Homo- and Heteronuclear Two-Dimensional NMR Studies of the Globular Domain of Histone H1: Full Assignment, Tertiary Structure, and Comparison with the Globular Domain of Histone H5[†,‡]

Corinne Cerf,[*,§] Guy Lippens,[§] V. Ramakrishnan,[∥] Serge Muyldermans,[⊥] Alain Segers,[⊥] Lode Wyns,[⊥] Shoshana J. Wodak,[§] and Klaas Hallenga[§]

Unité de Conformation des Macromolécules Biologiques (UCMB), CP 160-16, Université Libre de Bruxelles, Avenue P. Hêger, B-1050 Brussels, Belgium, Biology Department, Brookhaven National Laboratory, Upton, New York 11973, and Instituut voor Molekulaire Biologie, Vrije Universiteit Brussel, Paardenstraat 65, B-1640 Sint-Genesius-Rode, Belgium

Received May 10, 1994; Revised Manuscript Received June 27, 1994[⊗]

ABSTRACT: The globular domain of chicken histone H1 (GH1) has been studied by ^{1}H homonuclear and ^{1}H–^{15}N heteronuclear 2D NMR spectroscopy. After the full assignment of the proton and ^{15}N resonances, the tertiary structure of GH1 was determined by an iterative procedure using distance geometry and restrained simulated annealing. The secondary structure elements of GH1, three helices (S5–A16, S24–A34, N42–K56) followed by a β-hairpin (L59–L73), are folded in a manner very similar to the corresponding parts of the globular domain of chicken histone H5 (GH5) [Clore et al. (1987) *EMBO J. 6*, 1833–1842; Ramakrishnan et al. (1993) *Nature 362*, 219–223]. However, subtle differences are detected between the two structures and between the electrostatic potentials surrounding the molecules. The most important differences are located in the loop between the second and third helices, a region that could be responsible for the different affinity for DNA. The most positively charged regions are not found in exactly the same position in GH1 and GH5. Nevertheless, their location seems to agree with the model where nucleosome binding takes place through contact points located at one DNA terminus and close to the dyad axis of the nucleosome [Schwabe & Travers (1993) *Curr. Biol. 3*, 628–630].

The linker histone H1 and its extreme variant H5 play important roles in the nuclei. Beyond their structural role of inducing the folding of chromatin into the 30-nm filament (Thoma et al., 1979; Graziano et al., 1994), it is now known that these proteins play a more specific role in the regulation of gene expression (Sun et al., 1989; Zlatanova, 1990; Garrard, 1991). This involves binding of the histone to the nucleosome, probably in a sequence-specific manner although it is still a matter of debate whether the specificity is directed toward G-C or A-T base pairs (Croston et al., 1991; Zlatanova & Yaneva, 1991; Wellman et al., 1994; Muyldermans & Travers, 1994; Hendrickson & Cole, 1994). The globular central part of the linker histones (GH1/GH5)[1] is essential for the binding to the nucleosome. It is responsible for the specific positioning of the protein on the nucleosome, and it confers the same nucleosome protection from micrococcal nuclease digestion as the complete protein (Allan et al., 1980; Buckle et al., 1992). The N- and C-terminal tails, very basic and largely unstructured (except for a possible α-helix in the C-terminal tail; Clark et al., 1988; Hill et al., 1989), are probably involved in nonspecific interactions with linker DNA (Staynov & Crane-Robinson, 1988; Segers et al., 1991).

H5 partially replaces H1 during the maturation of avian erythroid cells (Affolter et al., 1987). Both linker histones globally play the same role, but H5 binds more strongly to the nucleosome, thereby increasing the stability of chromatin, which leads to a decrease of gene expression (Sun et al., 1990). The isolated GH5 also binds more strongly to the nucleosome than GH1 (Thoma et al., 1983).

The tertiary structure of chicken erythrocyte GH5 was determined first by NMR spectroscopy (Clore et al., 1987), and recently by X-ray crystallography, to a resolution of 2.5 Å (Ramakrishnan et al., 1993). We have studied the solution structure of chicken erythrocyte GH1 by NMR spectroscopy. This 75-residue domain has about 40% sequence identity with GH5. In this paper, we briefly present the full assignment of the molecule and then describe its 3D structure determination. The GH5 structure was not used in the structure determination but was merely consulted as a guide. The comparison of the independently determined GH1 structure to the known structure of GH5 is presented and discussed in the context of binding of linker histones to DNA.

MATERIALS AND METHODS

NMR Spectroscopy. The NMR experiments were performed at 295 K on two recombinant GH1 samples, one of which was ^{15}N-labeled, dissolved in a buffer composed of

[†] C.C. and G.L. are respectively Research Assistant and Senior Research Assistant of the National Fund for Scientific Research (Belgium). V.R. is supported by NIH Grant GM42796 and the Office of Health and Environmental Research of the U.S. Department of Energy. Part of this work was financially supported by the National Fund for Scientific Research (Belgium). This text presents research results of the Belgian programme on Interuniversity Poles of attraction initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. The scientific responsibility is assumed by its authors.

[‡] The coordinates of the 14 final structures have been deposited in the Brookhaven Protein Data Bank. The accession number is 1GHC.

* To whom correspondence should be addressed.

§ Université Libre de Bruxelles.

∥ Brookhaven National Laboratory.

⊥ Vrije Universiteit Brussel.

⊗ Abstract published in *Advance ACS Abstracts,* September 1, 1994.

[1] Abbreviations: GH1, globular domain of histone H1; GH5, globular domain of histone H5; 1D, one dimensional; 2D, two dimensional; 3D, three dimensional; NMR, nuclear magnetic resonance; DQF-COSY, double-quantum-filtered *J*-correlated spectroscopy; DQSY, double-quantum spectroscopy; TOCSY, total correlation spectroscopy; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; HSQC, heteronuclear single-quantum coherence; RMSD, root mean square deviation; HNF-3γ, hepatocyte nuclear factor 3γ.

0006-2960/94/0433-11079$04.50/0

210 mM Na_2SO_4, 245 mM NaH_2PO_4, and 35 mM Na_2HPO_4 (pH 5.8). Spectra were recorded on a Varian Unity 600-MHz spectrometer [for details, see Cerf et al. (1993)].

NOE Constraints Used in Structure Calculations. NOE constraints were extracted from both NOESY and HSQC-NOESY spectra and separated into four groups depending on their intensity (peak height). The intense peaks were assumed to correspond to the distance range 2–3 Å. The medium peaks were assigned to the distance range 2–4 Å. For the weak peaks connecting a proton of residue i to the amide proton of residue $i + 1$, distance bounds of 2–5 Å were adopted. For all other weak peaks, distance bounds were set to 2–5.5 Å.

Instead of using pseudoatoms representing methyl groups or methylene groups with degenerate proton resonances, we preferred to constrain the parent carbon atoms and to add 1 Å to the upper distance limit. For aromatic δ and ϵ protons, whose respective resonances were always degenerate in our case, we substituted the constraint involving each couple of indistinguishable protons by two constraints, applied to both protons, the upper distance limit for those new constraints being increased by 4.8 Å (corresponding to the distance between both δ or ϵ protons).

Due to overlaps in the spectra, the stereospecific assignments of all prochiral groups of protons could not been obtained from the analysis of intraresidual NOEs and coupling constants, without risking errors. For diastereotopic protons or methyl pairs, only one constraint with increased distance was applied to the parent atom, as described above. However, in the last simulated annealing refinement of the structures, a high-dimensional potential (Habazettl et al., 1990) was used in addition to the usual potential, in order to achieve the stereospecific assignments automatically during the structure calculations. This high-dimensional potential forces the side chains to adopt the conformation that fits all NOE distances to stereospecifically assignable protons or methyl groups. To use this potential, it is necessary to specify the minimum and maximum differences between the distances from a given proton to both diastereotopic protons or methyl groups. In order to do this, we classified the intensities in the following way: class 1 for the distance range 2–3 Å, class 2 for the distance range 2–4 Å, class 3 for the distance ranges 2–5 and 2–5.5 Å, and class 4 for absent peaks in nonoverlapping regions. This classification allowed us to attribute a minimum difference between distances of a given proton toward two diastereotopic groups as a function of the class difference: 0.5 Å for a class difference of 1, 1 Å for a class difference of 2, and 2 Å for a class difference of 3. For the maximum differences between distances, we chose 1.8 Å for diasterotopic protons and 2.5 Å for diastereotopic carbon atoms of methyl groups from valines and leucines.

Protocol for Structure Calculations. Using the X-PLOR package (Brünger et al., 1992), the protocol to calculate structures was the following: (A) distance geometry generation of substructures including amide N, amide H, carbonyl C, Cα, Hα, Cβ, Hβ, Cγ, and C/Nδ, (B) building of complete side chains and subsequent simulated annealing regularization, (C) simulated annealing refinement of the structures, and (D) analysis of the structures and generation of a subfamily of "acceptable" structures. The following acceptance criteria were used: no NOE distance outside the distance bounds plus 0.5 Å, RMSD for bond deviations from ideality smaller than 0.01 Å, and RMSD for angle deviations from ideality smaller than 2°. The four stages of calculation (A–D) constitute the standard protocol except that the definition of the substructures in (A) was extended in order to include more long-range constraints in this distance geometry stage. In fact, for each constraint involving Hγ or Hδ, a constraint involving the parent atom was added, the upper distance limit being increased by 1 Å for all amino acid types except for the aromatics. These constraints were thus taken into account already in stage A of the calculation. The NOE connectivities were constrained using a split harmonic energy term with a force constant of 50 kcal mol^{-1} $Å^{-2}$.

Electrostatic Potential Calculations. The electrostatic potential surfaces around GH1 and GH5 were calculated by the program DELPHI (Gilson & Honig, 1987) which solves the linearized Poisson–Boltzmann equation at the nodes of a cubic lattice. The calculation parameters were 65 nodes per cube dimension, successively 30%, 50%, 75%, and 90% of occupancy of the cube by the protein, 2 for the dielectric constant of the protein, 78.6 for the dielectric constant of the solvent, 0.2 M for the ionic strength (corresponding to the intracellular ionic strength), 2 Å for the exclusion radius of ions, and 1.4 Å for the exclusion radius of the solvent.

Structural Analysis and Display. All structures were displayed and graphically analyzed using the BRUGEL software (Delhaise et al., 1985). In particular, it allowed us to view selected constrained distances on the graphics screen, with indication about the extent of their violation, or to display the calculated electrostatic potential surfaces.

RESULTS

GH1 Full Assignment. The previously obtained sequential assignment of the resonances of GH1 up to β protons (Cerf et al., 1993) was readily extended up to the end of the side chains by a careful analysis of the TOCSY and HSQC-TOCSY spectra. For aromatic residues and asparagine, the NOESY connectivities between Hβ and Hδ (HNδ for the asparagine) were also used. The resulting resonance assignments are listed in Table 1. A first observation is the large number of resonance degeneracies. This is due to the amino acid composition (9 alanines, 9 glycines, 10 leucines, 11 lysines, 10 serines, and only 2 aromatics). The number of degeneracies is considerably greater than for GH5, which was assigned by Zarbock et al. in 1986, explaining the need to label GH1 with ^{15}N.

GH1 Structure Calculations. After a manual peak picking of the interresidual NOE connectivities, we automatically generated all possible assignments for these peaks by referring to the list of resonance assignments. The large number of resonance degeneracies caused many ambiguities in the assignment of NOE connectivities. To overcome them, an iterative approach was used, similar to that of Güntert et al. (1993). First we considered only the nonambiguous peaks to generate a family of preliminary structures. Given this family, we checked each possibility of assignment for ambiguous peaks by looking at the distance between the protons considered in the structures generated and eliminating clearly wrong assignments. Some peaks then became unambiguous and could be included in a second run of structure calculations. Four such runs were executed (I, II, III, and IV), leading to increasingly precise families (see Table 2). Before each run, we checked the more frequently violated NOE connectivities and sometimes reassigned them. We also examined the accepted structures on the graphics screen, in order to eliminate obviously wrong structures. For example, in the first run, some helices were misfolded, being left-handed in a given segment and right-handed everywhere else. This phenomenon disappeared in the following runs.

The diagonal plots of Figure 1 show that the number of new interresidual constraints (represented in black) in each run tends to decrease and that they connect mostly residues close

Table 1: ^{15}N and ^{1}H Resonance Assignments of GH1 in H_2O at 295 K, pH 5.8[a]

residue	^{15}NH	HN	Hα/Hα′	Hβ/Hβ′	others	residue	^{15}NH	HN	Hα/Hα′	Hβ/Hβ′	others
M1	117.4	7.88	3.82			K41	118.7	7.92	4.07	1.70/1.46	Hγ/Hγ′ 1.30/1.30, Hδ/Hδ′ 1.67/1.67, Hε/Hε′ 2.99/2.99
A2	118.1	8.50	4.97	1.80							
G3	109.7	7.83	3.95/3.81								
P4	133.1[b]		4.50	2.25/1.75	Hγ/Hγ′ 2.04/2.01, Hδ/Hδ′ 3.66/3.66	N42	110.3	7.23	5.25	3.35/2.65	HNδ/HNδ′ 7.09/6.51
						N43	119.8	7.30	4.35	3.14/2.90	HNδ/HNδ′ 7.74/7.05
S5	116.1	8.68	4.59	4.21/3.96		S44	112.2	9.06	4.18	4.00/4.00	
V6	122.9	9.14	3.66	2.17	Hγ(m)/Hγ′(m)[c] 0.91/0.77	R45	119.9	7.55	4.11	2.17/1.97	Hγ/Hγ′ 1.88/1.75, Hδ/Hδ′ 3.30/3.30
T7	109.6	7.79	3.81	4.16	Hγ(m) 1.29	I46	119.5	8.21	3.49	2.10	$H\gamma_1/H\gamma_1'$ 2.02/2.02, $H\gamma_2$(m) 0.67, $H\delta_1$(m) 1.02
E8	123.7	7.54	4.07	2.20/2.06	Hγ/Hγ′ 2.32/2.30						
L9	117.4	7.94	3.88	1.69/1.10	Hγ 1.60, Hδ(m)/Hδ′(m) 0.70/0.44	K47	117.3	7.81	3.84	1.94/1.94	Hγ/Hγ′ 1.37/1.37, Hδ/Hδ′ 1.73/1.73, Hε/Hε′ 2.97/2.97
I10	117.4	8.46	3.44	1.69	$H\gamma_1/H\gamma_1'$ 1.96/1.96, $H\gamma_2$(m) 0.63, $H\delta_1$(m) 0.70	L48	117.8	8.22	4.11	1.80/1.80	Hγ 1.61, Hδ(m)/Hδ′(m) 0.92/0.92
T11	112.7	8.02	3.53	4.23	Hγ(m) 1.16						
K12	120.8	8.18	3.98	1.86/1.86	Hγ/Hγ′ 1.40/1.40, Hδ/Hδ′ 1.57/1.57, Hε/Hε′ 2.96/2.96	G49	108.8	8.34	3.80/3.29		
						L50	121.1	8.36	3.74	1.68/1.15	Hγ 1.33, Hδ(m)/Hδ′(m) 0.34/−0.11
A13	120.1	7.77	4.12	1.32		K51	116.1	7.66	3.98	1.98/1.98	Hγ/Hγ′ 1.64/1.48, Hδ/Hδ′ 1.73/1.73, Hε/Hε′ 3.02/3.02
V14	115.3	8.44	3.64	2.02	Hγ(m)/Hγ′(m) 1.01/1.01						
S15	114.5	8.44	4.22	4.05/4.00		S52	113.8	8.28	4.27	4.05/4.05	
A16	119.5	7.51	4.35	1.49		L53	119.7	8.18	4.09	2.05/2.05	Hγ 1.29, Hδ(m)/Hδ′(m) 0.85/0.81
S17	111.3	7.33	4.33	4.01/4.01							
K18	123.1	8.71	4.36	2.03/1.85	Hγ/Hγ′ 1.57/1.50, Hδ/Hδ′ 1.74/1.74, Hε/Hε′ 3.05/3.05	V54	119.7	8.48	4.37	2.23	Hγ(m)/Hγ′(m) 1.12/1.12
						S55	116.7	8.51	4.30	4.11/4.11	
E19	118.4	8.21	4.35	2.14/1.96	Hγ/Hγ′ 2.32/2.32	K56	116.1	8.03	4.34	1.98/1.67	Hγ/Hγ′ 1.12/1.12, Hδ/Hδ′ 1.63/1.63, Hε/Hε′ 2.98/2.98
R20	123.4	8.89	4.07	1.93/1.93	Hγ/Hγ′ 1.73/1.62, Hδ/Hδ′ 3.25/3.25						
K21	114.3	8.41	4.22	1.82/1.82		G57	104.9	7.83	4.35/3.94		
G22	105.6	7.66	4.23/3.84			T58	116.4	8.01	4.09	3.97	Hγ(m) 1.26
L23	118.4	8.71	4.44	1.62/1.62	Hγ 1.81, Hδ(m)/Hδ′(m) 1.01/0.94	L59	115.0	7.06	5.29	1.45/1.39	Hγ 1.51, Hδ(m)/Hδ′(m) 0.81/0.72
S24	120.9	8.68	4.93	4.43/4.43		V60	116.3	9.21	4.34	1.87	Hγ(m)/Hγ′(m) 0.86/0.84
L25	120.1	9.09	4.17	2.03/1.75	Hγ 1.66, Hδ(m)/Hδ′(m) 1.01/0.93	Q61	125.0	8.88	4.83	1.74/1.63	Hγ/Hγ′ 1.99/1.99, $^{15}N\epsilon$ 109.8, HNε/HNε′ 7.08/6.91
A26	118.3	8.69	4.08	1.47							
A27	121.0	8.19	4.16	1.64							
L28	121.6	9.14	4.06	2.15/1.63	Hγ 1.48, Hδ(m)/Hδ′(m) 0.83/0.75	T62	119.0	8.66	4.33	4.13	Hγ(m) 1.10
						K63	121.0	8.41	4.30	1.84/1.79	Hγ/Hγ′ 1.41/1.41, Hδ/Hδ′ 1.64/1.64, Hε/Hε′ 2.98/2.98
K29	117.1	8.76	3.82	2.03/1.92	Hγ/Hγ′ 1.18/1.18, Hδ/Hδ′ 1.78/1.70, Hε/Hε′ 2.94/2.81						
						G64	108.4	8.40	4.01/4.01		
K30	117.0	7.74	4.09	1.97/1.97	Hγ/Hγ′ 1.64/1.48, Hδ/Hδ′ 1.73/1.73, Hε/Hε′ 3.00/3.00	T65	111.3	8.16	4.39	4.25	Hγ(m) 1.19
						G66	111.3	8.75	4.04/3.97		
						A67	123.7	8.63	4.07	1.46	
A31	121.3	8.03	4.26	1.54		S68	111.8	8.15	4.54	4.01/4.01	
L32	118.0	8.73	4.06	1.91/1.30	Hγ 1.78, Hδ(m)/Hδ′(m) 0.61/0.51	G69	108.1	8.20	3.91/3.91		
						S70	113.4	7.87	4.75	3.42/3.42	
A33	122.3	8.65	4.86	1.65		F71	118.4	8.68	5.34	2.83/2.77	Hδ/Hδ′ 7.06/7.06, Hε/Hε′ 7.15/7.15, Hζ 7.28
A34	121.1	8.19	4.25	1.58							
G35	101.6	7.47	4.47/3.86								
G36	106.0	7.97	4.42/3.76			R72	116.8	8.83	4.88	1.72/1.63	Hγ/Hγ′ 1.52/1.41, Hδ/Hδ′ 3.10/3.04
Y37	122.1	7.80	4.40	2.85/2.45	Hδ/Hδ′ 6.85/6.85, Hε/Hε′ 6.60/6.60						
						L73	120.7	8.76	4.60	1.77/1.63	Hγ 1.72, Hδ(m)/Hδ′(m) 1.00/0.93
D38	124.6	8.12	4.61	2.59/2.44							
V39	120.6	7.88	3.72	2.20	Hγ(m)/Hγ′(m) 1.23/1.14	S74	116.1	8.06	4.21	3.82/3.82	
						K75	126.2	7.90	4.21	1.86/1.75	Hγ/Hγ′ 1.41/1.41, Hδ/Hδ′ 1.63/1.63, Hε/Hε′ 3.01/3.01
E40	117.7	8.16	4.25	2.09/2.09	Hγ/Hγ′ 2.34/2.25						

[a] Chemical shifts (ppm) refer to the HSQC-NOESY spectrum with 150-ms mixing time. ^{1}H frequencies were measured relative to an internal reference of sodium trimethylsilylpropionate (0 ppm), and ^{15}N frequencies were measured relative to a fictive ^{15}N reference calculated by the method of Live et al. (1984). [b] Determined from the direct ^{15}N 1D spectrum. [c] (m) stands for methyl.

to those already constrained. On the other hand, after these four runs, the number of nonassigned or ambiguously assigned peaks was sufficiently low to revert to a manual spectrum analysis, in order to include more qualitative features like peak shape to assign the NOE peaks. We also specifically searched for contacts between secondary structure elements, most useful for determining precise 3D structures. They were the most difficult to assign by our iterative approach based on frequency lists and compatibility with preliminary structures: the relative position of the secondary structure elements in these preliminary structures was often not precise enough, leading to multiple possibilities for the given cross peak or to none at all. This manual analysis led us to add new interresidual constraints of great significance for the tertiary structure (e.g., between residue 53 and residues 7, 10, 11, and 29 or between residue 73 and residues 11 and 14) with which we executed a fifth run of structure calculations (see Figure 1, V).

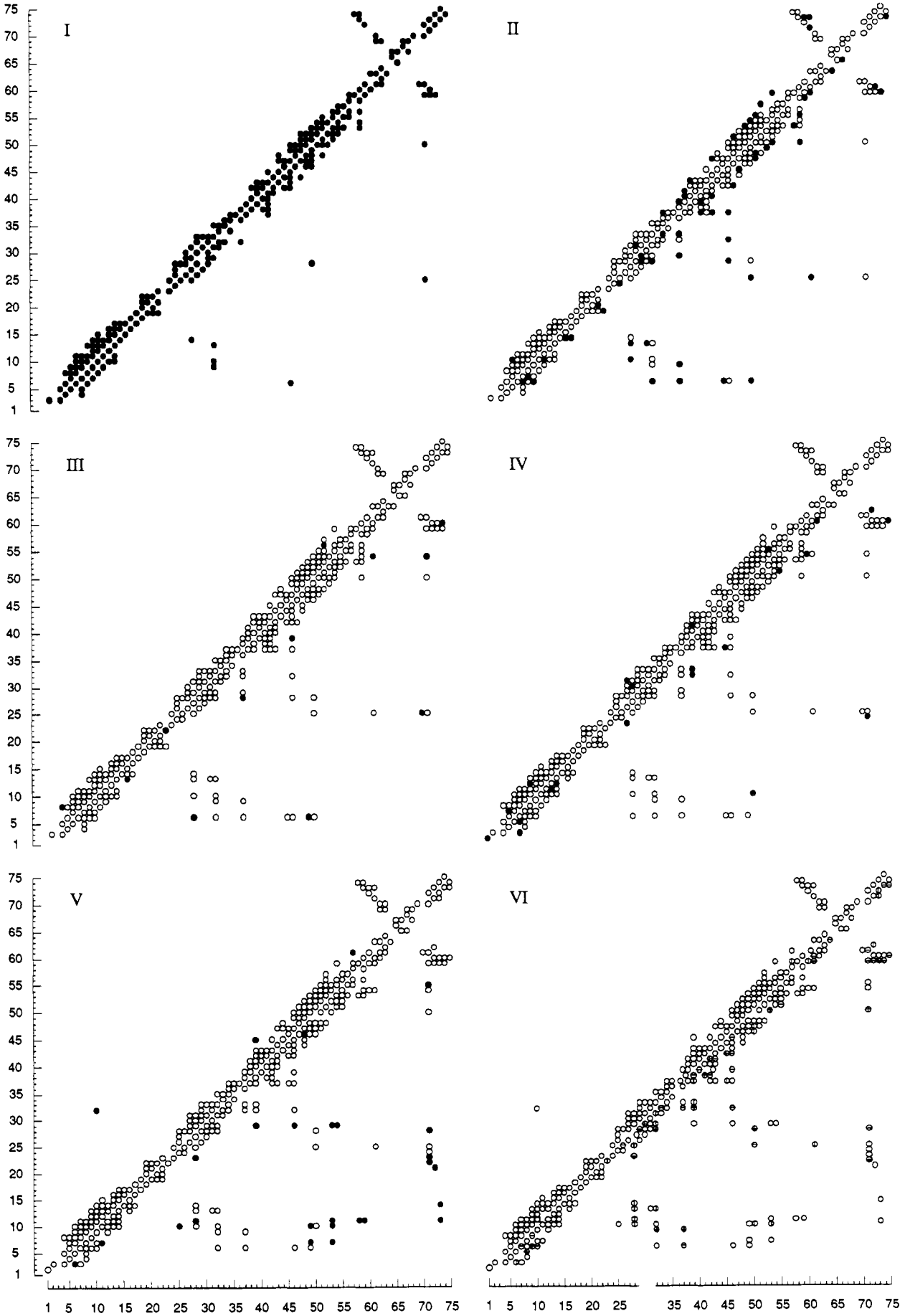

FIGURE 1: Diagonal plots of the interresidual distance constraints used in the six structure calculation runs (labeled I–VI). Backbone–backbone NOEs are shown above the diagonal whereas backbone–side chain and side chain–side chain NOEs are shown below the diagonal. New constraints are represented as black filled circles; preexisting constraints are represented as open circles. In run VI, stereospecific assignments are indicated by line segments superimposed on the open circles and pointing toward the residue containing the prochiral group.

Table 2: Protocol Followed for the GH1 Structure Determination

	no. of starting structures	stages in structure calculations[a] A	B	C	D	no. of accepted structures	manual selection	no. of selected structures	av pairwise backbone RMSD[b] (Å)
run I	50	yes	yes	yes	yes	20	yes	9	4.86
run II	100	yes	yes	yes	yes	10	no	10	3.93
run III	10 accepted structures from run II	no	no	yes	yes	8	no	8	3.78
run IV	8 accepted structures from run III	no	no	yes	yes	6		0	
	30	yes	yes	yes	yes	13	no	13	3.38
run V	20	yes	yes	yes	yes	19	no	19	2.67
run VI	20	yes	yes	yes	yes	1		0	
	100	yes	yes	yes	yes	14	no	14	2.51

[a] A, distance geometry; B, regularization; C, refinement, and D, acceptance; as defined in the text. [b] Backbone definition includes N, Cα, and C atoms.

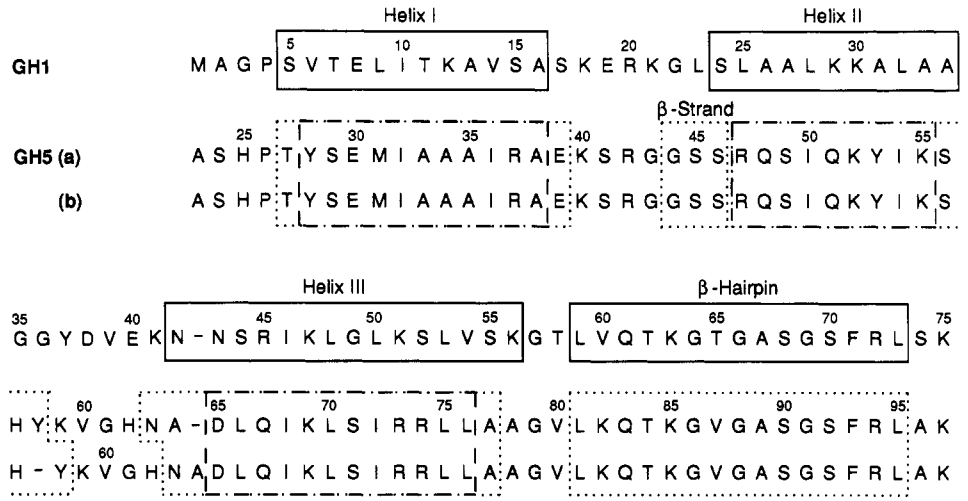

FIGURE 2: Sequence alignment of GH1 and GH5 (a) according to Coles et al. (1987) and (b) according to Crane-Robinson and Ptitsyn (1989). The secondary structure elements are represented as boxes: solid line for GH1 (this work), dashed line for GH5 as determined by Clore et al. (1987), and dotted line for GH5 as determined by Ramakrishnan et al. (1993).

Until that stage, all calculations were performed without stereospecific assignments. The last run was executed using a high-dimensional potential (Habazettl et al., 1990), in order to automatically reach the stereospecific assignments (see Figure 1, VI).

The entire procedure, together with the average pairwise backbone RMSD of the resulting families of structures, is summarized in Table 2. The structural determination of GH1 could not be continued much further, considering the fact that all remaining nonassigned peaks in the spectra are very weak or questionable. The final 14 structures were calculated using 921 interresidual constraints (344 sequential, 275 medium range up to $i-i+4$, 302 long range) in the standard potential and 134 pairs of constraints toward stereotopic groups (24 sequential, 44 medium range, 66 long range) in the high-dimensional potential.

Description of the GH1 Structures. The secondary structure elements described in Cerf et al. (1993), namely, three helices followed by a β-hairpin, are confirmed and now precisely delimited as follows: S5–A16 for helix I, S24–A34 for helix II, N42–K56 for helix III, and L59–L73 for the hairpin (see Figure 2). Their positions in space are represented in Figure 3.[2] The conformation of each of the three helices and of the loop connecting helices II and III is well defined (the average pairwise backbone RMSDs in the 14 structures are 1.16, 1.11, 0.60, and 0.88 Å for helix I, helix II, helix III, and loop II–III, respectively). On the other hand, the loop connecting helices I and II and the β-hairpin are less well determined, probably as a consequence of their mobility (the average pairwise backbone RMSDs are respectively 1.63 and 2.32 Å).

DISCUSSION

Comparison between GH1 and GH5 Structures. (A) Overall Conformation. In addition to the similarity between the secondary structures of GH1 and GH5 (Cerf et al., 1993; Clore et al., 1987; Ramakrishnan et al., 1993; see Figure 2), the tertiary structures are also very similar although a β-hairpin present in both the GH1 NMR structure and the GH5 X-ray structure was not seen in the GH5 NMR structure (Cerf et al., 1993). Since the GH5 X-ray structure has the highest resolution, we have used it for detailed comparison. On Figure 3 (top view) are superimposed the 14 final GH1 NMR structures and the GH5 X-ray structure (Ramakrishnan et al., 1993). The backbone RMSD between GH5 (molecule A in the asymmetric unit) and the GH1 average structure is quite high (3.43 Å), but this deviation drops to 1.71 Å when only the helices are considered.

(B) Primary DNA Binding Site. The primary DNA binding site of GH5 probably includes three basic residues, namely, K69, R73, and K85 (Ramakrishnan et al., 1993). The first two residues are located in helix III, which is thought to constitute the DNA recognition helix by analogy to the helix–turn–helix motif in other DNA-binding proteins (Brennan, 1991). This was confirmed by the recent cocrystal structure of an eukaryotic transcription factor (HNF-3γ) with DNA, very similar to the GH5 structure, where the helix corresponding to helix III of GH5 is presented to the major groove of DNA (Clark et al., 1993). In GH1, the three basic residues of GH5 are conserved, with one arginine-to-lysine substitu-

[2] The electrostatic potentials are expressed in kT/e, where k is the Boltzmann constant, T is the temperature in kelvin, and e is the charge of the electron. With T set to 298 K, kT/e amounts to 25.6 mV or 0.6 kcal/mol of electrons.

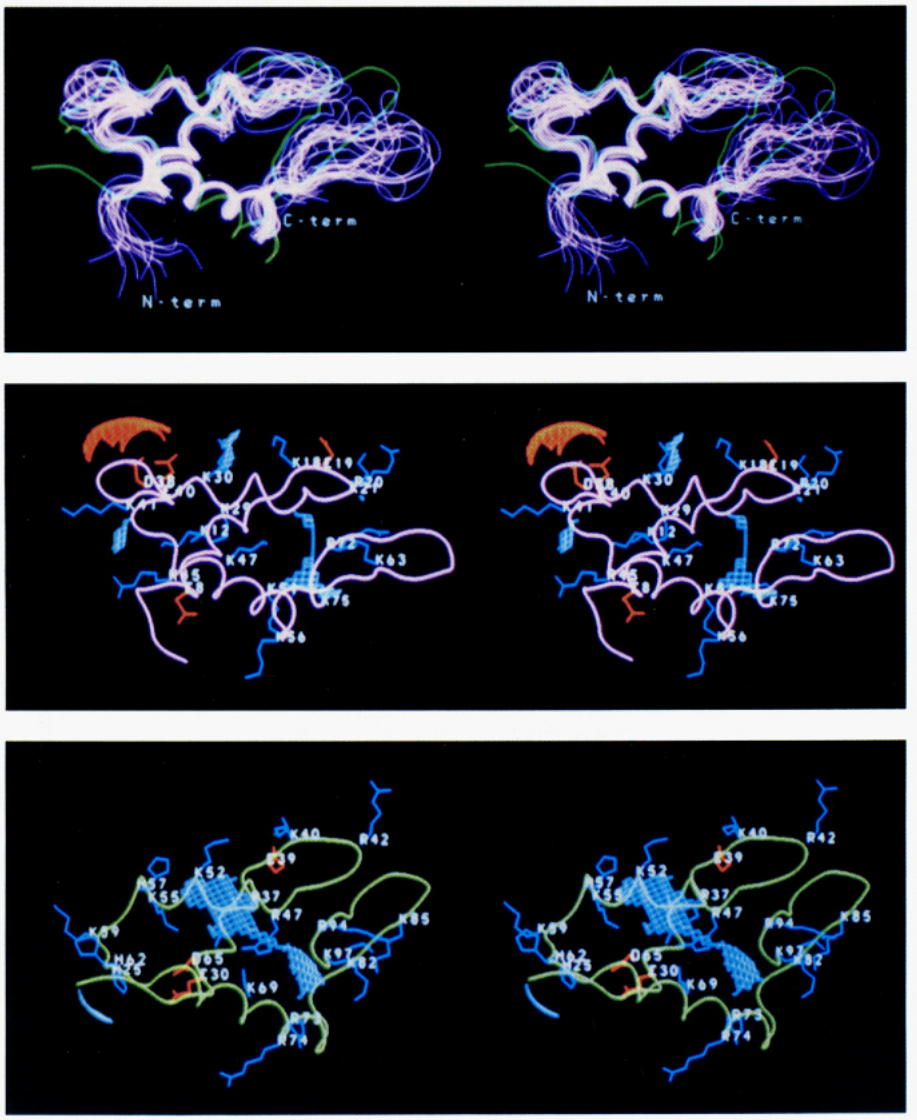

FIGURE 3: Stereoviews of GH1 (NMR structure; this work) and GH5 (X-ray structure). Colors are purple for the GH1 backbone, green for the GH5 backbone, blue for positively charged side chains, and orange for negatively charged side chains. Top view: overlay of the final 14 GH1 structures superposed on the GH5 structure fitted to minimize the backbone RMSD on the helices. Middle view: electrostatic potential surfaces around one GH1 structure (potential +4 in light blue, potential −2 in orange). Bottom view: electrostatic potential surfaces around GH5 (potential +4 in light blue).

tion: K47, K51, and K63. The side chains of these residues are oriented in the same way as in GH5 (data not shown). This suggests that the primary DNA-binding site in both molecules is the same.

(C) β-Hairpin. This is the less well determined part of the GH1 structure. At this stage, it is impossible to say if it is due to flexibility or to a lack of constraints. Distance geometry calculations can create the impression of flexibility when fewer NOE constraints exist, which is the case for small peptides and for hairpins or loops of proteins extending into the solvent [see Lippens et al. (1993)]. On the other hand, in the GH5 crystal (Ramakrishnan et al., 1993), both monomers constituting the asymmetric unit are in a nearly identical conformation except in this hairpin part. This is not the case in the crystal structure of the HNF-3γ/DNA complex (Clark et al., 1993), where the corresponding part is well fixed (and fits the A monomer of GH5). This tends to indicate that the β-hairpin of GH1 and GH5, whose amino acid sequence is very well conserved, would adopt a well-defined conformation only in the presence of DNA.

(D) Loop between Helices II and III. This region is very interesting because it is well determined in both GH1 and GH5 (average pairwise backbone RMSD for GH1 is 0.88 Å; Cα RMSD between A and B monomers in the GH5 crystal is 0.3 Å). However, the structure of the loop is very different in GH1 and GH5, with a backbone RMSD of 2.77 Å. It may not be by accident that there is a discrepancy between the two sequence alignments of GH1 and GH5 just in this region (see Figure 2). According to the sequence alignments, histidine-62 of GH5, which can be cross-linked to chromatin (Mirzabekov et al., 1989), is substituted either by glutamate-40 or by lysine-41 of GH1. Histidine-62 of GH5 and lysine-41 of GH1 could play the same role since they are both positively charged when in proximity to the DNA. The other basic amino acid of GH5 in this region, lysine-59, is substituted in GH1 either by tyrosine-37 or by aspartate-38, none of which is basic. Until now, lysine-59 was not invoked as being involved in DNA binding. Nevertheless, Thomas and Wilson (1986) showed that this lysine is slightly protected against chemical modification by the presence of the nucleosome. The loop in GH1 has thus a different conformation and is globally more negatively charged than in GH5. More precisely, if GH5 and GH1 are modeled in interaction with DNA, based on the structure of the HNF-3γ/DNA complex (Clark et al., 1993), it can be seen that in GH5 the loop follows one of the DNA strands while in GH1 it is somewhat away from the DNA

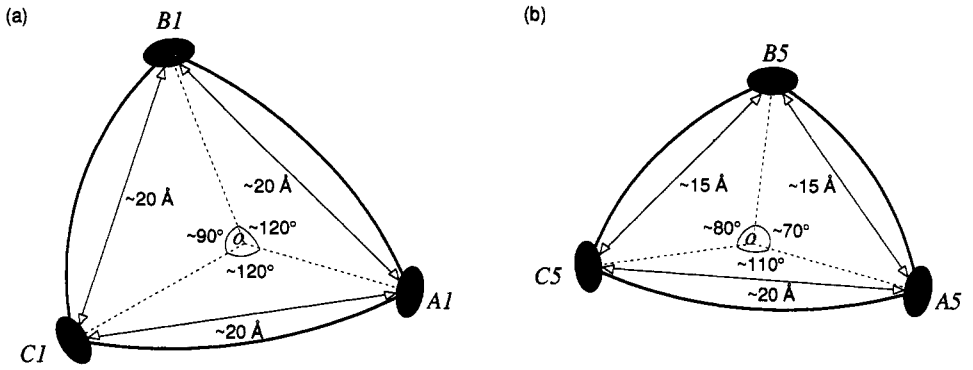

FIGURE 4: Geometric parameters describing the relative positions of regions with an electrostatic potential of +4. (a) In GH1, *A1* represents the probable primary DNA binding site, *B1* and *C1* constitute possible secondary binding sites, and *o* is the center of the molecule. (b) In GH5, *A5* represents the probable primary binding site, *B5* and *C5* constitute possible secondary binding sites, and *o* is the center of the molecule.

(data not shown). This may be the chief reason why GH5 and GH1 differ in their affinity for DNA and it could have biological relevance in the sense that H5 is thought to be a "permanent" seal for the nucleosome while H1 is a "temporary" or removable one.

(E) Limited Stability of GH1. Two experimental facts tend to indicate a limited stability of GH1. First, GH1 only adopted a stable structure under very specific buffer conditions, including sulfate ions (see Materials and Methods). For GH5, a standard phosphate buffer was used (Zarbock et al., 1986; Clore et al., 1987). Moreover, during our proton exchange NMR experiments (protonated protein dissolved in D_2O), all amide protons of GH1 exchanged in less than half an hour, indicating that none was involved in a stable hydrogen bond. This was not the case for GH5 where most of the amide protons belonging to the helices exchanged with deuterium in more than 24 h (Zarbock et al., 1986). This difference is too dramatic to be explained by the difference in pH during the exchange experiments (pH 5.8, respectively 3.7).

(F) Electrostatic Potentials. Binding of linker histone to DNA is believed to involve interaction between positively charged and negatively charged groups and thus essentially electrostatic stabilizations. This is confirmed by the ready dissociation of linker histone from DNA induced by increasing salt concentration (Kumar & Walker, 1980). In order to investigate the role of the electrostatic contributions, the electrostatic potentials surrounding the GH1 and GH5 molecules were calculated and compared. Among the 14 GH1 NMR structures, the one having a conformation closest to the average NMR conformation was selected. For GH5, the A monomer of the X-ray structure (Ramakrishnan et al., 1993) was chosen. The only difference between the A and B monomers is in the β-hairpin. The hairpin of monomer A has the best defined electronic density, was used by the authors (Ramakrishnan et al., 1993) for the modeling with DNA, and is in a closer conformation to the corresponding part in the HNF-3γ/DNA complex (Clark et al., 1993). It is thus likely to more closely represent the structure of GH5 when bound to DNA. These choices of structures were made after we checked that all GH1 NMR structures on the one hand, and both monomers of the GH5 X-ray structure on the other hand, displayed qualitatively the same electrostatic potentials. For GH5, histidines were taken as positively charged since in the strong negative electrostatic field of the DNA their pK_a should be raised above neutral. On the other hand, we used uncharged NH_2 and COOH polypeptide termini in order to avoid singularities in the potential. This choice was not crucial for the biological relevance, as both isolated globular domains and complete histones bind to DNA in the same way (Allan et al., 1980; Buckle et al., 1992).

Each basic residue (lysine, arginine, histidine) gives rise to a positive potential, while each acid residue (aspartate or glutamate) gives rise to a negative one. Due to the excess of basic residues in GH1 and GH5, a slightly positive electrostatic potential surface (around a potential of +1) surrounds both molecules at a few angstroms of the protein surface. In regions including basic residues and closer to the protein surface, GH1 and GH5 are surrounded by increasingly positive potential surfaces. At the level of +4, only three regions can be identified in both molecules (labeled as *A1, B1, C1* for GH1 and *A5, B5, C5* for GH5; see Figure 3). In GH1, the regions *A1* and *B1* surround two clusters of lysines (K47/K51/K63 and K18/K30), very well conserved in the linker histone family, which belong to the putative primary and secondary DNA-binding sites, respectively (Ramakrishnan et al., 1993). On the other hand, the *C1* region surrounds a basic cluster (K12/R45) not invoked for DNA binding and not conserved in the family. In GH5, the *A5* region is equivalent to *A1*. It surrounds residues K69, R73, and K85, which constitute the putative primary DNA-binding site. The *B5* region, nonequivalent to *B1*, surrounds a cluster comprising R47, K52, and K55, where only K52 is conserved in the family and belongs to the putative secondary DNA-binding site. Finally, the *C5* region surrounds H25, a residue that can be cross-linked to DNA (Mirzabekov et al., 1989).

(G) Implications for Nucleosome Binding. Assuming that regions *A1* and *A5* constitute the primary DNA-binding site of GH1 and GH5 and that, for each molecule, one of the other regions could constitute a secondary binding site, geometric parameters were calculated in order to see if indications about the positioning of GH1/5 on the nucleosome could be found. The resulting distances between the different regions, and the angles between the vectors taking their origin at the center of the molecule and pointing to the different regions, are represented in Figure 4. All distances are in the range 15–20 Å; the angles vary from 70° to 120°.

GH1/5 is known to contact the nucleosome at the entry/exit point of the DNA which makes two turns around a central histone octamer (Thoma et al., 1979), but the exact points of interaction are not known. The first models suggested three binding sites, at the dyad axis and at both DNA termini (Allan et al., 1980; Crane-Robinson & Ptitsyn, 1989). More recent data suggest that GH1/5 binds to the nucleosome at two sites only (Thomas et al., 1992; Draves et al., 1992; Ramakrishnan et al., 1993). After the crystal structure of GH5 was solved by Ramakrishnan et al. (1993), the primary DNA-binding site (namely, helix III plus a lysine at the beginning of the β-hairpin) was proposed to interact with one of the DNA exit regions, while the second binding site would bind to the other exit. This is likely since helix III probably lies in the major

groove of the DNA which is accessible in the DNA termini. From the nucleosome architecture, this type of binding would imply a distance of 25–30 Å between the two sites on the protein, with an angle of about 180°, and also a geometric constraint on the DNA terminal double-helical turns. An alternative model would keep one binding site on a DNA exit region but would place the second binding site around the dyad. In this case, helix III would lie in the major groove, either at the level of the DNA terminus or at the level of the central DNA gyre, where the major groove is exposed not at the dyad axis point but close to it. The other DNA exit would be indirectly protected, possibly by the stabilization of histone H3 (Belyavsky et al., 1980; Bavykin et al., 1990). This second model is less constraining for the DNA and more consistent with neutron scattering data [Lambert et al., 1991; see Schwabbe and Travers (1993) and references therein]. The distance and angle between the two DNA-binding sites would then be shorter and less critical. Our data are in favor of this model since our calculated distances and angles (Figure 4) allow this type of binding, but not the 180° type, for all possibilities of secondary binding site (*B1* or *C1*, *B5* or *C5*).

(H) Differences between Electrostatic Potentials. In addition to the differences in the relative position of the positive surfaces around GH1 and GH5, the main difference between the electrostatic potentials is located in the region of the loop between helices II and III. At the level of −2, only one surface can be identified, surrounding this loop in GH1 (see Figure 3). In GH5, such a negative potential does not exist. This may be significant because it is not due to the difference in conformation of both loops but rather to the presence in this region of two acidic residues (D38 and E40) nonexistent in GH5 while largely conserved in the GH1 family (Crane-Robinson & Ptitsyn, 1989). This negative potential could create a repulsion against DNA, thereby weakening the binding. However, in several DNA/protein complexes, acidic residues are found at the DNA/protein interface, neutralized by positive ions [e.g., Winkler (1992)]. This possibility must be considered for GH1 but would still lead to a difference in DNA binding in comparison with GH5.

In conclusion, this study shows that the globular domains of the linker histones H1 and H5 adopt very similar 3D structures. It also describes some structural and electrostatic differences which could form the basis for the observed functional differences of these homologous proteins.

ACKNOWLEDGMENT

We thank R. Gordon-Beresford for his help with the electrostatic potential calculations, Dr. M. Froeyen for providing his routine of visualization of electrostatic potential surfaces, and Dr. D. Van Belle for providing his routine of visualization of NOE constraints.

REFERENCES

Affolter, M., Côté, J., Renaud, J., & Ruiz-Carrillo, A. (1987) *Mol. Cell. Biol. 7*, 3663–3672.

Allan, J., Hartman, P. G., Crane-Robinson, C., & Aviles, F. X. (1980) *Nature 288*, 675–679.

Bavykin, S. G., Usachenco, S. I., Zalensky, A. O., & Mirzabekov, A. D. (1990) *J. Mol. Biol. 212*, 495–511.

Belyavsky, A. V., Bavykin, S. G., Goguadze, E. G., & Mirzabekov, A. D. (1980) *J. Mol. Biol. 139*, 519–536.

Brennan, R. G. (1991) *Curr. Opin. Struct. Biol. 1*, 80–88.

Brünger, A. T. (1992) *X-PLOR version 3.1 Manual*, Yale University, New Haven, CT.

Buckle, R. S., Maman, J. D., & Allan, J. (1992) *J. Mol. Biol. 223*, 651–659.

Cerf, C., Lippens, G., Muyldermans, S., Segers, A., Ramakrishnan, V., Wodak, S. J., Hallenga, K., & Wyns, L. (1993) *Biochemistry 32*, 11345–11351.

Clark, D. J., Hill, C. S., Martin, S. R., & Thomas, J. O. (1988) *EMBO J. 7*, 69–75.

Clark, K. L., Halay, E. D., Lai, E., & Burley, S. K. (1993) *Nature 364*, 412–420.

Clore, G. M., Gronenborn, A. M., Nilges, M., Sukumaran, D. K., & Zarbock, J. (1987) *EMBO J. 6*, 1833–1842.

Coles, L. S., Robins, A. J., Madley, L. K., & Wells, J. R. E. (1987) *J. Biol. Chem. 262*, 9656–9663.

Crane-Robinson, C., & Ptitsyn, O. B. (1989) *Protein Eng. 2*, 577–582.

Croston, G. E., Kerrigan, L. A., Lira, L. M., Marshak, D. R., & Kadonaga, J. (1991) *Science 251*, 643–649.

Delhaise, P., Van Belle, D., Bardiaux, M., Alard, P., Hamers, P., Van Cutsem, E., & Wodak, S. J. (1985) *J. Mol. Graphics 3*, 116–119.

Draves, P. H., Lowary, P. T., & Widom, J. (1992) *J. Mol. Biol. 225*, 1105–1121.

Garrard, W. T. (1991) *BioEssays 13*, 87–88.

Gilson, M. K., & Honig, B. H. (1987) *Nature 330*, 84–86.

Graziano, V., Gerchman, S. E., Schneider, D. K., & Ramakrishnan, V. (1994) *Nature 368*, 351–354.

Güntert, P., Berndt, K. D., & Wüthrich, K. (1993) *J. Biomol. NMR 3*, 601–606.

Habazettl, J., Cieslar, C., Oschkinat, H., & Holak, T. A. (1990) *FEBS Lett. 268*, 141–145.

Hendrickson, F. M., & Cole, R. D. (1994) *Biochemistry 33*, 2997–3006.

Hill, C. S., Martin, S. R., & Thomas, J. O. (1989) *EMBO J. 8*, 2591–2599.

Kumar, N. M., & Walker, I. O. (1980) *Nucleic Acids Res. 8*, 3535–3551.

Lambert, S., Muyldermans, S., Baldwin, J., Kilner, J., Ibel, K., & Wyns, L. (1991) *Biochem. Biophys. Res. Commun. 179*, 810–816.

Lippens, G., Hallenga, K., Van Belle, D., Wodak, S. J., Nirmala, N. R., Hill, P., Russell, K. C., Smith, D. D., & Hruby, V. J. (1993) *Biochemistry 32*, 9423–9434.

Live, D. H., Davis, D. G., Agosta, W. C., & Cowburn, D. (1984) *J. Am. Chem. Soc. 106*, 1939–1941.

Mirzabekov, A. D., Pruss, D. V., & Ebralidse, K. K. (1989) *J. Mol. Biol. 211*, 479–491.

Muyldermans, S., & Travers, A. A. (1994) *J. Mol. Biol. 235*, 855–870.

Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L., & Sweet, R. M. (1993) *Nature 362*, 219–223.

Schwabe, J. W. R., & Travers, A. A. (1993) *Curr. Biol. 3*, 628–630.

Segers, A., Muyldermans, S., & Wyns, L. (1991) *J. Biol. Chem. 266*, 1502–1508.

Staynov, D. Z., & Crane-Robinson, C. (1988) *EMBO J. 7*, 3685–3691.

Sun, J.-M., Wiaderkiewicz, R., & Ruiz-Carrillo, A. (1989) *Science 245*, 68–71.

Sun, J.-M., Ali, Z., Lurz, R., & Ruiz-Carrillo, A. (1990) *EMBO J. 9*, 1651–1658.

Thoma, F., Koller, T., & Klug, A. (1979) *J. Cell Biol. 83*, 403–427.

Thoma, F., Losa, R., & Koller, T. (1983) *J. Mol. Biol. 167*, 619–640.

Thomas, J. O., & Wilson, C. M. (1986) *EMBO J. 5*, 3531–3537.

Thomas, J. O., Rees, C., & Finch, J. T. (1992) *Nucleic Acids Res. 20*, 187–194.

Wellman, S. E., Sittman, D. B., & Chaires, J. B. (1994) *Biochemistry 33*, 384–388.

Winkler, F. K. (1992) *Curr. Opin. Struct. Biol. 2*, 93–99.

Zarbock, J., Clore, G. M., & Gronenborn, A. M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7628–7632.

Zlatanova, J. (1990) *Trends Biochem. Sci.15*,273–276.

Zlatanova, J., & Yaneva, J. (1991) *Mol. Biol. Rep. 15*, 53–56.