could affect the ability of the homeodomain for induced fit on binding to DNA.

Our analysis of an extended data set with 754 sequences reveals a slightly more complex situation (Fig. 1*c*). For homeodomains with a salt bridge connecting 19/30, 26 can either be a branched aliphatic residue (usually leucine) or proline, depending on the salt bridge polarity. If the positive charge is at 30, 26 is a branched aliphatic amino acid in >96% of the cases. On the other hand, if position 30 is negatively charged, then 26 can be either leucine or proline, although leucine is clearly pre-

## Salt Versus Hydrophobic Bridges in Stability and DNA Binding

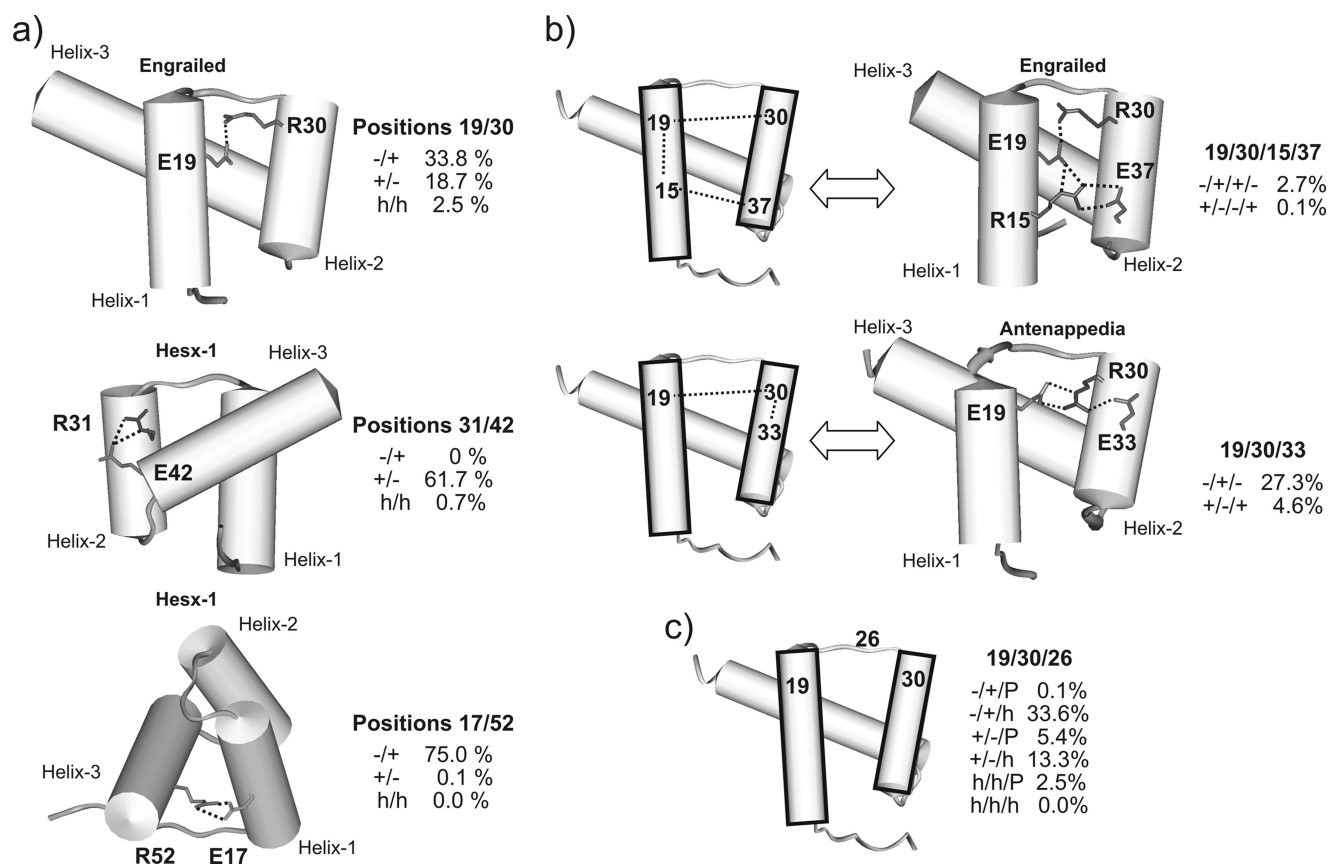


FIGURE 1. *a*, representation of the Engrailed (*upper panel*) and Hesx-1 (*middle and bottom panels*) homeodomain structures. Salt bridge interactions, conserved within this family of proteins, connecting positions 19/30 (*upper panel*), 31/42 (*middle panel*), and 17/52 (*bottom panel*) are indicated. In addition, frequencies for salt versus hydrophobic bridges at these sites, within a data set of 754 homeodomain sequences, are shown on the *right*. Symbols *h*, +, and – stand for aliphatic (Val, Leu, and Ile), positively (His, Arg, and Lys), and negatively (Glu and Asp) charged residues, respectively. *b*, schematic representation of the salt bridge networks involving residues 19/30 most commonly found in homeodomains. Populations for each pattern of interactions are represented on the *right*. *c*, observed correlation between positions 19/30 and the loop residue 26 in homeodomains (*h* stands for aliphatic and *P* stands for proline).

ferred. Interestingly, if residues 19 and 30 are aliphatic, 26 is always proline (see Fig. 1*c*).

Regarding positions 31/42, they form a salt bridge in the majority of homeodomains (61.7%). For a minor fraction of sequences, an aliphatic residue can be found either in 31 (14.7%) or 42 (3.8%). Finally, both are aliphatic in just five sequences (0.7%) of our data set (Fig. 1*a*).

The most conserved salt bridge interaction within the homeodomain family is that connecting positions 17/52, present in 75.1% of the sequences. Hydrophobic residues at either position, 17 or 52, can be found in 0.8 and 0.1% of the cases, respectively. However, a hydrophobic pair at this site is not present in any single protein within our data set (Fig. 1*a*). A selection of representative homeodomain structures with salt bridges 19/30, 31/42, and 17/52 is shown in [supplemental Fig. S2](#).

**Salt Versus Hydrophobic Bridges as Determinants of Homeodomain Stability, Methodological Aspects**—To analyze the relative efficacy of salt versus hydrophobic bridges at positions 19/30, 31/42, and 17/52 in the stabilization of the homeodomain fold, several single, double, triple, and quadruple mutants of Hesx-1 homeodomain were produced. The obtained variants (1-33 in Table 1) were subjected to both heat and chemically induced denaturation experiments.

To gain further insights into the different contributions to the observed changes in stability, thermal denaturation profiles

were collected at five different NaCl concentrations (0, 0.1, 0.4, 1.0, and 2.0 M), and the dependence of  $\Delta G$  with the ionic strength at a common temperature (55 or 77 °C, see Table 1) was analyzed.

Typical thermal and chemical denaturation profiles collected for Hesx-1 mutants are shown in Fig. 2*a* and [supplemental Figs. S3 and S4](#). It can be observed that the obtained variants are strongly stabilized by increasing NaCl concentrations. Moreover, in all cases,  $\Delta G$  exhibits a linear dependence with the square root of the ionic strength suggesting that the electrostatic screening of unfavorable interactions is the dominant mechanism for the observed salt stabilization (21).

Free energies derived at 0 and 2 M NaCl are referred as  $\Delta G_{0\text{ M NaCl}}$  and  $\Delta G_{2\text{ M NaCl}}$ , respectively. Following the methodology described by Schmid and co-workers (22–25), the contributions to stability still present in 2 M NaCl were assumed to reflect primarily packing and hydrophobic interactions. Other effects, such as desolvation penalties, the intrinsic helix propensities of the different residues, or their conformational entropies, might also contribute to this term. On the other hand, those contributions that could be screened by adding 2 M NaCl were attributed to coulombic interactions ( $\Delta G_{\text{coulomb}}$ ). Throughout this paper, residues relevant for the discussion in the different protein variants are indicated in parentheses.

TABLE 1

Stability data measured for wild-type Hesx-1 and variants 1–33 from chemical (left) and thermal (right) denaturation experiments (employing urea or guanidinium chloride as indicated)

Mutants are numbered according to their thermal stability at 0 M NaCl. For all variants a proline residue is present at the loop position 26 unless explicitly stated (in parentheses, variants 1 and 17).

Variant	Residues 19/30, 15/37	Chemical Denaturation		Thermal Denaturation				
		$\Delta G$ (kcal/mol) Urea	$m^a$	$T_m^b$ ( $^{\circ}C$ ) 0/2 M NaCl	$\Delta G^c$ (kcal/mol) at 55 $^{\circ}C$			$mI^d$
					0 M NaCl total	2 M NaCl non-polar	Coulombic	
1	V/I, V/K <sup>(L26)</sup>	-	-	-	-	-	-	-
2	D/R, V/K	-2.01	0.8	36.2/53.6	2.53±0.27	0.19±0.04	2.34±0.27	-1.7
3	R/I, E/R	-2.14	0.9	37.5/54.7	2.04±0.24	0.03±0.03	2.01±0.24	-1.5
4	R/I, V/K	-2.01	0.9	37.7/58.4	2.14±0.24	-0.39±0.04	2.53±0.24	-1.9
5	V/R, R/E	-2.52	0.8	38.3/62.8	2.18±0.23	-1.00±0.09	3.18±0.25	-2.3
6	E/I, R/E	-3.18	0.8	41.0/60.8	1.77±0.18	-0.76±0.07	2.53±0.19	-1.9
7	E/R, R/K	-2.54	0.9	41.2/58.6	1.89±0.18	-0.50±0.05	2.39±0.19	-1.7
8	V/R, V/K	-2.59	0.8	41.5/64.3	1.50±0.17	-1.01±0.10	2.51±0.20	-1.8
9	E/R, V/E	-3.39	0.8	41.8/64.8	2.18±0.18	-1.44±0.11	3.62±0.21	-2.6
10	V/I, V/E	-3.84	0.8	43.7/69.9	1.93±0.15	-1.63±0.17	3.56±0.23	-2.7
11	R/E, E/K	-3.12	0.9	43.7/58.7	1.50±0.14	-0.46±0.05	1.96±0.15	-1.4
12	R/E, E/R	-3.14	0.9	44.0/58.5	1.45±0.14	-0.43±0.05	1.88±0.15	-1.4
13	R/E, V/R	-3.23	0.9	44.3/62.2	1.39±0.14	-0.92±0.08	2.31±0.16	-1.7
14	E/I, V/K	-3.23	0.8	44.8/63.5	1.33±0.13	-1.05±0.09	2.38±0.16	-1.7
15	E/R, R/E	-3.57	0.9	45.2/62.8	1.45±0.13	-1.06±0.09	2.51±0.16	-1.8
16	R/E, V/K	-3.32	0.9	46.0/63.1	1.28±0.12	-1.09±0.09	2.37±0.15	-1.7
17	E/R, V/K <sup>(L26)</sup>	-3.53	0.8	46.3/65.0	1.11±0.11	-1.80±0.12	2.91±0.16	-2.1
18	V/I, R/K	-3.56	1.0	47.1/64.3	1.02±0.10	-1.05±0.10	2.07±0.14	-1.5
19	E/R, V/K	-3.69	0.9	48.4/66.6	0.89±0.09	-1.50±0.13	2.39±0.16	-1.7
20	V/I, R/E	-3.80	0.8	48.6/67.9	0.88±0.09	-1.51±0.15	2.39±0.17	-1.7
21	V/I, V/L	-4.60	0.8	48.9/70.8	0.64±0.07	-1.46±0.18	2.10±0.19	-1.5
22	V/E, E/R	-3.75	0.8	49.2/66.6	0.77±0.08	-1.43±0.13	2.20±0.15	-1.6
23	V/I, E/K	-4.17	0.9	52.3/68.8	0.33±0.05	-1.70±0.16	2.03±0.17	-1.4
24	V/I, E/R	-4.38	0.9	53.1/68.9	0.23±0.04	-1.53±0.16	1.76±0.17	-1.3
25	V/E, V/K	-4.20	0.9	53.2/70.2	0.23±0.04	-1.83±0.17	2.06±0.17	-1.5
26	V/I, V/R	-5.11	1.0	56.1/72.9	-0.13±0.03	-2.05±0.22	1.92±0.22	-1.4
wt	V/I, V/K	-4.43	0.9	56.3/72.0	-0.16±0.03	-1.94±0.20	1.78±0.20	-1.3
Variant	Residues 31/42	$\Delta G$ (kcal/mol) Urea	$m^a$	$T_m^b$ ( $^{\circ}C$ ) 0/2 M NaCl	$\Delta G^c$ (kcal/mol) at 55 $^{\circ}C$			$mI^d$
27	R/D	-3.70	0.9	50.1/66.1	0.56±0.06	-1.28±0.12	1.84±0.13	-1.3
wt	R/E	-4.43	0.9	56.3/72.0	-0.16±0.03	-1.94±0.20	1.78±0.20	-1.3
Variant	Residues 31/42	$\Delta G$ (kcal/mol) GdmCl	$m^a$	$T_m^b$ ( $^{\circ}C$ ) 0/2 M NaCl	$\Delta G^c$ (kcal/mol) at 77 $^{\circ}C$			$mI^d$
wt	R/E	-4.71	1.5	56.3/72.0	3.00±0.29	0.68±0.07	2.32±0.30	-1.7
28	R/L	-5.03	1.4	58.2/75.8	2.57±0.25	0.21±0.04	2.36±0.25	-1.7
29	L/E	-6.52	1.5	71.7/85.2	0.68±0.07	-1.04±0.09	1.72±0.11	-1.3
30	I/L	-8.44	1.7	82.8/91.6	-0.75±0.07	-1.91±0.16	1.16±0.17	-0.8
31	L/L	-9.48	1.6	88.7/95.9	-1.26±0.12	-2.43±0.22	1.17±0.25	-0.8
Variant	Residues 17/52	$\Delta G$ (kcal/mol) Urea	$m^a$	$T_m^b$ ( $^{\circ}C$ ) 0 M NaCl	$\Delta G^c$ (kcal/mol) at 55 $^{\circ}C$			$mI^d$
wt	E/R	-4.43	0.9	56.3	-0.16±0.03	-1.94±0.20	1.78±0.20	-1.3
32	L/Y	-	-	-	-	-	-	-
33	Y/Y	-	-	41.3	-	-	-	-

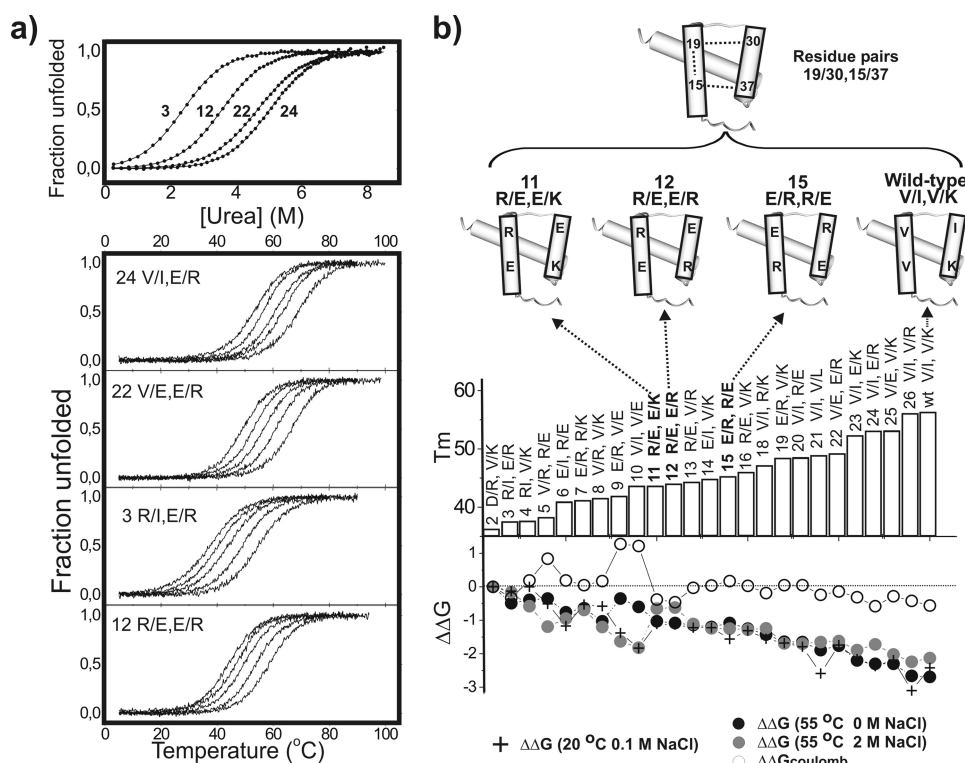
<sup>a</sup>  $m$  (kcal/(mol M)) slope of  $\Delta G$  versus denaturant concentration plots.

<sup>b</sup>  $T_m$  is the midpoint of the thermal unfolding transition. Values measured at 0 and at 2 M NaCl concentration are shown.

<sup>c</sup>  $\Delta G$  is the change in Gibbs free energy of folding at 55 or 77  $^{\circ}C$  (as indicated);  $\Delta G$  at 0 M NaCl is the total change,  $\Delta G$  at 2 M NaCl represents the nonpolar contribution, and the difference between 0 and 2 M NaCl represents the Coulombic contribution to  $\Delta G$ .

<sup>d</sup>  $mI$  (kcal/(mol M<sup>1/2</sup>)) slope of  $\Delta G$  versus  $[NaCl]^{1/2}$  determined as described by Rios and Plaxco (21). This parameter is proportional to the  $\Delta G_{\text{Coulomb}}$  term.

## Salt Versus Hydrophobic Bridges in Stability and DNA Binding



**FIGURE 2. Stability analysis of Hexs-1 mutants with different combinations of aliphatic and charged side chains at positions 19/30 and 15/37 (indicated for every mutant).** Residues at these positions in wild-type Hexs-1 are Val<sup>19</sup>/Ile<sup>30</sup> and Val<sup>15</sup>/Lys<sup>37</sup>. *a*, chemical denaturation curves measured for variants **24**, **22**, **3**, and **12** at 20 °C in 0.1 M NaCl, 10 mM phosphate, 0.5 mM DTT, pH 6.0. The corresponding thermal denaturation curves in presence of 0, 0.1, 0.4, 1, and 2 M NaCl are represented in the lower panel. In all cases, salt increases the thermal stability. *b*, thermodynamic parameters derived for the analyzed variants. Melting temperatures together with free energy increments (with respect to the more unstable mutant, **2**) derived from the chemical (at 20 °C) and thermal (at 55 °C in the presence of 0 and 2 M NaCl) denaturation experiments are represented. Differences in the  $\Delta G_{\text{coulomb}}$  term ( $\Delta\Delta G_{\text{coulomb}}$ ) are also represented. Those variants with a potential network of salt bridges involving 19/30 and 15/37, together with the wild-type protein, are schematically represented above.

**Salt Bridges Versus Hydrophobic Pairs at Positions 19/30—**First, the influence of pair 19/30 on the stability of Hexs-1 was analyzed. It is well established that formation of extensive cooperative networks can greatly enhance the stabilizing effect of salt bridges in proteins (26). Pair 19/30 has been shown to be involved in a polar network with residues 15/37 in the Engrailed homeodomain (Fig. 1*b*). Taking this into account, variants, including different combinations of charged/aliphatic side chains at these four positions, were considered. The obtained results are represented in Table 1 and summarized in Fig. 2*b* and supplemental Figs. S4–S7.

It can be observed that the single or double replacement of residues Val<sup>19</sup>/Ile<sup>30</sup> (present in the wild-type polypeptide) by charged side chains to build salt bridges of different polarities results, in all cases, in a clear destabilization of the homeodomain fold. Interestingly, this effect seems relatively independent on the context provided by pair 15/37. The salt dependence exhibited by  $\Delta G$  at 55 °C provides a possible explanation for this behavior. Thus, although the unfavorable influences of the mutations are, in most cases, slightly attenuated at 2 M NaCl (indicating a less optimized electrostatic balance in the mutants), to a large extent they are maintained at such ionic strength (see  $\Delta G_{0 \text{ M NaCl}}$ ,  $\Delta G_{2 \text{ M NaCl}}$ , and  $\Delta G_{\text{coulomb}}$  terms in Table 1 and Fig. 2*b*), suggesting that Val<sup>19</sup> and Ile<sup>30</sup> establish more optimized packing interactions with the protein hydro-

phobic core than the charged side chains. Even variants **11** (Arg<sup>19</sup>/Glu<sup>30</sup>, Glu<sup>15</sup>/Lys<sup>37</sup>), **12** (Arg<sup>19</sup>/Glu<sup>30</sup>, Glu<sup>15</sup>/Arg<sup>37</sup>), and **15** (Glu<sup>19</sup>/Arg<sup>30</sup>, Arg<sup>15</sup>/Glu<sup>37</sup>), in which a network of polar interactions involving positions 15, 19, 30, and 37 (similar to that observed in the Engrailed homeodomain) might be feasible, are severely destabilized with respect to the wild-type polypeptide, as a result of the unfavorable salt-independent contributions of the mutations.

Finally, the obtained results indicate a clear preference for the Glu<sup>19</sup>/Arg<sup>30</sup> combination. Indeed, variants **19** (Glu<sup>19</sup>/Arg<sup>30</sup>, Val<sup>15</sup>/Lys<sup>37</sup>) and **15** (Glu<sup>19</sup>/Arg<sup>30</sup>, Arg<sup>15</sup>/Glu<sup>37</sup>) are more stable than **16** (Arg<sup>19</sup>/Glu<sup>30</sup>, Val<sup>15</sup>/Lys<sup>37</sup>) and **12** (Arg<sup>19</sup>/Glu<sup>30</sup>, Glu<sup>15</sup>/Arg<sup>37</sup>), respectively (see Table 1). It is likely that this particular distribution of amino acids allows a better packing of the side-chain methylene groups with the protein hydrophobic core, and/or a more optimized interaction between position 30 and the loop connecting helices I and II. Interestingly, the observed preference is reflected in the relative abundance of both types of contacts in the

homeodomain resource data base (salt bridges of equivalent polarity to those present in mutants **19** and **16** are present in 33.8 and 18.7% of the sequences, respectively, see Fig. 1*a*). In conclusion, simple hydrophobic interactions established by aliphatic residues at positions 19 and 30 provide a more effective stabilization of the Hexs-1 fold than salt bridges of any polarity either isolated or networked.

**Interaction between Pair 19/30 and the Loop Residue 26, Influence on Homeodomain Stability—**As indicated under “Occurrence of Salt Bridges within the Homeodomain Family,” the homeodomain resource data base reveals an intriguing correlation between pair 19/30 and residue 26, located on the loop between helices I and II.

Wild-type Hexs-1 includes a proline at position 26, and residues 19/30 are both aliphatic (Val<sup>19</sup>/Ile<sup>30</sup>, see Fig. 3). This pair was replaced by a salt bridge Glu<sup>19</sup>/Arg<sup>30</sup> in the previously analyzed mutants **19** and **15**, resulting in a rather unusual combination (Glu<sup>19</sup>/Pro<sup>26</sup>/Arg<sup>30</sup>). To assess the relative effect of aliphatic versus salt bridges connecting sites 19/30 on the protein stability, in the context of both proline or leucine at position 26, two additional protein variants were produced and analyzed (Fig. 3*a*). First, the triple mutant **17** (Glu<sup>19</sup>/Leu<sup>26</sup>/Arg<sup>30</sup>), which contains the triad more frequently found in homeodomains, was obtained. Second, Pro<sup>26</sup> was replaced by a leucine in the

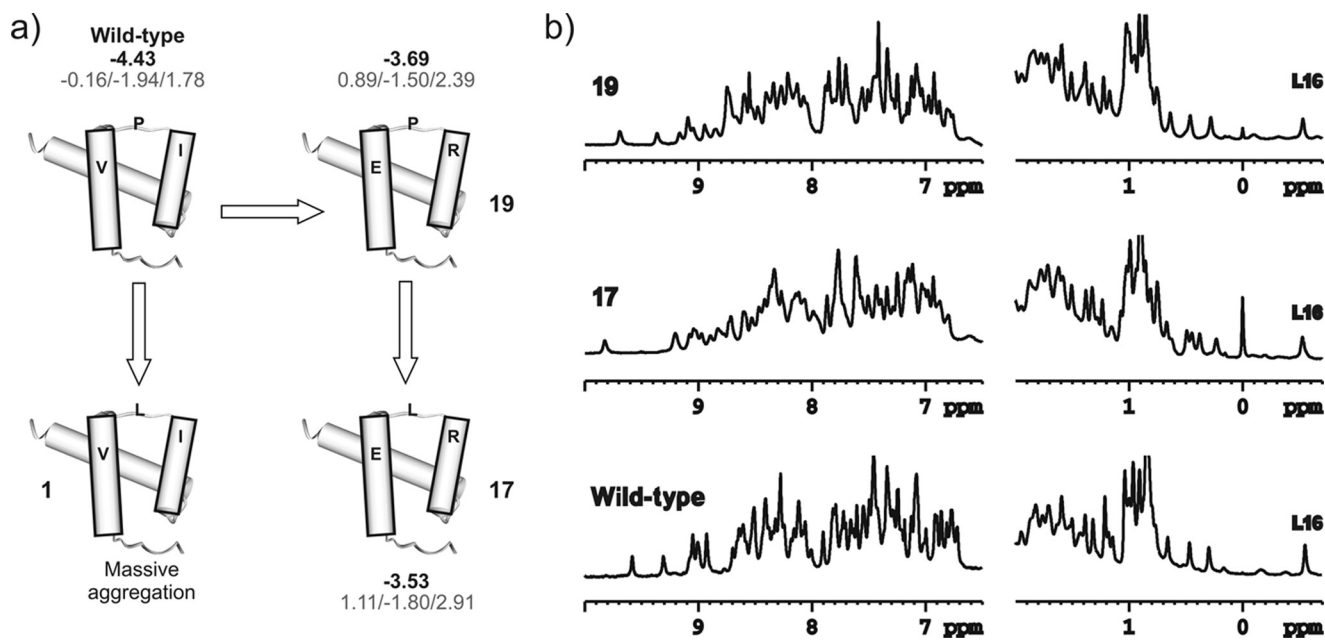


FIGURE 3. *a*, schematic representation of wild-type Hexs-1 together with variants 1, 19, and 17, characterized by different *natural* and *non-natural* combinations of residues at positions 19, 30, and 26.  $\Delta G$  values at 20 °C measured from chemical denaturation experiments are represented in *black*. In addition,  $\Delta G_{0, M NaCl} / \Delta G_{2, M NaCl} / \Delta G_{coulomb}$  terms at 55 °C, derived from thermal unfolding experiments, are shown in *gray*. For variant 1, stability could not be analyzed due to strong aggregation and precipitation of the protein. Free energy contributions are given in kcal/mol. *b*, one-dimensional NMR spectra measured for wild-type Hexs-1 and variants 17 and 19.

wild-type protein to produce mutant **1** characterized by the *unnatural* triad Val<sup>19</sup>/Leu<sup>26</sup>/Ile<sup>30</sup>.

As a preliminary analysis, high resolution one- and two-dimensional NMR spectra were collected for mutants **17** (Glu<sup>19</sup>/Leu<sup>26</sup>/Arg<sup>30</sup>) and **19** (Glu<sup>19</sup>/Pro<sup>26</sup>/Arg<sup>30</sup>) and compared with that corresponding to the wild-type polypeptide (Fig. 3*b*). All spectra exhibit sharp signals with a similar dispersion. Moreover, distinct chemical shifts and key long range NOEs, which are hallmarks of this family of proteins, could also be identified for the two mutants, which rule out the existence of major structural differences between the three homeodomains. Next, both **17** (Glu<sup>19</sup>/Leu<sup>26</sup>/Arg<sup>30</sup>) and **19** (Glu<sup>19</sup>/Pro<sup>26</sup>/Arg<sup>30</sup>) were subjected to denaturation studies. The results obtained are shown schematically in Fig. 3*a*. It can be observed that the P26L substitution in **19** (Glu<sup>19</sup>/Pro<sup>26</sup>/Arg<sup>30</sup>), to give **17** (Glu<sup>19</sup>/Leu<sup>26</sup>/Arg<sup>30</sup>), has a minor effect on the homeodomain stability. In fact, **17** (Glu<sup>19</sup>/Leu<sup>26</sup>/Arg<sup>30</sup>) is slightly less stable, according to both chemical and thermal denaturation experiments, mainly reflecting a less optimized electrostatic balance (see  $\Delta G_{coulomb}$  terms in Table 1). A speculative but plausible explanation for this result is that the P26L mutation induces minor conformational differences in the loop between helices I and II (a flexible region in homeodomains), which in turn might lead to a poorer electrostatic interaction between charged residues Glu<sup>19</sup> and Arg<sup>30</sup>.

A different behavior was observed for variant **1** (Val<sup>19</sup>/Leu<sup>26</sup>/Ile<sup>30</sup>). Strikingly, the P26L substitution in the wild-type protein strongly promoted the formation of inclusion bodies under all the expression conditions tested. In addition, the minor fraction of protein expressed, which was soluble, showed a strong tendency to aggregate and precipitate and was lost during the purification in all attempts. Therefore, this mutant could not be subjected to a detailed thermodynamic analysis. However, tak-

ing into account that residue 26 is not solvent-exposed in the native protein (and therefore it is not likely to significantly affect the protein solubility), this behavior suggests that the P26L mutation strongly destabilizes the protein. This conclusion might seem surprising considering that it appears conservative at first glance. However, it would be in agreement with previous studies on protein design (27–30). More specifically, it has been shown that although some variations in the buried positions of a protein are allowed, there are limits on the sequences that result in stable native-like folds (27–30). Taking this into account, specific packing between residues 19/26/30 seems critical for the stability of the homeodomain fold.

*Salt Bridges Versus Hydrophobic Pairs at Positions 31/42, Three-dimensional Solution Structure of a Thermostable Homeodomain Variant*—To determine the influence of pair 31/42 on the homeodomain structure and stability, additional single and double mutants were prepared and analyzed. The results obtained are schematically shown in Fig. 4*a*. It can be observed that the single replacement E42L (variant **28** (Arg<sup>31</sup>/Leu<sup>42</sup>)) slightly increases the protein stability (see Table 1 and Fig. 4*a*). This result was unexpected given the exposed location of residue 42 and considering that the E42L substitution disrupts the salt bridge present in the wild-type polypeptide (and observed in most homeodomain structures). A much larger stabilization (up to 2.32 kcal/mol at 77 °C) is achieved by the single mutation R31L (variant **29** (Leu<sup>31</sup>/Glu<sup>42</sup>)). Finally, the double substitution R31L/E42L leads to variant **31** (Leu<sup>31</sup>/Leu<sup>42</sup>), a thermophilic version of Hexs-1 homeodomain, stabilized more than 4 kcal/mol with respect to the wild-type polypeptide (see Table 1 and Fig. 4*a*) and characterized by melting temperatures in the 89–95 °C range. The origin of this effect can be understood by examining the thermodynamic data in detail. Thus, despite the disruption of the Arg<sup>31</sup>/Glu<sup>42</sup> interaction, replace-

## Salt Versus Hydrophobic Bridges in Stability and DNA Binding

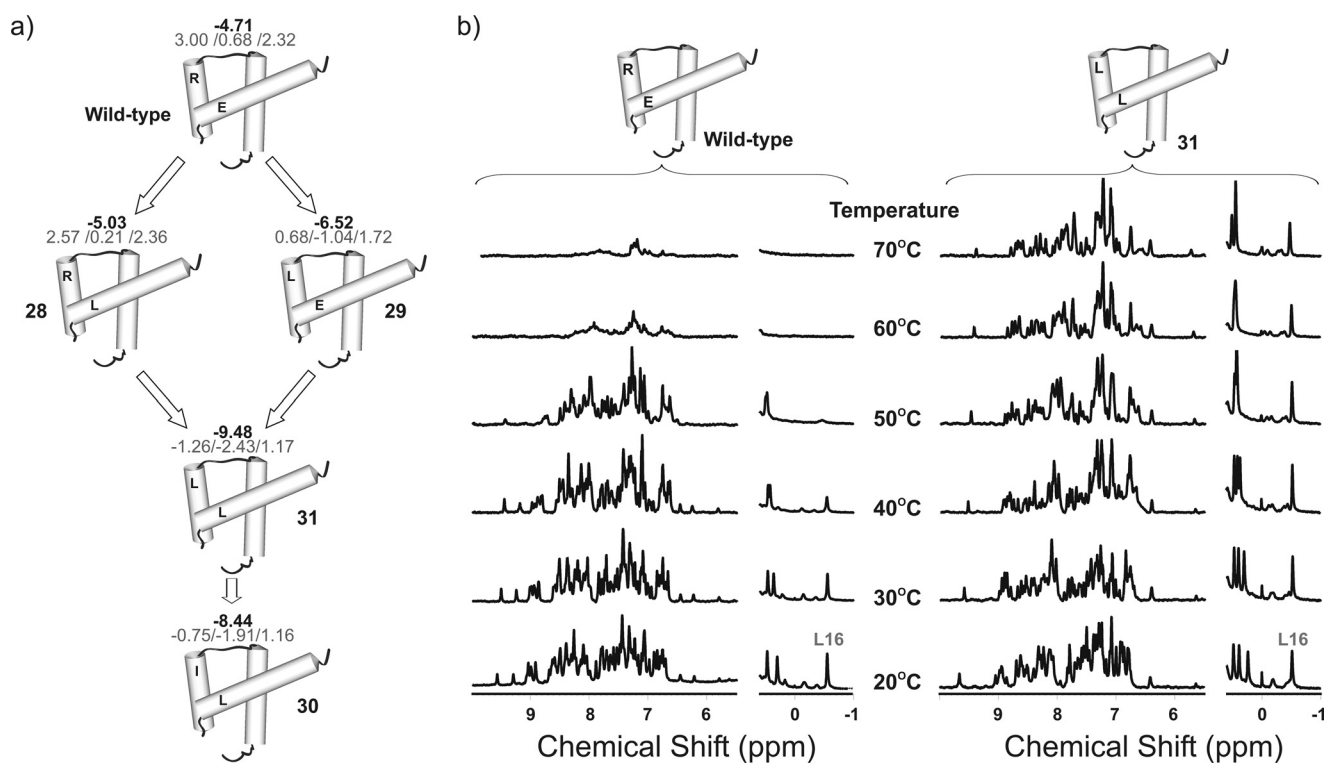


FIGURE 4. *a*, schematic representation of wild-type Hesx-1 together with variants 28, 29, 30, and 31 with substitutions at positions 31/42. For every variant,  $\Delta G$  values at 20 °C, measured from chemical denaturation experiments, are represented in black. In addition,  $\Delta G_{0\text{ M NaCl}}/\Delta G_{2\text{ M NaCl}}/\Delta G_{\text{coulomb}}$  terms at 77 °C, derived from thermal unfolding experiments, are shown in gray. *b*, thermal denaturation experiments for the wild-type polypeptide and the thermostable variant **31** monitored by one-dimensional NMR.

ment of both side chains by aliphatic residues results in an improvement of the homeodomain electrostatic balance (as deduced from the  $\Delta G_{\text{coulomb}}$  term). In addition, these substitutions lead to a very favorable  $\Delta G_{2\text{ M NaCl}}$  term, suggesting a more optimized packing or hydrophobic contacts in the mutated proteins. Indeed, energy differences at 0 M NaCl are to a large extent maintained in the presence of very high ionic strength (2 M NaCl, see Fig. 4*a*). In principle,  $\Delta G_{2\text{ M NaCl}}$  should be rather sensitive to the packing of the aliphatic side chain at position 31. To test this point, the leucine residue present in variant **31** (Leu<sup>31</sup>/Leu<sup>42</sup>) was replaced by an isoleucine to give mutant **30** (Ile<sup>31</sup>/Leu<sup>42</sup>). As expected, **30** (Ile<sup>31</sup>/Leu<sup>42</sup>) and **31** (Leu<sup>31</sup>/Leu<sup>42</sup>) present identical coulombic contributions to  $\Delta G$  at 77 °C. In contrast, the salt-independent contribution is slightly reduced in variant **30** (Ile<sup>31</sup>/Leu<sup>42</sup>). The observed destabilization might reflect partially the lower intrinsic helix propensity of the isoleucine with respect to that of leucine. However, it also suggests that a leucine side chain establishes more optimized packing interactions within the protein hydrophobic core than an isoleucine.

As a next step, one-dimensional NMR spectra were collected for the wild-type protein and variant **31** (Leu<sup>31</sup>/Leu<sup>42</sup>) at different temperatures (Fig. 4*b*). It can be observed that the thermophilic polypeptide exhibits good dispersion in the proton one-dimensional spectra, a characteristic of well folded proteins. Moreover, the lack of significant broadening of the signals rules out the existence of aggregation under the experimental conditions employed. Finally, the NMR data are in agreement with the CD denaturation experiments confirming the enhanced

stability of **31** (as revealed by the persistence of its proton NH signals at high temperatures). To determine the influence of pair Leu<sup>31</sup>/Leu<sup>42</sup> on the homeodomain conformation, its solution structure was solved by employing NMR methods. An ensemble of 30 structures was calculated on the basis of 703 unambiguous nuclear Overhauser effects. The structures have very small deviations from ideal geometry and reasonable non-bonded contacts (supplemental Tables S1 and S2). A schematic illustration of the 25 simulated annealed structures is depicted in Fig. 5 and supplemental Fig. S8.

It can be observed that, despite the mutations, the global structure of **31** (Leu<sup>31</sup>/Leu<sup>42</sup>) is very similar to that reported for wild-type Hesx-1 (backbone pairwise root mean square deviation of 0.51 Å between residues 8 and 55). According to the NMR data, Leu<sup>31</sup> is buried, presenting clear contacts with side chains of Leu<sup>42</sup>, Pro<sup>26</sup>, and Phe<sup>49</sup> (Fig. 5, *a* and *b*). Minor structural differences between the two polypeptides are located on the loop between helices I and II (Fig. 5*c*). For example, in the wild-type protein, Pro<sup>26</sup> stacks on the aromatic ring of Phe<sup>49</sup>. As a result of this interaction Pro<sup>26</sup> H $\gamma$  protons present unusual chemical shifts (1.04/0.12 ppm). For variant **31** (Leu<sup>31</sup>/Leu<sup>42</sup>), the Leu<sup>31</sup>/Phe<sup>49</sup> contact induces a slight displacement in the loop residue Pro<sup>26</sup>, which is clearly reflected in its side-chain resonances (*i.e.* the proton H $\gamma$ 2 is shifted downfield 0.85 ppm with respect to wild-type Hesx-1). The refined structure of the thermophilic polypeptide conclusively shows that the interaction between Pro<sup>26</sup> and Phe<sup>49</sup> (present in the wild-type homeodomain) is partially disrupted in **31** by the burial of Leu<sup>31</sup> side chain in the hydrophobic core. In contrast to residue 31, the



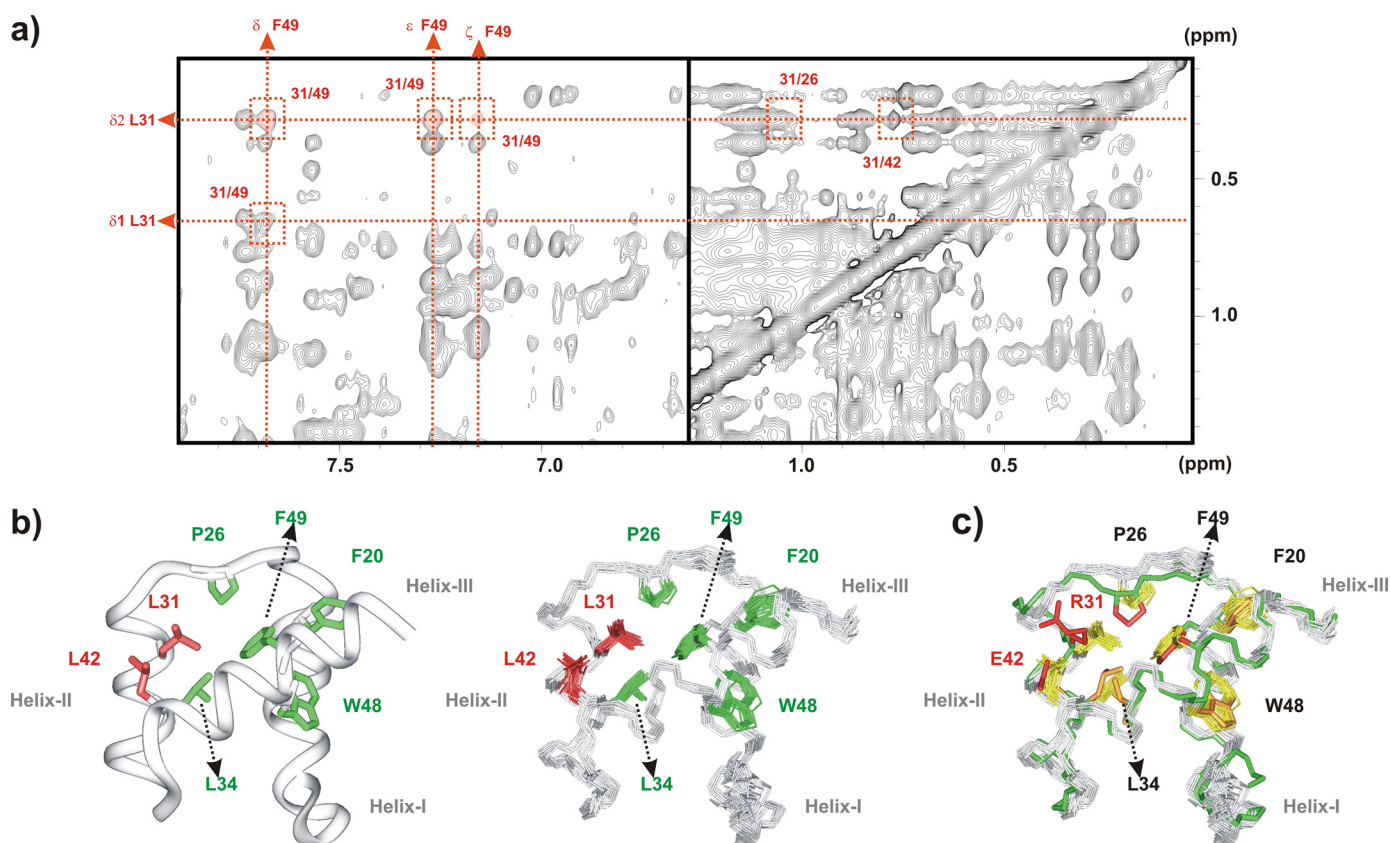


FIGURE 5. *a*, two different regions of a homonuclear two-dimensional NOESY spectrum measured for the thermostable variant 31. Key contacts between side chain of Leu<sup>31</sup> and residues Phe<sup>49</sup>, Pro<sup>26</sup>, and Leu<sup>42</sup> are represented in red. *b*, ensemble of 25 NMR structures (right) obtained for variant 31. A single representative structure is shown on the left for clarity. In all cases, side chains of residues 31/42 are represented in red and those of the core residues 20, 26, 34, 48, and 49 are shown in green. *c*, ensemble of 25 NMR structures of variant 31 (backbone is represented in gray and key side chains in yellow) superimposed on the average structure of wild-type Hesx-1 (backbone and side chains are shown in red and red, respectively). The Pro<sup>26</sup>/Phe<sup>49</sup> interaction, observed in the wild-type homeodomain, is partially disrupted in 31 by the burial of Leu<sup>31</sup> side chain in the hydrophobic core.

leucine at position 42 is exposed to the solvent. However, this residue contributes significantly to protein stabilization through its contact with Leu<sup>31</sup>. The close proximity between both side chains is evidenced by several nuclear Overhauser effect cross-peaks. In fact, the three-dimensional structure obtained shows that Leu<sup>42</sup> aliphatic side chain significantly reduces the solvent-accessible surface of Leu<sup>31</sup>, which is almost totally buried in mutant 31.

In conclusion, aliphatic side chains at positions 31 and 42 provide a more efficient stabilization of the homeodomain fold than the highly conserved Arg<sup>31</sup>/Glu<sup>42</sup> contact. This is achieved without promoting significant aggregation of the polypeptide even at high concentrations. The NMR data are consistent with a local repacking of the hydrophobic core in the thermophilic protein, which is in agreement with the stabilizing  $\Delta G_{2\text{ M NaCl}}$  terms measured for variants 30 (Ile<sup>31</sup>/Leu<sup>42</sup>) and 31 (Leu<sup>31</sup>/Leu<sup>42</sup>).

*Salt Bridges Versus Hydrophobic Pairs at Positions 17/52*—The salt bridge 17/52 has been subjected to several studies in the past. For example, a stabilization of 0.5 kcal/mol has been reported for the K52E variant of the Engrailed homeodomain (in which a bridge Lys<sup>17</sup>/Glu<sup>52</sup> can be formed) with respect to the single mutant K52A (9). Moreover, it has been shown that H52R replacement in vnd/NK-2 (allowing the electrostatic interaction between side chains of Glu<sup>17</sup> and Arg<sup>52</sup>) increases the thermal stability of

the polypeptide by 7.7 °C at 0 M NaCl (12). Overall, these studies indicate that a salt bridge connecting 17 and 52 is moderately stabilizing. Interestingly, close inspection of the sequence information available for homeodomains reveals a very low occurrence of hydrophobic residues at sites 17 or 52. Aliphatic (usually Val or Leu) or aromatic (Tyr in all cases) side chains can be found at position 17 in less than 2% of the sequences included in our data set. Although a minor fraction of polypeptides exhibits aromatic residues (Phe, Trp, or Tyr) at position 52 (<0.5%) aliphatic amino acids (Val, Leu, or Ile) are totally absent from this site. Finally, hydrophobic bridges (involving either aliphatic or aromatic residues) connecting 17/52 were not found in any sequence. This observation argues against the viability of such interaction in homeodomains. To test this prediction additional mutants were designed. Thus, variants 32 and 33, with pairs Leu<sup>17</sup>/Tyr<sup>52</sup> and Tyr<sup>17</sup>/Tyr<sup>52</sup>, respectively, were produced and analyzed. Inspection of molecular models suggests that both pairs should be sterically tolerated and could provide stabilizing interactions with the hydrophobic core of the protein. The obtained results are schematically represented in Fig. 6a and Table 1. A minor amount of mutant 33 (Tyr<sup>17</sup>/Tyr<sup>52</sup>), insufficient for a detailed thermodynamic analysis was purified after several attempts. However, the measured thermal denaturation profiles are consistent with a significant decrease in stability (−15 °C at 0 M NaCl). Regarding variant 32 (Leu<sup>17</sup>/Tyr<sup>52</sup>), this protein was not expressed soluble under

## Salt Versus Hydrophobic Bridges in Stability and DNA Binding

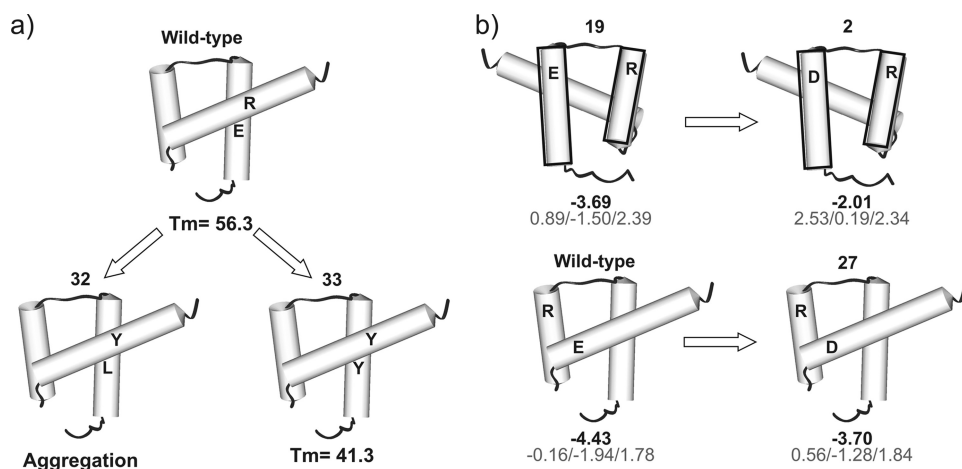


FIGURE 6. *a*, schematic representation of wild-type Hexs-1, together with variants 32 and 33 with substitutions at positions 17/52. For variant 32, stability could not be analyzed due to strong aggregation and precipitation of the protein. Measured  $T_m$  values (in °C) for the wild-type protein and mutant 33 are shown. *b*, schematic representation of wild-type Hexs-1 and variant 19 together with variants 2 and 27, with Glu → Asp substitutions at positions 19 and 42, respectively. For every single mutation,  $\Delta G$  values at 20 °C measured from chemical denaturation experiments are indicated. In addition,  $\Delta G_{0\text{ M NaCl}}/\Delta G_{2\text{ M NaCl}}/\Delta G_{\text{Coulomb}}$  terms at 55 °C, derived from thermal unfolding experiments, are shown in gray. Free energies are given in kcal/mol.

TABLE 2

### Measured $pK_a$ values for glutamate and aspartate side chains in mutants 12 and 15

NMR titrations were performed at 5 °C in 10 mM phosphate, 100 mM NaCl. Estimated errors are indicated.

Residue	Variant 12	Variant 15
Glu <sup>14</sup>	4.41 ± 0.10	4.25 ± 0.10
Glu <sup>15</sup>	3.83 ± 0.14	
Glu <sup>17</sup>	3.75 ± 0.10	3.95 ± 0.11
Glu <sup>19</sup>		<3.9
Asp <sup>29</sup>	3.52 ± 0.08	3.45 ± 0.09
Glu <sup>30</sup>	4.60 ± 0.15	
Glu <sup>32</sup>	5.13 ± 0.12	5.16 ± 0.13
Asp <sup>33</sup>	3.50 ± 0.06	3.65 ± 0.07
Glu <sup>37</sup>		4.52 ± 0.10
Glu <sup>41</sup>	3.52 ± 0.13	3.47 ± 0.11
Glu <sup>42</sup>	4.39 ± 0.10	4.40 ± 0.10
Asp <sup>43</sup>	3.77 ± 0.08	3.67 ± 0.10
Glu <sup>62</sup>	4.00 ± 0.11	4.10 ± 0.10
Met <sup>67</sup>	3.40 ± 0.06	3.50 ± 0.08

any of the conditions tested. It seems that the Leu<sup>17</sup>/Tyr<sup>52</sup> combination strongly promotes protein aggregation. Altogether, these observations suggest that the nonpolar side chains connecting 17/52 reduce the protein stability and might also affect its solubility. In conclusion, only polar residues seem to be tolerated at positions 17/52.

**Stabilizing Influence of Side Chain Packing in Conserved Salt Bridges**—In addition to their electrostatic contribution to the energy of folding, charged side chains could also contribute to protein stability by means of packing interactions involving methylene groups. To analyze this effect, two additional variants were prepared. In both cases, a glutamic acid involved in a salt bridge interaction was replaced by the shorter aspartic acid (see Fig. 6b). Thus, the Glu<sup>19</sup>/Arg<sup>30</sup> bridge present in variant 19 was replaced by pair Asp<sup>19</sup>/Arg<sup>30</sup> (variant 2). In a similar way, residue Glu<sup>42</sup>, which is involved in a salt bridge with Arg<sup>31</sup> in the wild-type protein, was substituted by Asp (to give 27, see Table 1 and Fig. 6b).

According to our analysis, the salt bridge Asp<sup>19</sup>/Arg<sup>30</sup>, present in 2, is strongly destabilizing, with respect to pair Glu<sup>19</sup>/Arg<sup>30</sup> present in variant 19 (see Table 1). This remarkable

destabilization is maintained in the presence of 2 M NaCl ( $\Delta\Delta G_{2\text{ M NaCl}} = 1.69$  kcal/mol). A fraction of the observed difference in  $\Delta G_{2\text{ M NaCl}}$ , estimated in about 0.2 kcal/mol (32), could be attributed to the more unfavorable contribution of the aspartic acid side chain to the intrinsic helix propensity of the polypeptide. The remaining 1.49 kcal/mol difference is suggestive of less optimized packing or hydrophobic interactions in 2 (Asp<sup>19</sup>/Arg<sup>30</sup>). In contrast,  $\Delta G_{\text{Coulomb}}$  terms are similar in 2 (Asp<sup>19</sup>/Arg<sup>30</sup>) and 19 (Glu<sup>19</sup>/Arg<sup>30</sup>), which indicates that the electrostatic interaction between 19 and 30 is not altered significantly by the shortening of the acidic side chain.

In a similar way, the single substitution E42D in the wild-type polypeptide (to give variant 27) leads to a moderate decrease in stability (around 0.7 kcal/mol at both 20 and 55 °C; see Table 1). Again, the observed difference in free energy between the wild-type protein (with pair Arg<sup>31</sup>/Glu<sup>42</sup>) and variant 27 (Arg<sup>31</sup>/Asp<sup>42</sup>) is rather insensitive to the ionic strength, suggesting that the reduced number of hydrophobic and van der Waals contacts established by the shorter Asp<sup>42</sup> side chain is responsible for this effect. In conclusion, our results indicate that the packing of methylene groups from salt-bridging side chains provide significant favorable contributions to the global  $\Delta G$ .

**NMR Analysis of  $pK_a$  Values for Selected Multiple Mutants**—To further dissect the electrostatic contributions to homeodomain stability, we measured the  $pK_a$  values of acid residues (Glu and Asp) in the folded state for selected protein variants, employing NMR spectroscopy. By monitoring the chemical shift dependence of resonances adjacent to the titrating groups, the degree of stabilization of the charged state may be estimated (33–36). Our analysis was focused on quadruple mutant 12 and 15, given that they include all the potential salt bridge interactions conserved within this family of proteins (involving pairs Arg<sup>19</sup>/Glu<sup>30</sup>, Glu<sup>15</sup>/Arg<sup>37</sup>, Arg<sup>31</sup>/Glu<sup>42</sup>, and Glu<sup>17</sup>/Arg<sup>52</sup> in 12 and Glu<sup>19</sup>/Arg<sup>30</sup>, Arg<sup>15</sup>/Glu<sup>37</sup>, Arg<sup>31</sup>/Glu<sup>42</sup>, and Glu<sup>17</sup>/Arg<sup>52</sup> in 15). The results obtained are represented in Table 2 and supplemental Fig. S9.

The measured  $pK_a$  values for Glu<sup>17</sup> in both 12 and 15 are 3.75 and 3.95, respectively, which is consistent with a moderate stabilizing effect of the charged residue (in the 0.6–0.8 kcal/mol range assuming a  $pK_a = 4.4$  in the unfolded state). In contrast, the side chain of Glu<sup>42</sup> shows a  $pK_a$  close to those typical for random coils (4.39 and 4.40 for 12 and 15, respectively) suggesting that, overall, a negative charge at this position has a minor contribution to protein stability.

According to our analysis, negatively charged Glu<sup>15</sup> in quadruple mutant 12 is clearly favorable ( $pK_a = 3.83$ ). On the contrary, glutamate Glu<sup>30</sup> has a minor destabilizing influence on the polypeptide conformation ( $pK_a = 4.60$ ).

The extreme overlapping of Glu<sup>19</sup> proton signals in **15** did not allow an accurate measurement of the corresponding  $pK_a$ . However, an upper limit of 3.9 could be established. This reflects the stabilizing electrostatic interaction between Glu<sup>19</sup> and both Arg<sup>30</sup> and Arg<sup>15</sup>.

Finally, the negative charge of residue 37 in variant **15** seems to have a minor effect on the homeodomain stability as judged by the measured  $pK_a$  value ( $pK_a = 4.52$ ). A plausible explanation is that its favorable contact with Arg<sup>15</sup> is counterbalanced by a repulsive interaction with Asp<sup>33</sup>.

It should be pointed out that these conclusions are in qualitative agreement with those derived from the thermal denaturation studies previously described (a comparison is shown in [supplemental Fig. S10](#)). Thus, the electrostatic influence of a negatively charged side chain on the protein stability could be estimated from the effect of neutralizing mutations on the  $\Delta G_{\text{coulomb}}$  term. For example, the E42L substitution in the wild-type protein, to give variant **28**, entails the removal of a negatively charged residue. As expected from the measured  $pK_a$  value (4.4), this mutation does not alter significantly the electrostatic balance of the homeodomain ( $\Delta\Delta G_{\text{coulomb}} = 0.04$  kcal/mol, see [supplemental Fig. S10a](#)). In another example, the neutralization of Glu<sup>15</sup> in **12** by the E15V mutation, to give variant **13**, produces significant changes in the salt-dependent  $\Delta G_{\text{coulomb}}$  contribution (0.43 kcal/mol), which also is in agreement with the  $pK_a$  measured for this residue ( $pK_a = 3.8$ , see [supplemental Fig. S10c](#)). Similar conclusions can be obtained for the other acidic residues analyzed (see [supplemental Fig. S10](#)). Unfortunately, a more quantitative comparison between the results obtained employing both approaches is not possible given that the  $pK_a$  values in the Hexs-1 denatured state are not known at the moment.

Finally, the analysis of the  $pK_a$  values suggests the presence of additional salt bridges in Hexs-1 not conserved within the homeodomain family. Thus, Glu<sup>41</sup> ( $pK_a = 3.52$  and  $3.47$  for **12** and **15**, respectively) seems to participate in a polar intra-helical contact with Arg<sup>44</sup>. More interestingly, Asp<sup>43</sup> ( $pK_a = 3.77/3.67$ ) might interact electrostatically with Arg<sup>3</sup>, located at the disordered N terminus of the homeodomain. Indeed, both residues are in close proximity according to the NMR structure of wild-type Hexs-1. This dynamic protein region is known to participate in DNA recognition. One could speculate that the Asp<sup>43</sup>/Arg<sup>3</sup> interaction might contribute to DNA binding by orienting the N-terminal region and also by reducing its conformational entropy.

*Influence of Conserved Salt Bridges on the Homeodomain DNA Binding Properties*—As a final step, the specific binding of selected Hexs-1 variants to a 14-bp DNA was analyzed, employing microcalorimetry. Selected experimental conditions (pH 6.0 in 150 mM NaCl, 20 mM phosphate, and 5 mM MgCl<sub>2</sub>) were proximate to those employed in the stability studies previously described, to allow a direct correlation between binding and stability.

First, the influence of pair 31/42 on the thermodynamic parameters of the association was determined. The structural data available indicate that residue Arg<sup>31</sup> usually participates in DNA recognition by forming a salt bridge with a DNA phosphate. Interestingly, in most cases, such interaction is estab-

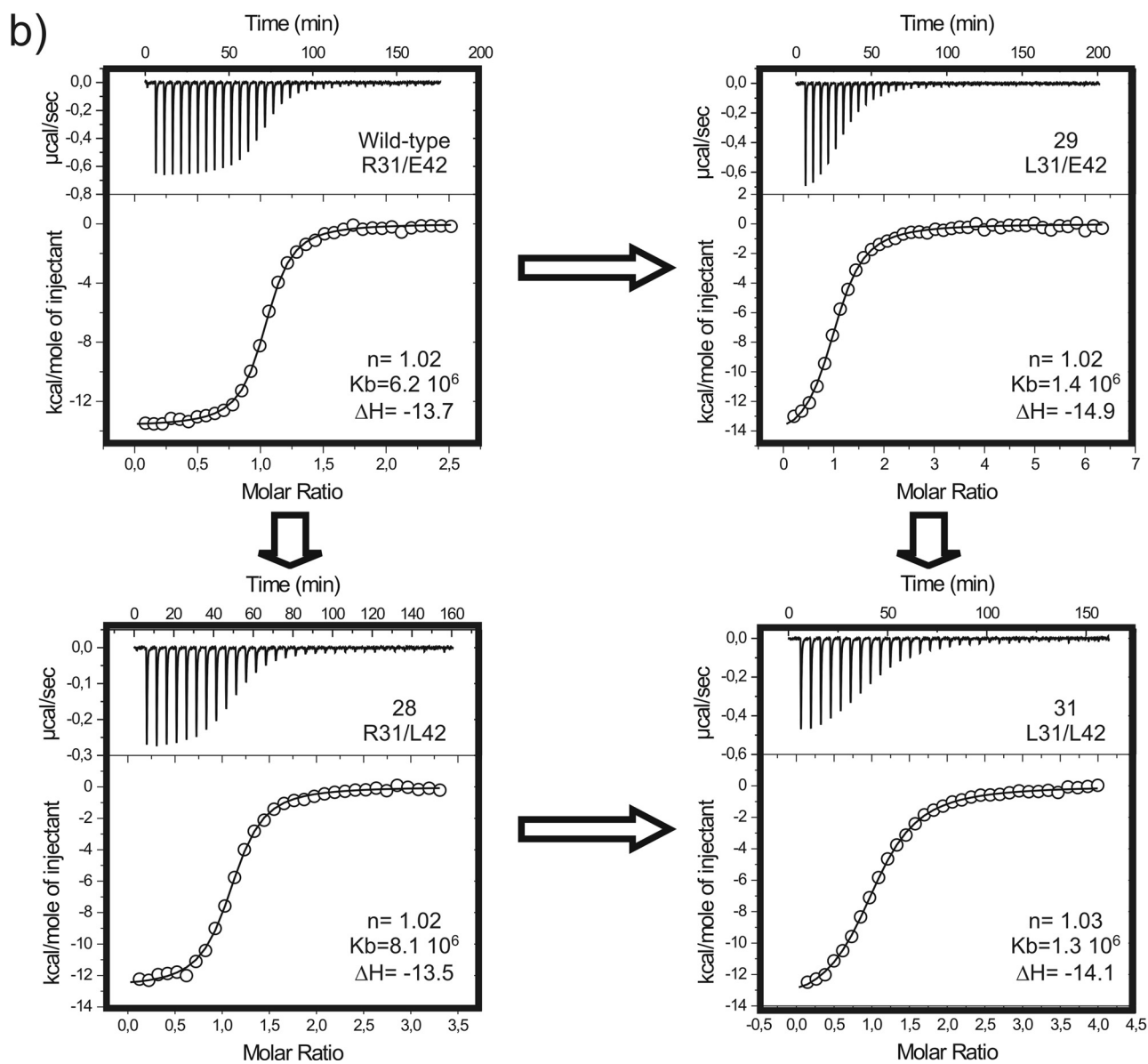
lished without disrupting the previously existing contact Arg<sup>31</sup>/Glu<sup>42</sup> (as illustrated by the x-ray structures of the MSX-1 and Engrailed homeodomains bound to a DNA duplex; see Fig. 7a). The thermodynamic influence on complex formation of such pattern of interactions is unclear. On the one hand, it could be hypothesized that Arg<sup>31</sup>/Glu<sup>42</sup> salt bridge fixes the Arg<sup>31</sup> side chain in the right orientation for DNA binding (pre-organizing the protein receptor and reducing the entropic cost of the recognition process). On the other hand, secondary repulsive interactions between Glu<sup>42</sup> and the DNA backbone might destabilize the complex. However, this side chain is in most cases relatively far from the phosphates (6.5–9.0 Å), and therefore, under physiological conditions, its influence would presumably be small.

ITC titrations performed with the wild-type protein and variants **28** (Arg<sup>31</sup>/Leu<sup>42</sup>), **29** (Leu<sup>31</sup>/Glu<sup>42</sup>), and **31** (Leu<sup>31</sup>/Leu<sup>42</sup>) at 25 °C are represented in Fig. 7b, and the derived thermodynamic parameters are shown in Table 3. It can be observed that, in all cases, complex formation is enthalpically driven, whereas entropy opposes binding. This is in contrast with the results described for the Nkx-2.5 (37) and vnd-NK-2 (12) homeodomains where entropy was found to also be favorable. According to our data, the mutation E42L (to give variant **28**) has an almost negligible influence on the homeodomain binding properties, indicating that the electrostatic interaction between Glu<sup>42</sup> side chain and the DNA receptor, in the wild-type polypeptide, is small. In contrast, the single substitution of Arg<sup>31</sup> by leucine (to give **29**) has a moderate effect on the complex stability. Thus, the affinity constant ( $K_b$ ) is reduced by a factor of 4.4, corresponding to a free energy increment ( $\Delta\Delta G$ ) of 0.9 kcal/mol. This destabilization probably reflects the absence of the Arg<sup>31</sup>/phosphate interaction in the complex formed by the mutated protein. However, it should be pointed out that the observed effect is modest in comparison with that reported for comparable contacts in other homeodomains. Thus, previous studies have shown that, in Hexs-1, the single replacement of Arg<sup>53</sup> (a highly conserved arginine, also involved in a salt bridge with the DNA backbone) by a cysteine decreases binding affinity by more than 2 orders of magnitude (19). This observation highlights the context-dependent character of such interactions and their role on DNA binding.

Finally, it should be noticed that the influence of both mutations on the protein-DNA complex stability is almost additive, *i.e.* the effect of R31L substitution in the presence and absence of Glu<sup>42</sup> is almost identical (see Fig. 7b). It can be concluded from these data that the salt bridge Arg<sup>31</sup>/Glu<sup>42</sup> exerts a minor influence on the homeodomain DNA binding properties.

As a next step, the role of salt bridges 19/30 and 15/37 on DNA binding was analyzed. In contrast with pair Arg<sup>31</sup>/Glu<sup>42</sup>, none of these sites are involved in direct contacts with the DNA duplex. In fact, they are all located far from the protein/DNA interface, as deduced from the inspection of the homeodomain-DNA complex structures available (see Fig. 8). However, they might still have an indirect influence on DNA binding, modulating the dynamic properties of the polypeptide and its adaptability. The thermodynamic parameters collected for selected mutants, with different combinations of charged or hydrophobic residues at these

# Salt Versus Hydrophobic Bridges in Stability and DNA Binding



**TABLE 3**

Thermodynamic parameters measured for the binding of different Hesx-1 variants to a 14-mer DNA duplex (5'-GTCTAATTGACGCG-3' and its complementary 5'-CGCGTCAATTAGAC-3') at 25 °C and pH 6.0 in 20 mM phosphate, 150 mM NaCl, 5 mM MgCl<sub>2</sub>, employing microcalorimetry (ITC)

Variant	31/42	<i>n</i>	<i>K<sub>b</sub></i>	Δ <i>G</i>	Δ <i>H</i>	TΔ <i>S</i>
			<i>M</i> <sup>-1</sup>	kcal/mol	kcal/mol	kcal/mol
31	Leu/Leu	1.03	1.3 10 <sup>6</sup>	-8.3 ± 0.2	-14.1 ± 0.6	-5.8 ± 0.6
29	Leu/Glu	1.02	1.4 10 <sup>6</sup>	-8.3 ± 0.2	-14.9 ± 0.5	-6.5 ± 0.5
28	Arg/Leu	1.02	8.1 10 <sup>6</sup>	-9.4 ± 0.2	-13.5 ± 0.5	-4.1 ± 0.5
Wild-type	Arg/Glu	1.02	6.2 10 <sup>6</sup>	-9.2 ± 0.1	-13.7 ± 0.4	-4.5 ± 0.4
Variant	19/30, 15/37	<i>n</i>	<i>K<sub>b</sub></i>	Δ <i>G</i>	Δ <i>H</i>	TΔ <i>S</i>
			<i>M</i> <sup>-1</sup>	kcal/mol	kcal/mol	kcal/mol
21	Val/Ile, Val/Leu	1.04	6.5 10 <sup>6</sup>	-9.3 ± 0.1	-13.8 ± 0.5	-4.5 ± 0.5
Wild-type	Val/Ile, Val/Lys	1.02	6.2 10 <sup>6</sup>	-9.2 ± 0.1	-13.7 ± 0.4	-4.5 ± 0.4
2	Asp/Arg, Val/Lys	1.01	6.7 10 <sup>6</sup>	-9.3 ± 0.2	-15.3 ± 0.6	-6.0 ± 0.6
16	Arg/Glu, Val/Lys	1.03	5.8 10 <sup>6</sup>	-9.2 ± 0.1	-13.6 ± 0.6	-4.4 ± 0.6
17	Glu/Arg, Val/Lys <sup>(Leu26)</sup>	1.01	4.7 10 <sup>6</sup>	-9.1 ± 0.2	-14.1 ± 0.5	-5.0 ± 0.5
7	Glu/Arg, Arg/Lys	1.02	5.5 10 <sup>6</sup>	-9.2 ± 0.1	-14.8 ± 0.5	-5.6 ± 0.5
11	Arg/Glu, Glu/Lys	1.02	4.0 10 <sup>6</sup>	-9.0 ± 0.1	-14.1 ± 0.4	-5.1 ± 0.4
15	Glu/Arg, Arg/Glu	0.92	5.3 10 <sup>6</sup>	-9.1 ± 0.1	-13.2 ± 0.5	-4.1 ± 0.5

sites, shown in Table 3, do not support this hypothesis. It can be observed that differences in binding strength ( $\Delta\Delta G_b$ ), with respect to the wild-type polypeptide, are in the 0–0.2 kcal/mol range. Although somewhat larger differences in enthalpy were measured, they are partially cancelled by the entropic contribution (enthalpy/entropy compensation). Under the employed experimental conditions, neither the conformational stability of the polypeptide (maximum differences in the free energy of folding for the selected mutants amounts to 2.6 kcal/mol at 20 °C; see Table 1), nor the protein total charge (that varies from -1 to +1 with respect to the wild-type homeodomain) seems to have a significant effect on complex formation.

To further analyze the DNA binding properties of the mutated homeodomains, the temperature dependence of the thermodynamic parameters was measured, by performing ITC titrations at three different temperatures (15, 25, and 35 °C) with a reduced set of variants (see Table 4 and supplemental Figs. S11 and S12). It has been shown that the specific interaction of homeodomains with DNA is accompanied by folding of their intrinsically disordered N- and C-terminal regions. In fact, coupling of local folding and binding is common between DNA-binding proteins (38). Considering that all protein variants included in our analysis present mutations far from these unstructured regions, it might be anticipated that they would have a minor influence on the induced fit of the protein upon binding to DNA. However, it has been shown that the homeodomain core is dynamic leading to the hypothesis that this character is essential for the adaptability of the polypeptide. In addition, it has been suggested that salt bridge 19/30 is part of a co-varying network of interacting residues that plays a role in the induced fit of the protein on complex formation (9). The estimated changes in heat capacity that accompany the specific

binding of the selected variants to DNA ( $\Delta C_p$ , one of the thermodynamic signatures of the induced fit) do not provide support for this hypothesis. Thus, for the wild-type protein it can be observed (Table 4 and supplemental Figs. S11 and S12) that the binding process becomes more exothermic with temperature. The measured variation in  $\Delta H$  is consistent with a negative  $\Delta C_p$  value of -0.38 kcal/(mol K). A similar behavior was deduced for variants 16, 11, and 7 (including salt bridges 19/30 of opposite polarities). Our experimental data suggest a minor increase in  $\Delta C_p$  for the mutated homeodomains that might correlate with their thermal stability. This would be consistent with a slightly more pronounced adjustment of the polypeptides in the more unstable variants. However, the observed differences are, in all cases,  $\leq 0.1$  kcal/(mol K). Overall, the obtained data rule out a major effect of pairs 19/30 and 15/37 on DNA recognition and suggest that the induced fit of the protein is mainly restricted to residues in or close to the protein/DNA interface.

## DISCUSSION

According to our analysis, of the three salt bridges conserved in homeodomains, two (19/30 and 31/42) are clearly destabilizing with respect to aliphatic pairs found in a minor fraction of sequences. Thus, we have shown that replacement of the highly conserved salt bridge 31/42 by an aliphatic pair (in variants 30 and 31) leads to a thermophilic version of the Hesx-1 homeodomain. Moreover, for sites 19/30 the stabilizing influence of the aliphatic side chains is maintained even in those cases in which a polar network involving tetrad 15/19/30/37 could be established. Our analysis shows that stabilization by the aliphatic residues reflects, in all cases, the improvement in packing interactions with the protein hydrophobic core (as judged from the  $\Delta G_{2 M NaCl}$  terms). The

FIGURE 7. *a*, representation of the protein-DNA contacts involving pair Arg<sup>31</sup>/Glu<sup>42</sup> in MSX-1 (PDB code 1IG7) and Engrailed (PDB code 1DU0) homeodomains according to x-ray studies. The shorter distance between the acidic side chain and the phosphate backbone is shown in both cases (in Å). The salt bridge Arg<sup>31</sup>/Glu<sup>42</sup> is also present in wild-type Hesx-1. *b*, ITC profiles for the titration of Hesx-1 variants with single or double substitutions at positions 31/42, into solutions containing duplex DNA (5'-GTCTAATTGACGCG-3' and its complementary 5'-CGCGTCAATTAGAC-3') at 25 °C, pH 6.0, in 20 mM phosphate, 150 mM NaCl, 5 mM MgCl<sub>2</sub>. The upper panel shows in all cases the raw data (thermal power) corresponding to the injection of the proteins into the DNA solution. The integrated signal, after subtraction of the dilution heat, yielded the heat per mol of protein injected, which is plotted against the protein/DNA molar ratio in the lower panel. The solid line in bottom panels represents the least squares fitting of the data assuming a 1:1 complex. The estimated thermodynamic parameters are shown (see also Table 3).

## Salt Versus Hydrophobic Bridges in Stability and DNA Binding

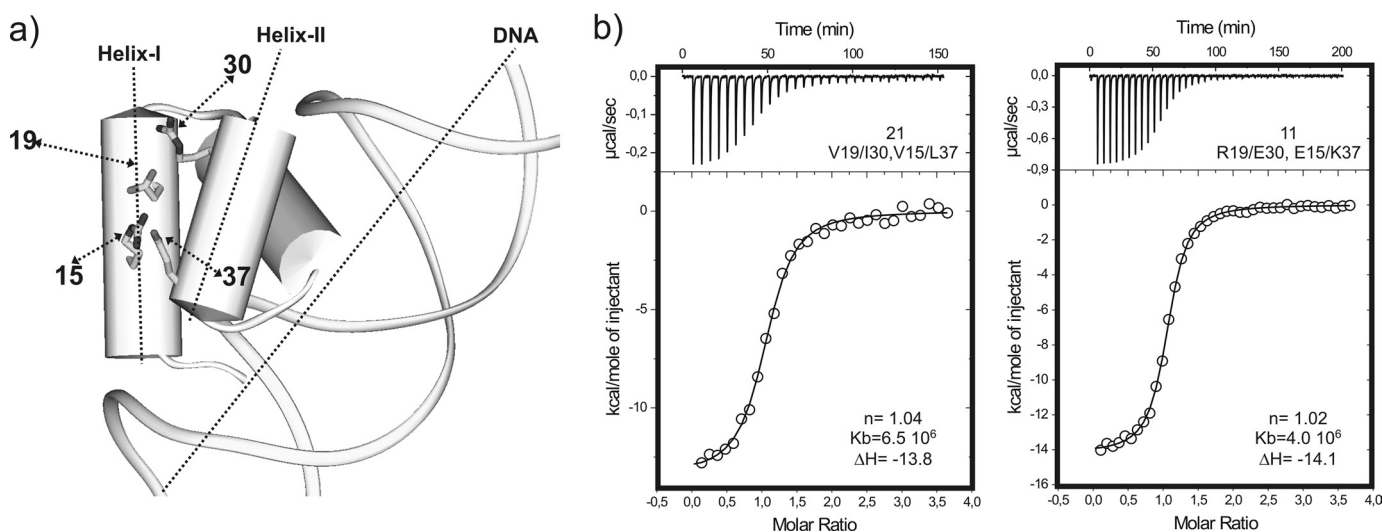


FIGURE 8. *a*, representation of side chains at positions 19/30 and 15/37 in the specific complex between the Engrailed homeodomain and DNA, according to x-ray data (PDB code 1DU0). Residues at these positions in wild-type Hex-1 are Val<sup>19</sup>/Ile<sup>30</sup> and Val<sup>15</sup>/Lys<sup>37</sup>. *b*, ITC profiles for the titration of Hex-1 variants **21** and **11**, with hydrophobic or charged side chains at these sites into solutions containing duplex DNA (at 25 °C, pH 6.0, in 20 mM phosphate, 150 mM NaCl, 5 mM MgCl<sub>2</sub>). The estimated thermodynamic parameters are shown (see also Table 3).

**TABLE 4**

Thermodynamic parameters for the binding of wild-type Hex-1 and variants 7, 11, and 16 to a 14-mer DNA duplex at 15, 25, and 35 °C and pH 6.0 in 20 mM phosphate, 150 mM NaCl, 5 mM MgCl<sub>2</sub> employing microcalorimetry (ITC)

Variant	19/30, 15/37	Temperature	<i>n</i>	<i>K<sub>b</sub></i>	$\Delta H$
		°C			
Wild-type	Val/Ile, Val/Lys	15	1.04	1.7 10 <sup>7</sup>	-10.3 ± 0.3
		25	1.02	6.2 10 <sup>6</sup>	-13.7 ± 0.4
		35	0.99	2.6 10 <sup>6</sup>	-18.0 ± 0.6
16	Arg/Glu, Val/Lys	15	0.92	1.6 10 <sup>7</sup>	-10.0 ± 0.3
		25	1.03	5.8 10 <sup>6</sup>	-13.6 ± 0.6
		35	1.01	3.0 10 <sup>6</sup>	-19.7 ± 0.6
11	Arg/Glu, Glu/Lys	15	1.02	1.2 10 <sup>7</sup>	-10.6 ± 0.3
		25	1.02	4.0 10 <sup>6</sup>	-14.1 ± 0.4
		35	1.02	2.0 10 <sup>6</sup>	-19.3 ± 0.6
7	Glu/Arg, Arg/Lys	15	0.99	1.3 10 <sup>7</sup>	-10.5 ± 0.3
		25	1.02	5.5 10 <sup>6</sup>	-14.8 ± 0.5
		35	0.99	2.7 10 <sup>6</sup>	-20.7 ± 0.8

three-dimensional NMR structure obtained for thermophilic mutant **31** is consistent with this view.

Despite the observed influence of salt *versus* hydrophobic bridges, charged side chains at positions 19, 30, 31, and 42 contribute to the stability of the homeodomain fold, through packing interactions between the side-chain methylene groups and the protein hydrophobic core. The destabilization induced by single Glu/Asp substitutions in protein variants **2** and **27** is illustrative of this effect.

In addition, our work provides clues about the correlation between triad 19/26/30, deduced from the statistical analysis of homeodomain sequences. Thus, a proline at position 26 is compatible with both aliphatic and polar (neutral or charged) residues at 19/30. In contrast, the sterically larger and conformationally more mobile leucine is incompatible with aliphatic side chains at positions 19/30. Violations of this rule lead to a severe destabilization and/or aggregation of the protein, which is in agreement with the absence of this particular combination from any natural sequence.

It should be emphasized that, apart from position 26, the hydrophobic environment of residues 19/30 is highly conserved

and shows no relevant differences in homeodomains with salt or hydrophobic bridges connecting these sites (supplemental Fig. S13). Taking this into account, the observed preference for aliphatic residues at positions 19/30 is likely to be general for those polypeptides with a proline in 26.

Overall, the results obtained are in agreement with those reported by other groups for different systems (39) and correspond well with theoretical predictions by Hendsch and Tidor (31).

In contrast with sites 19/30 and 31/42, only polar residues seem to be tolerated at positions 17 and 52. This particular behavior might have its origin in the dynamic properties of the polypeptide. Whereas sites 19, 30, 31, and 42 are located in well structured regions, residue 52 is located close to the end of helix III. This segment of the protein (especially from residue 55), although not totally disordered in solution, is characterized by a low helical content that might increase solvent exposure of pair 17/52. Thus, a hydrophobic pair at these sites is likely to affect the protein stability and promote aggregation.

Finally, the influence of positions 15, 19, 30, 31, 37, and 42 on the homeodomain DNA binding properties has been tested employing microcalorimetry. First, our analyses confirm the participation of Arg<sup>31</sup> in the recognition of DNA. However, its effect on the binding strength (only a factor of 4.4 in *K<sub>b</sub>*) is relatively small in comparison with that reported for other comparable contacts in homeodomains. In contrast, the influence of positions 15, 19, 30, 37, and 42 on the DNA binding properties of the polypeptide is almost negligible. Interestingly, the thermodynamic parameters derived seem almost independent of the protein conformational stability. Indeed, folding free energies of the analyzed homeodomains range from -2.01 to -4.60 kcal/mol (at 20 °C, see Table 1). Despite this fact, all of them exhibit similar affinities for DNA ( $\Delta\Delta G_b \leq 0.3$  kcal/mol). Moreover, binding enthalpies ( $\Delta H$ ) and the changes in heat capacity ( $\Delta C_p$ ) are also comparable. It has been hypothesized that the dynamic character of the homeodomain hydrophobic core plays a role in the induced fit of the protein upon

complex formation. Considering that protein stability and dynamics are usually strongly correlated, the observed lack of influence of the former on DNA binding argues against this proposal. Instead, it strongly suggests that the homeodomain core does not play a significant role in the induced fit mechanism, which is probably restricted to the protein/DNA interface. This conclusion is probably meaningful in the context of evolution. Thus, if the dynamics of the homeodomain hydrophobic core was important for complex formation, many of the mutations affecting the conformational stability of the polypeptide would also have an effect on DNA binding. In contrast, the observed lack of correlation between binding and stability allows an independent modulation of both properties by evolution. In this context, it should be considered that homeodomains are present in both cold-blooded and warm-blooded organisms, under different environments, and therefore should be able to satisfy different stability requirements. Rapid protein turnover ensures that some regulatory proteins are degraded so that the cell can respond to constantly changing conditions. Such proteins usually have evolved to be relatively unstable under physiological conditions. We propose that salt/hydrophobic bridges at positions 19/30, 15/37, and 31/42 probably contribute to adjust the thermal stability of the polypeptide for a particular environment while maintaining binding affinity and specificity. In agreement with this hypothesis, an analysis of the sequence information available reveals that stabilizing hydrophobic pairs at positions 31/42 in homeodomains can be found exclusively in warm-blooded animals. Moreover, these sequences exhibit also a significantly greater abundance of hydrophobic bridges connecting sites 19 and 30 (3.5% versus 1% in sequences from cold-blooded animals).

In summary, the studies presented herein provide a detailed description of the role of conserved salt bridge interactions on the homeodomain stability and DNA binding properties together with new insights into the relationship between sequence, structure, stability, and function within this important family of proteins.

**Accession Number**—The coordinates corresponding to the NMR structure of the thermostable Hesx-1 homeodomain R31L/E42L double mutant have been deposited in the Protein Data Bank (PDB) under code 2k40.

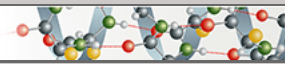
**Acknowledgments**—We thank Centro de Supercomputación de Galicia for computer support. We are grateful to Dr. D. V. Laurents for helpful discussion and English style corrections.

## REFERENCES

- Chi, Y. I. (2005) *Hum. Genet.* **116**, 433–444
- Cillo, C., Faiella, A., Cantile, M., and Boncinelli, E. (1999) *Exp. Cell Res.* **248**, 1–9
- Billeter, M. (1996) *Prog. Biophys. Mol. Biol.* **66**, 211–225
- Yang, Q. S. (1996) *Prog. Biochem. Biophys.* **23**, 29–33
- Banerjee-Basu, S., Moreland, T., Hsu, B. J., Trout, K. L., and Baxevanis, A. D. (2003) *Nucleic Acids Res.* **31**, 304–306
- Banerjee-Basu, S., Ryan, J. F., and Baxevanis, A. D. (2000) *Nucleic Acids Res.* **28**, 329–330
- Banerjee-Basu, S., Sink, D. W., and Baxevanis, A. D. (2001) *Nucleic Acids Res.* **29**, 291–293
- Clarke, N. D. (1995) *Protein Sci.* **4**, 2269–2278
- Stollar, E. J., Mayor, U., Lovell, S. C., Federici, L., Freund, S. M., Fersht, A. R., and Luisi, B. F. (2003) *J. Biol. Chem.* **278**, 43699–43708
- Santoro, M. M., and Bolen, D. W. (1988) *Biochemistry* **27**, 8063–8068
- Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) *Protein Sci.* **4**, 2138–2148
- Gonzalez, M., Weiler, S., Ferretti, J. A., and Ginsburg, A. (2001) *Biochemistry* **40**, 4923–4931
- Kumar, A., Ernst, R. R., and Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun.* **95**, 1–6
- Bax, A., and Davis, D. G. (1985) *J. Magn. Reson.* **65**, 355–360
- Marion, D., and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.* **113**, 967–974
- Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) *J. Mol. Biol.* **273**, 283–298
- Scheek, R. M., van Gunsteren, W. F., and Kaptein, R. (1989) *Methods Enzymol.* **177**, 204–218
- Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) *J. Am. Chem. Soc.* **117**, 5179–5197
- de la Mata, I., Garcia, J. L., González, C., Menéndez, M., Cañada, J., Jiménez-Barbero, J., and Asensio, J. L. (2002) *ChemBioChem* **3**, 726–740
- Fraenkel, E., Rould, M. A., Chambers, K. A., and Pabo, C. O. (1998) *J. Mol. Biol.* **284**, 351–361
- de Los Rios, M. A., and Plaxco, K. W. (2005) *Biochemistry* **44**, 1243–1250
- Martin, A., Kather, L., and Schmid, F. X. (2002) *J. Mol. Biol.* **318**, 1341–1349
- Max, K. E., Wunderlich, M., Roske, Y., Schmid, F. X., and Heinemann, U. (2007) *J. Mol. Biol.* **369**, 1087–1097
- Perl, D., Mueller, U., Heinemann, U., and Schmid, F. X. (2000) *Nat. Struct. Biol.* **7**, 380–383
- Wunderlich, M., Martin, A., and Schmid, F. X. (2005) *J. Mol. Biol.* **347**, 1063–1076
- Horovitz, A., Serrano, L., Avron, B., Bycroft, M., and Fersht, A. R. (1990) *J. Mol. Biol.* **216**, 1031–1044
- Dahiyat, B. I., and Mayo, S. L. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10172–10177
- Lim, W. A., Farruggio, D. C., and Sauer, R. T. (1992) *Biochemistry* **31**, 4324–4333
- Lim, W. A., and Sauer, R. T. (1989) *Nature* **339**, 31–36
- Lim, W. A., and Sauer, R. T. (1991) *J. Mol. Biol.* **219**, 359–376
- Hendsch, Z. S., and Tidor, B. (1994) *Protein Sci.* **3**, 211–226
- Chakrabarty, A., Kortemme, T., and Baldwin, R. L. (1994) *Protein Sci.* **3**, 843–852
- Bosshard, H. R., Marti, D. N., and Jelesarov, I. (2004) *J. Mol. Recognit.* **17**, 1–16
- Wyman, J., Jr. (1948) *Adv. Protein Chem.* **4**, 407–531
- Marti, D. N., and Bosshard, H. R. (2003) *J. Mol. Biol.* **330**, 621–637
- Marti, D. N., Jelesarov, I., and Bosshard, H. R. (2000) *Biochemistry* **39**, 12804–12818
- Fodor, E., Mack, J. W., Maeng, J. S., Ju, J. H., Lee, H. S., Gruschus, J. M., Ferretti, J. A., and Ginsburg, A. (2005) *Biochemistry* **44**, 12480–12490
- Spolar, R. S., and Record, M. T., Jr. (1994) *Science* **263**, 777–784
- Waldburger, C. D., Schildbach, J. F., and Sauer, R. T. (1995) *Nat. Struct. Biol.* **2**, 122–128

**Protein Structure and Folding:  
Role of Conserved Salt Bridges in  
Homeodomain Stability and DNA Binding**

PROTEIN STRUCTURE  
AND FOLDING



Mario Torrado, Julia Revuelta, Carlos  
Gonzalez, Francisco Corzana, Agatha Bastida  
and Juan Luis Asensio

*J. Biol. Chem.* 2009, 284:23765-23779.

doi: 10.1074/jbc.M109.012054 originally published online June 26, 2009

---

Access the most updated version of this article at doi: [10.1074/jbc.M109.012054](https://doi.org/10.1074/jbc.M109.012054)

Find articles, minireviews, Reflections and Classics on similar topics on the [JBC Affinity Sites](https://www.jbc.org/).

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

Supplemental material:

<http://www.jbc.org/content/suppl/2009/06/26/M109.012054.DC1.html>

This article cites 39 references, 6 of which can be accessed free at  
<http://www.jbc.org/content/284/35/23765.full.html#ref-list-1>