Complex of linker histone H5 with the nucleosome and its implications for chromatin packing

Li Fan and Victoria A. Roberts*

Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037

Edited by Robert M. Stroud, University of California, San Francisco, CA, and approved April 12, 2006 (received for review October 12, 2005)

Linker histones are essential for chromatin filament formation, and they play key roles in the regulation of gene expression. Despite the determination of structures of the nucleosome and linker histones, the location of the linker histone on the nucleosome is still a matter of debate. Here we show by computational docking that the globular domain of linker histone variant H5 (GH5) has three distinct DNA-binding sites, through which GH5 contacts the DNA at the nucleosome dyad and the linker DNA strands entering and exiting the nucleosome. Our results explain the extensive mutagenesis and crosslinking data showing that side chains spread throughout the GH5 surface interact with nucleosomal DNA. The nucleosome DNA contacts positively charged side chains that are conserved within the linker histone family, indicating that our model extends to linker histone–nucleosome interactions in general. Furthermore, our model provides a structural mechanism for formation of a dinucleosome complex specific to the linker histone H5, explaining its efficiency in chromatin compaction and transcription regulation. Thus, this work provides a basis for understanding how structural differences within the linker histone family result in functional differences, which in turn are important for gene regulation.

computational docking | DNA–protein interactions | DOT | linker DNA | winged-helix protein

The interaction of the linker histone with the nucleosome is an ongoing controversial issue (1–3). Histone H5 has been the focus of linker-histone-related studies in recent years. H5 consists of a central globular domain (GH5) that is essential for nucleosome binding and is flanked by basic N- and C-terminal tails (4). Binding of either H5 or GH5 to nucleosomes protects an additional 20 bp of linker DNA from micrococcal nuclease digestion (5). Early studies based on micrococcal nuclease digestion and DNase I footprinting proposed a symmetrical model in which GH5 contacts the dyad of the nucleosome and both the entering and exiting DNA duplexes (also called DNA arms) (6, 7). A "bridging" model was later proposed based on experiments that mapped the binding site of GH5 on mixed-sequence chicken chromatosomes by conjugating a crosslinking reagent to specific Ser-to-Cys substitutions (8). In this model, GH5 forms a bridge between one DNA arm and the dyad. A radically different "off-axis" model was developed from studies on a DNA fragment containing the *Xenopus borealis* somatic 5S RNA gene (9). Based on this model, GH5 is positioned ≈65 bp away from the dyad and is bound inside the DNA superhelix. The bridging and off-axis models imply there may be two equivalent linker histone-binding sites per nucleosome (1), but a 1:1 ratio of linker histone and nucleosome is observed (10, 11).

There have been multiple proposals for defining the DNA-binding sites on the linker histone. GH5 has a winged-helix fold (12), consisting of a three-helix bundle in which helices H2 and H3 are part of a helix–turn–helix motif that is followed by a β-hairpin, termed the wing (13). Possible modes of GH5 binding to the nucleosome have been proposed based on comparisons with transcription factors also in the winged-helix family (12, 14–16). Helix H3, often called the recognition helix, is very likely involved in DNA binding. Comparisons of crystallographic structures of DNA-bound winged-helix transcription factors (14, 16, 17), however, show a substantial variation in DNA positioning and mode of DNA binding. Thus, it may be risky to derive a model for DNA–GH5 interactions by assuming a specific mode of binding, such as insertion of the recognition helix into the DNA major groove (18).

In this report, we used computational docking to investigate the interaction of GH5 with the nucleosome. No assumptions of specific types of DNA–protein contacts were made. Instead, predictions of interactions were based on the sum of van der Waals and electrostatic intermolecular energies. The resulting predicted complex provides both a model for linker histone–nucleosome interactions in general and reveals interactions specific to GH5 that may be responsible for its ability to compact chromatin structure and to repress transcription.

Results and Discussion

Analysis of the GH5 Structure. We analyzed the known biochemical data in the context of the GH5 structure and found that GH5 is fundamentally different from winged-helix transcription factors. To select coordinates for analysis, we examined GH5 [Protein Data Bank (PDB) ID code 1HST] (12), which forms a dimer in the crystal structure. Molecules A and B of the dimer differ in the conformation of the winged loop. Molecule A has been typically used for modeling DNA–GH5 interactions (8, 12, 16, 18, 19). The winged loop of molecule A extends away from the molecule because of hydrophobic intermolecular contacts with the wing of another molecule A in the crystal lattice (12). This wing conformation exposes hydrophobic residues Ile-72 (from helix H3) and Phe-93 (at the base of the wing) to solvent. These residues would be part of a hydrophobic core between the wing and helix H3 in winged-helix proteins. In molecule B, there are no intermolecular contacts of the wing in the crystal. Instead, the top of the wing (residues 88–91) folds against the hydrophobic core, excluding Leu-72 and Phe-93 side chains from solvent. The wing conformation of molecule B is anchored by two hydrogen bonds between the Gln-83 side chain (at the base of the wing) and the main chain of Gly-88 at the top of the wing. Because the molecule B structure is less perturbed by crystal contacts, this molecule was used in our structure analysis and docking studies.

GH5 side chains that have been implicated in nucleosome binding by radiolabeling (20), mutagenesis (21, 22), and crosslinking experiments (8, 23) are widely dispersed over the protein's surface, as are the positively charged side chains conserved in the H1 family (Fig. 1*A*). This distribution of positively charged residues in GH5 results in positive electrostatic potential over most of the molecular surface. Virtually the entire electrostatic potential field created by GH5 is positive

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviation: NCP, nucleosome core particle.

*To whom correspondence should be sent at the present address: San Diego Supercomputer Center, University of California, San Diego, 9500 Gilman Drive #0505, La Jolla, CA 92093-0505. E-mail: vickie@sdsc.edu.

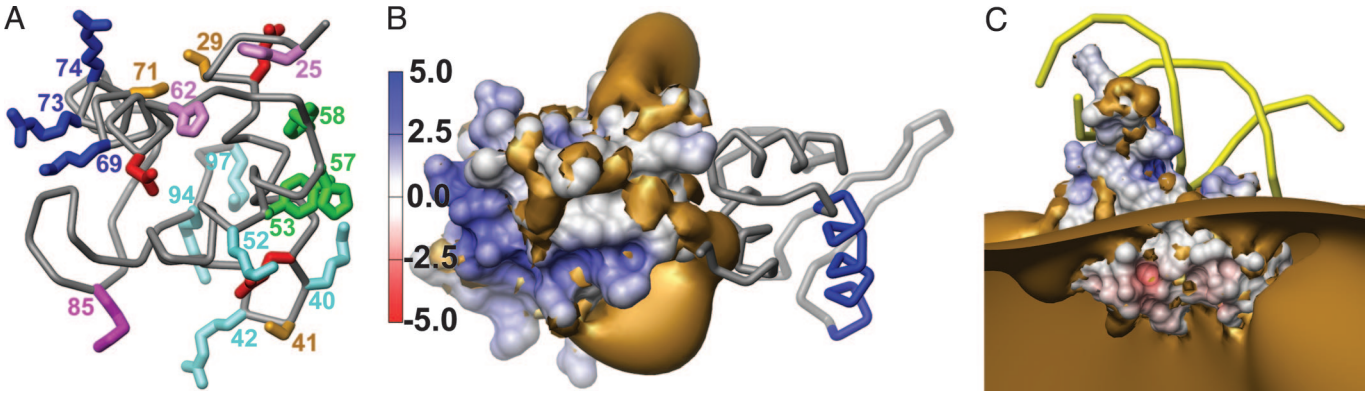

Fig. 1. DNA-contacting residues are widely distributed over the GH5 surface. (*A*) Side chains implicated in chromatosome binding include the essential Lys-85 (magenta) on the wing; helix H3 residues Lys-69, Arg-73, and Arg-74 (blue); and Lys-40, Arg-42, Lys-52, Arg-94, and Arg-97 (light blue). Residues His-25 and His-62 (lavender) (23) and residues Ser-29, Ser-41, and Ser-71 (gold) (8) crosslink to nucleosomal DNA. Unlikely to contact DNA are Tyr-53, His-57, and Tyr-58 (green), which lie in the GH5 dimer interface, and the three GH5 acidic residues (red). (*B*) Positive electrostatic potential extends out into solvent from most of the GH5 surface. The electrostatic potential (blue positive and red negative) at the solvent-accessible surface is mapped onto the GH5 molecular surface (oriented as in *A*). The backbone of the other GH5 molecule in the crystallographic dimer (12) is displayed with helix H3 in blue. (*C*) Human RFX1 (PDB ID code 1D7) is strongly dipolar. Positive electrostatic potential (blue) extends out from the half of the protein that includes the DNA-binding site. The phosphate backbone of the bound DNA is yellow. The gold surfaces in *B* and *C* represent where the electrostatic potential field is 0.0 kcal·mol^{-1}·e^{-1}. All figures are rendered with the graphics program AVS (Advanced Visual Systems, Waltham, MA) (46).

(Fig. 1*B*), as shown by examining the surface (gold in Fig. 1*B*) that represents the electrostatic potential at 0.0 kcal·mol^{-1}·e^{-1}. The only regions of negative electrostatic potential lie around the three acidic side chains (Glu-30, Glu-39, and Asp-65) of GH5 (under the gold surface) and extend only a short distance past the molecular surface into solvent. In contrast, the DNA-binding domain of the human transcription factor RFX1, which has been suggested as having electrostatic properties similar to GH5 (16), is strongly dipolar (Fig. 1*C*). The electrostatic potential surface at 0.0 kcal·mol^{-1}·e^{-1} (gold in Fig. 1*C*) divides the volume surrounding RFX1 into two regions: positive potential emanating from the face containing the DNA-binding site (upper half of Fig. 1*C*) and negative potential from the other half of the protein (lower half of Fig. 1*C*). The electrostatic differences between GH5 and winged-helix transcription factors may explain a property unique to GH5: its preference for binding four-way junctions (24) and the nucleosome over linear duplex DNA.

Docking B-DNA to GH5. To understand how the extensive positive potential created by GH5 influences DNA binding, we performed rigid-body dockings with DOT, a computational docking tool for macromolecular interactions (25). Studies on winged-helix transcription factors show that DOT is an effective tool for predicting protein–DNA interactions (17). DOT performs a complete search of all orientations between two macromolecules by systematically rotating and translating a moving molecule about a stationary molecule over all space. Interactions energies are computed as the sum of van der Waals and electrostatic terms (see *Methods*) for >60 billion configurations of the two molecules.

Docking linear B-DNA (12 bp) (the moving molecule) to GH5 (the stationary molecule) identified three distinct DNA-binding sites in the 200 top-ranked solutions (Fig. 2*A*), with each site represented in the top 30 solutions. For B-DNA solutions docked at site I (11 of the 30 solutions), the side chains of highly conserved Lys-69 and Arg-73 in helix H3 and Arg-47 in helix H2 are inserted into the DNA major groove. Arg-74 and Lys-52 may also contact the DNA backbone. For B-DNA solutions docked at site II (12 of the 30 solutions) (Fig. 2*B*), Lys-85 is inserted in the DNA major groove, providing the primary protein–DNA interaction, with additional contacts of Arg-42, Lys-82, and Arg-94 to the DNA backbone. B-DNA solutions at site III (5 of the 30 solutions) were

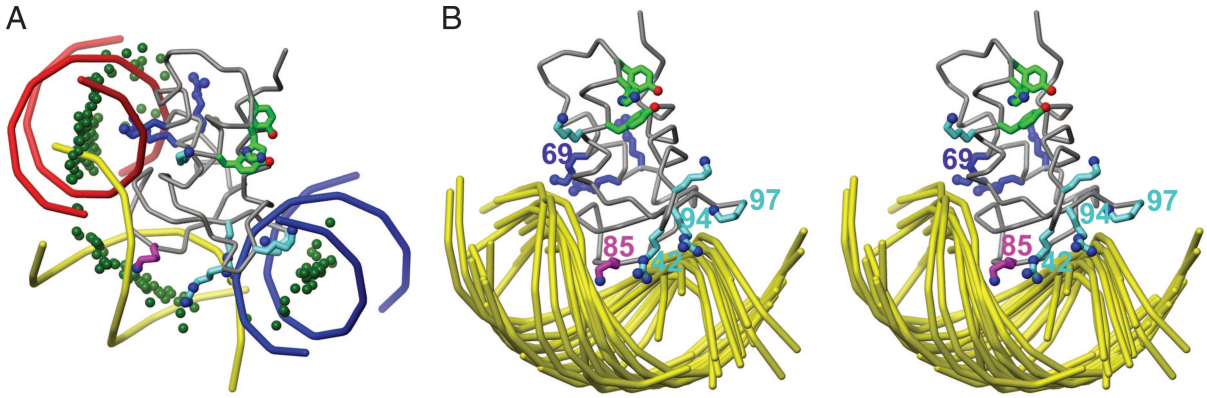

Fig. 2. Linear B-DNA fragments docked to GH5. (*A*) GH5 has three distinct DNA-binding sites. The 200 top-ranked DNA solutions (centers shown as green spheres) found by DOT cluster at site I (representative docked DNA fragment, rank 3, red DNA backbone), site II (rank 1, yellow), and site III (rank 5, blue). Side chains of GH5 are colored as in Fig. 1*A*. (*B*) DNA binding at site II (stereo pair). All DNA fragments docked at site II in the 30 top-ranked solutions (shown by their phosphate backbone, yellow) show similar orientations in which the Lys-85 side chain (magenta with blue NZ atom) lies in the DNA major groove.

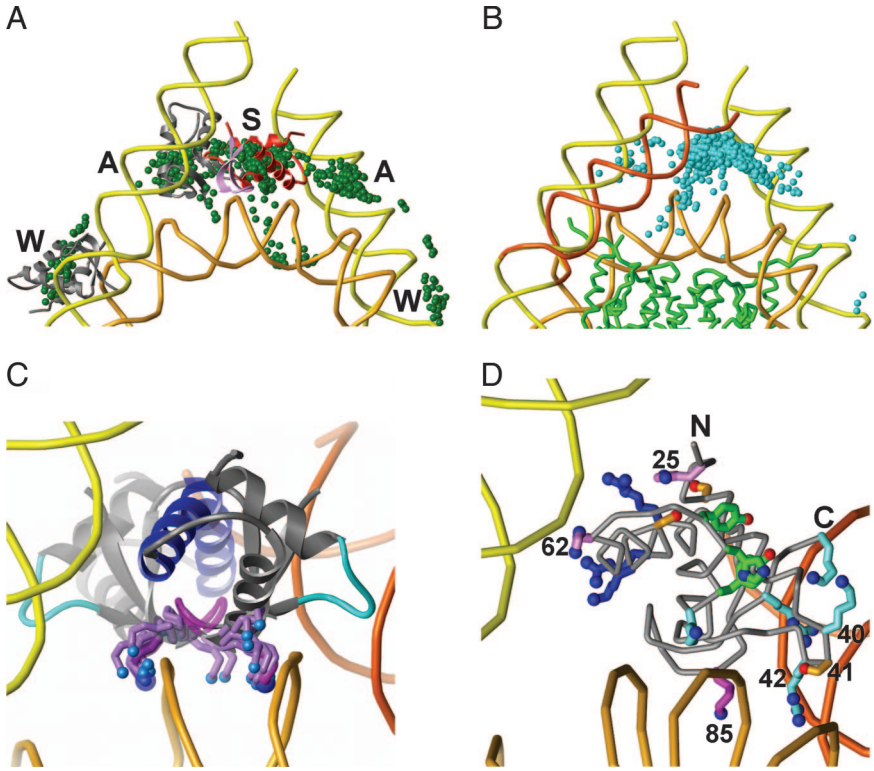

Fig. 3. GH5 docked to the nucleosome. (*A*) Nucleosome model with both DNA arms extended [shown are the DNA arms (yellow) and DNA at the dyad (light orange)]. The spread of the 1,000 top-ranked GH5 solutions (centers shown as green spheres) is also seen in the top 30 solutions, which include dockings far from the nucleosome dyad axis (represented by rank 15, site W, gray), near the dyad axis but contacting only one arm (rank 2, site A, gray), and over the dyad axis (rank 10, site S, red with lavender wing). (*B*) Nucleosome model with one DNA arm bent (orange). The 1,000 top-ranked GH5 solutions (centers shown as aqua spheres) are concentrated over the dyad axis. For comparison, the left straight arm (yellow) also is displayed. (*C*) Lys-85 lies at the nucleosome dyad axis. In 27 of the 30 top-ranked solutions, the Lys-85 side chain (lavender with blue NZ atom) lies in the DNA minor groove at the dyad axis (represented by GH5 rank 1, gray with blue helix H3; a magenta wing; and a light blue site III loop) or slightly to one side of the phosphate backbone at the dyad axis (represented by GH5 rank 5, same coloring but paler). (*D*) Interactions of GH5 side chains with the nucleosome. The top-ranked GH5 solution has Lys-69, Arg-73, and Arg-74 (blue) from helix H3, site I, contacting one arm, with His-25 and His-62 (lavender) and Ser-29 and Ser-71 (gold with red OG) nearby. Lys-85 (magenta) and its wing (site II) are centered in the DNA minor groove at the dyad axis; and Arg-40, Lys-42, Arg-94, and Lys-97 (light blue, right, site III) contact the other DNA arm. The Ser-41 side chain (gold, lower right) extends toward DNA at the dyad axis. The N and C termini of GH5 are indicated.

more diverse, but all had the DNA backbone clamped between Lys-40 and Lys-97. Usually only two GH5 DNA-binding sites are discussed (12, 19, 22): a primary site contributed by helix H3 and Lys-85 and a secondary site containing Lys-40, Arg-42, Lys-52, and Arg-94. These analyses were based on molecule A of the GH5 structure in which the perturbed wing conformation causes the Lys-85 side chain to extend toward helix H3. Our docking using molecule B separates the so-called primary site into sites I and II (red and yellow in Fig. 2*A*). A distinct DNA-binding site centered around Lys-85 is consistent with Lys-85 being singly required to maintain protection of 20 bp of linker DNA (21) and Lys-85 being maximally protected from proteolysis when GH5 is bound to nucleosomal DNA (20). In addition, three GH5 DNA-binding sites are consistent with a long-held view that the linker histone contacts the nucleosomal dyad and both the entering and the exiting linker DNA (6). Three distinct DNA-binding sites have been previously proposed (26) based on an early NMR structure of GH5 (27) and sequence conservation among members of linker histone H1/H5 family.

Docking GH5 to the Nucleosome. To investigate GH5 interactions with nucleosomal DNA, GH5 (the moving molecule) was docked to a nucleosome model derived from the nucleosome core particle (NCP) crystal structure (28). Our nucleosome model used the half of the NCP that contains the nucleosome dyad (see *Methods*), which included residues from histones H2A, H3, and H4 and the associated DNA. Because GH5 protects ≈20 bp DNA beyond the 146-bp DNA in the NCP, we extended each DNA end of the NCP by 16 bp of linear B-DNA (see *Methods*). Phosphate backbone geometry and base stacking were maintained from the NCP DNA to the added DNA arms. This nucleosome model has all GH5 binding positions proposed by the current three models. When GH5 was docked to this nucleosome model, the top 1,000 GH5 solutions clustered at three distinct sites named S, A, and W (Fig. 3*A*), corresponding to the three proposed models for H5 binding (2, 29). The energies at these three sites are similar, suggesting that these three binding modes could be observed experimentally. Certain conditions or experimental methods may favor one or another binding mode, possibly explaining why different groups have obtained alternative results. GH5 bound at site S appears to be consistent with the symmetrical model, in which GH5 contacts the dyad DNA and both DNA arms, resulting in symmetrical protection of 2 × 10-bp linker DNA (6, 7). Close examination of GH5 (red in Fig. 3*A*) at site S, however, indicated that, although the critical Lys-85 side chain lies near the dyad DNA, GH5 cannot simultaneously contact both DNA arms because they are too far apart. GH5 located at site A is consistent with the bridging model in which GH5 binds the dyad and one DNA arm (8). This GH5 position, however, appears capable of contacting no more than 15 bp DNA. GH5 located at site W seems to be consistent with the third off-axis model (9) in which GH5 binds

far away from the dyad, but no linker DNA would be protected by GH5 unless the other DNA arm bends over and interacts with GH5. Thus, in all three cases, GH5 can protect only 20 bp of linker DNA if the DNA arms approach more closely, presumably by bending of the linker DNA (30). We note that local conformational changes in the DNA or rearrangement of the histone core upon GH5 binding, neither of which is included in our model, could lead to the protection of additional 20 bp or to the stabilization of alternative GH5 binding modes.

To bring the DNA arms closer, we built a second nucleosome model in which one DNA arm continued the observed curvature of the NCP-bound DNA ≈11 bp from one end, resulting in an approximate 18° bend (Fig. 3*B*). After we completed our docking studies, a low resolution (9 Å) crystal structure of the tetranucleosome was reported (31) in which the two linker DNA arms extend from the nucleosome at different orientations: one arm oriented in a fashion similar to those of the single NCP structure and the other bent more toward the dyad DNA. These orientations are similar to those in our model (Fig. 5, which is published as supporting information on the PNAS web site), verifying that the linker DNA arms in our second nucleosome model have reasonable orientations. Docking GH5 to this nucleosome model improved the best energy by ≈3 kcal/mol. Of the top 1,000 solutions, 90% form a cluster over the dyad (Fig. 3*B*). The best-energy solution outside this cluster (rank 20) is 2.8 kcal/mol less favorable than the top-ranked solution. In 27 of the top 30 solutions, Lys-85 lies either in the minor groove of the dyad (when helix H3 contacts the straight DNA arm) or just to one side of the minor groove phosphate backbone (when helix H3 contacts the bent DNA arm) (Fig. 3*C*). Thus, Lys-85 shows a unique placement, consistent with its key role (20, 21). In these 27 solutions, GH5 helix H3 side chains Lys-69, Arg-73, and Arg-74 contact ≈15 bp of one DNA arm, whereas the loop that includes Lys-40 and Arg-42 and the side chains of Arg-94 and Lys-97 contact ≈5 bp of the second DNA arm. Together, these contacts account for the observed protection of 20 bp of linker DNA against nuclease digestion (5, 6). GH5 uses the same three sites to interact with nucleosomal DNA as found in our B-DNA fragment docking. In the more complex nucleosomal environment, however, the local DNA orientation at each site is different from the optimal alignment found for a single DNA fragment (Fig. 2). In the top-ranked GH5–nucleosome complex (Fig. 3*D*), the three sites have key contacts with the DNA minor groove, a common feature of nonsequence-specific DNA-histone interactions (28).

These GH5–DNA interactions are likely to apply to linker histone H1. The globular domain of chicken erythrocyte linker histone H1 has ≈40% sequence identity with GH5 and retains a very similar three-dimensional structure (32). The wing, including the key residue Lys-85, has high sequence identity (>87%). Positively charged Lys-69 and Arg-73 of helix H3 are well conserved by Lys-47 and Lys-51 in GH1. The cluster of Lys-40, Arg-42, Arg-94, and Lys-97 also is conserved in GH1 through Lys-18, Arg-20, Arg-72, and Lys-75. Sequence conservation of these basic residues and the wing region within the H1/H5 family (26) indicates that our results apply to linker histone–DNA interactions in general.

The top-ranked GH5–nucleosome complex shows excellent agreement to known biochemical data. Contacts are made by the positively charged residues implicated in nucleosome binding, including essential Lys-85, Lys-69, and Arg-73 of helix H3 and Lys-40 and Arg-42 (Fig. 3*D*). Perhaps the best validation of the bound GH5 model is its agreement with biochemical data concerning side chains with little influence on the electrostatic interactions that guided the docking of GH5. His-25 and His-62 (neutral in docking) lie near the end of the DNA arm contacted by helix H3 (Fig. 3*D*), consistent with their crosslinking to the DNA arm (23). The positions (Fig. 3*D*) of Ser-29 and Ser-71 are consistent with their crosslinks near the end of the DNA arm (8). Although the backbone of Ser-41 lies near one DNA arm in our model, its side chain extends toward the dyad (Fig. 3*D*), allowing the azidophenacyl derivative of Ser-41 to form the observed crosslink with DNA close to the dyad (8). This crosslink was interpreted as evidence for adjacent residues Lys-40 and Arg-42 contacting the dyad DNA (8), leading to the bridging model in which GH5 contacts only one linker DNA arm and the dyad DNA (8). In our model, the side chains of Lys-40 and Arg-42 extend toward the linker DNA arm not contacted by helix H3, consistent with the symmetrical model in which GH5 contacts both DNA arms and the dyad DNA. Neither the residues in the GH5 dimer interface nor the three acidic residues (Fig. 1*A*) lie close to the nucleosomal DNA. Thus, our model for GH5 docked to the nucleosome includes all of the contacts suggested by experimental data and explains how such a large percentage of the GH5 surface is involved in nucleosome contacts. The model also explains the transient protection of ≈12 bp of linker DNA by the GH5 mutant in which residues 40, 42, 52, and 94 are replaced by Ala (22). This mutant would retain adjacent DNA-binding sites I and II, sufficient for weak binding to one arm and the dyad.

The N- and C-terminal ends of GH5 (indicated in Fig. 3*D*) extend away from the dyad DNA. This positioning of the C terminus is consistent with electrophoretic mobility and electron microscopy studies indicating that the basic C-terminal tail interacts with both entering and exiting linker DNA (10).

Nucleosome Dimerization by GH5. Linker histone H5 incurs greater compaction of chromatin and is more inhibitory toward transcription than H1 (33). Unlike H1, H5 and GH5 can form dimers in solution (34). The GH5 dimer is likely that found in the GH5 crystal (12), formed mainly by three aromatic residues (Tyr-53, His-57, and Tyr-58) in each monomer (Fig. 1*A*) that create an electrostatically neutral interaction surface. Two of these residues (Tyr-53 and His-57) are replaced by Ala-31 and Gly-35 in H1 (32), explaining why H1 does not form a stable dimer. To investigate the possible role of H5 dimerization in chromatin compaction, we superposed the predicted GH5–nucleosome complex onto each GH5 monomer of the crystallographic GH5 dimer. Remarkably, the resulting dinucleosome complex has no steric clashes between NCPs (Fig. 4*A*). Many of the 30 top-ranked docked complexes gave similar models, with minor variations in the relative orientation of the two linked nucleosomes. This structural model suggests how the replacement of the monomeric H1 linker histone by the dimeric H5 linker histone can induce greater chromatin compaction, leading to transcription repression. Linker histone H5 may, through the dimerization, bring two adjacent nucleosomes closer together (zigzag chain) (Fig. 4*B*), consistent with the popular model for partially condensed chromatin suggested by many experiments (35, 36) and the recent tetranucleosome structure (31). Two distant segments of nucleosomes upon H5 dimerization could create one dense array (Fig. 4*C*), as found by gel electrophoresis experiments in which two 12-nucleosome chains dimerized in the presence of histone H5 (37). Thus, from our predicted GH5–nucleosome complex, we have derived a specific structural mechanism for the compaction of chromatin structure by histone H5.

Our results on the GH5–nucleosome interaction explain extensive experimental data obtained over the past two decades by many different methods under various conditions and put them into a single structural context. The results provide a common theory about how linker histones interact with the nucleosome core and the linker DNA and explain the different properties of linker histone variants. The combination of computational docking and structural analysis used here to investigate GH5–DNA interactions at different levels, from linear DNA fragments to dinucleosome condensation, shows promise for predicting and interpreting complex DNA–protein interactions (38).

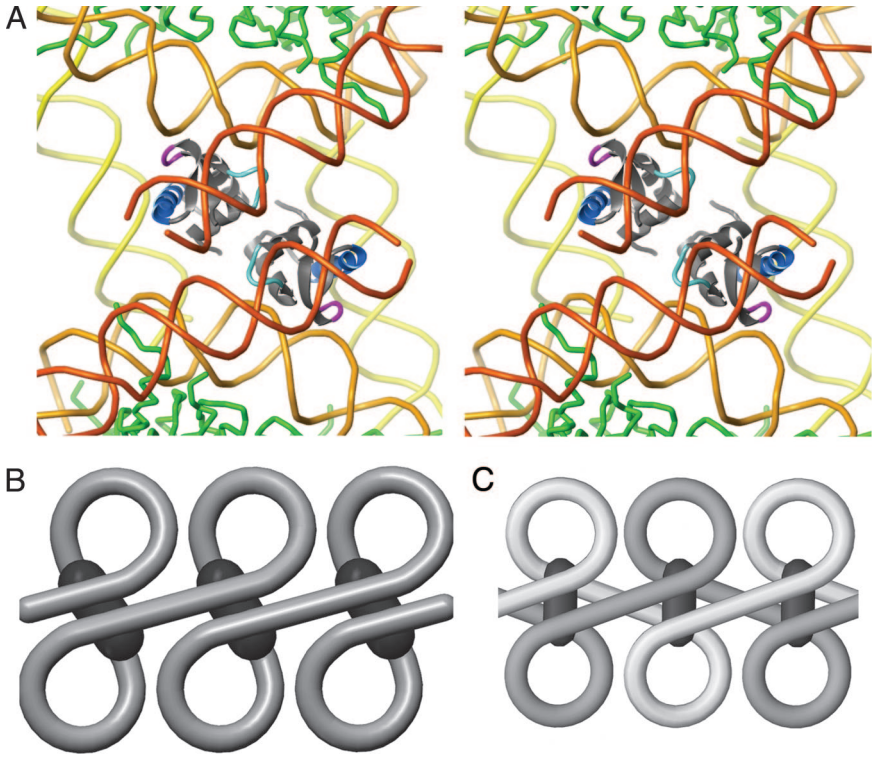

Fig. 4. Models for chromatin condensation by GH5 dimerization. (*A*) The dinucleosome model in which two copies of the GH5–nucleosome complex were superposed on the GH5 dimer (stereo pair). Depending on the angle of the DNA arms, the two arms (one from each nucleosome) can crossover (front) or not crossover (back). (*B*) Model of GH5 dimerization for condensing two adjacent nucleosomes. The geometry for each pair is as in *A*, but the two DNA arms meet in back to form a continuous DNA strand. (*C*) Model of GH5 dimerization for condensing two distant segments (light and dark gray) of nucleosomes. In the model, the two DNA arms crossover both in front and in back for each dinucleosome. Both models shown in *B* and *C* use ≈200 bp per nucleosome. Duplex DNA is represented by gray tubes, and GH5 dimers are indicated by black ellipsoids. The width of the DNA tube is shown to scale within the NCP. Both models are shown as lying in a plane, but adjacent dinucleosomes are likely twisted relative to each other.

Methods

Preparation of Protein and DNA Coordinates. Coordinates for GH5 were taken from molecule B of the GH5 dimer structure (PDB ID code 1HST) (12). Linear B-DNA fragments with lengths of 12 bp and 18 bp were constructed with the "bdna" command of the NUCLEIC ACID BUILDER (NAB) program (39). Sequences were 5′-GTTCAGCTGAAT-3′ for the 12-bp fragment and 5′-TTCAGCGTTCAGCTGAAT-3′ for the 18-bp fragment.

To build the nucleosome models, we started with the NCP structure (PDB ID code 1AOI) (28) and extracted the half of the structure that includes the nucleosome dyad (residues 38–63 and 99–135 of H3 and H3′, residues 20–60 of H4, residues 25–56 of H4′, residues 107–118 of H2A, and residues 106–118 of H2A′ and the associated DNA). The size of the resulting model provided ample search space within the cubic grid (128 Å on each side, see below) used in the docking calculations to fit GH5 around the nucleosome model. The H3 N-terminal tail (residues preceding 38) was not included in the model because of its conformational variability observed in crystal structures (28, 40).

To build the first nucleosome model with extended DNA arms, the linear 18-bp DNA fragments made with NAB were superposed on the last 2 bp of each DNA terminus of the NCP. The resulting 16-bp extensions continued the phosphate backbone geometry and base stacking observed in the NCP structure. For the second nucleosome model, we continued the curve of the DNA in the NCP near one terminus of the DNA to create one bent arm. To build this bent DNA arm, we searched within the NCP DNA for a DNA fragment that would maintain the curvature of the NCP-bound DNA near the DNA terminus at the position I134–I138/J155–J159 yet continue appropriate base stacking, maintain reasonable geometry for the phosphate backbone, and keep the terminal DNA in the plane of the NCP DNA. The segment I72–I76/J217–J221 satisfied these criteria. The DNA fragment I134–I146/J147–J159 was replaced by I72–I84/J209–J221 by superposition of corresponding phosphate atoms. The DNA arm was then extended by 16 bp by superposition of an 18-bp linear DNA fragment, as described above.

Polar hydrogen atoms were added to all molecules with the computer graphics program INSIGHT (Accelrys, San Diego). The imidazole ring of histidine residues was kept neutral, with a single proton on Nε.

The center for each moving molecule in the DOT calculation (see below) was defined as the center of geometry of the non-hydrogen atoms, which is the midpoint between the minimum and maximum values in the *x*, *y*, and *z* directions.

The DOT Calculation. In the DOT calculation, one molecule (the moving molecule) is systematically moved about a second molecule (the stationary molecule) in a complete translational and rotational search (25, 41). Interaction energies for all configurations of the two molecules are evaluated as correlation functions, which are efficiently computed with fast Fourier transforms. The properties of both molecules are mapped onto grids. For each orientation of the moving molecule, the moving molecule is centered at each grid point and the interaction energy is calculated as the sum of the electrostatics and van der Waals terms. A cubic grid 128 Å on a side with 1-Å grid spacing (≈2.1 million points) and a set of 28,800 orientations for the moving molecule (≈7.5° spacing) gave over 60 billion configurations of the two molecules. The top several thousand configurations were retained. DOT also gives a complete grid

of interaction energies with the energy mapped to the grid point at which the moving molecule is centered, so that the distribution of favorable energy configurations can be easily determined (17, 25). Each docking calculation took ≈4 h, using 25 Sun Ultra-10 desktop workstations (Sun Microsystems, Mountain View, CA) running in parallel. The DOT program is distributed by the Computational Center for Macromolecule Structure at the San Diego Supercomputer Center.

Van der Waals Energy Term for the DOT Calculation. The DOT van der Waals energy is proportional to the number of moving molecule atoms that lie within a favorable interaction layer surrounding the stationary molecule (41). The shape potential of the stationary molecule is represented by an excluded volume, defined as all grid points inside the molecular surface calculated by the program MS (42) or MSMS (43), surrounded by a 3.0-Å favorable layer (17). The shape of the moving molecule is represented by its atomic centers, including those of polar hydrogen atoms. Thus, a moving molecule atom can lie as close as a van der Waals radius from an atomic center of the stationary molecule. This soft fit allows for grid effects and small conformational changes that may be induced upon intermolecular interactions. Each moving molecule atom that lies in the favorable region surrounding the stationary molecule contributes −0.1 kcal/mol to the van der Waals energy. A configuration is eliminated if any moving molecule atom lies within the stationary molecule's excluded core.

Electrostatic Energy Term for the DOT Calculation. The electrostatic energy in DOT is calculated as the set of point charges representing the moving molecule placed within the electrostatic potential of the stationary molecule (41). Partial atomic charges for the molecules were taken from the AMBER library that includes polar hydrogen atoms (44). The stationary molecule was positioned on the cubic grid exactly as in the shape potential calculation. The electrostatic potential was then calculated with the program UHBD (45), which uses finite-difference methods to solve the linearized Poisson–Boltzmann equation, thereby taking solvent and ionic strength effects into account. A dielectric of 3 for the protein, a dielectric of 80 for the surrounding environment, an ion exclusion radius of 1.4 Å, and an ionic strength of 150 mM NaCl were used. The electrostatic potential of the stationary molecule was clamped so that the values at all grid points lay within the range of maximum negative and positive electrostatic potential values observed at the molecule's solvent-accessible surface (out 1.4 Å from the molecular surface). This modification makes the electrostatic potential compatible with the approximate van der Waals potential (17). The ranges of the electrostatic potential were -1.5 to $+3.5$ kcal·mol^{-1}·e^{-1} for GH5 as the stationary molecule and -4.5 to $+4.5$ kcal·mol^{-1}·e^{-1} for the nucleosome as the stationary molecule.

The Dinucleosome Complex. To build the dinucleosome complex, two copies of the top-ranked GH5–nucleosome complex were superposed onto the structure of the GH5 dimer. The C^{α} atoms of the three GH5 helices (residues 27–39, 48–56, and 64–81) were used for the superposition. These helical regions are very similar in molecules A and B of the dimer (rms deviation = 0.31 Å). Residues 53–58 involved in the dimer interface show an excellent fit with this superposition.

We thank Prof. Karolin Luger for comments and critical reading of the manuscript and Michael E. Pique for help preparing Fig. 4 *B* and *C*. L.F. thanks John Tainer for his support. This work was supported by National Science Foundation Grant DBI 9904559, National Institutes of Health Grant 5R01GM070996, and National Cancer Institute Grant CA81967.

1. Thomas, J. O. (1999) *Curr. Opin. Cell Biol.* **11,** 312–317.
2. Travers, A. (1999) *Trends Biochem. Sci.* **24,** 4–7.
3. Vignali, M. & Workman, J. L. (1998) *Nat. Struct. Biol.* **5,** 1025–1028.
4. Crane-Robinson, C., Dancy, S. E., Bradbury, E. M., Garel, A., Kovacs, A. M., Champagne, M. & Daune, M. (1976) *Eur. J. Biochem.* **67,** 379–388.
5. Puigdomenech, P., Jose, M., Ruiz-Carrillo, A. & Crane-Robinson, C. (1983) *FEBS Lett.* **154,** 151–155.
6. Simpson, R. T. (1978) *Biochemistry* **17,** 5524–5531.
7. Staynov, D. Z. & Crane-Robinson, C. (1988) *EMBO J.* **7,** 3685–3691.
8. Zhou, Y. B., Gerchman, S. E., Ramakrishnan, V., Travers, A. & Muyldermans, S. (1998) *Nature* **395,** 402–405.
9. Hayes, J. J., Pruss, D. & Wolffe, A. P. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 7817–7821.
10. Hamiche, A., Schultz, P., Ramakrishnan, V., Oudet, P. & Prunell, A. (1996) *J. Mol. Biol.* **257,** 30–42.
11. Thomas, J. O. & Khabaza, A. J. (1980) *Eur. J. Biochem.* **112,** 501–511.
12. Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L. & Sweet, R. M. (1993) *Nature* **362,** 219–223.
13. Gajiwala, K. S. & Burley, S. K. (2000) *Curr. Opin. Struct. Biol.* **10,** 110–116.
14. Clark, K. L., Halay, E. D., Lai, E. & Burley, S. K. (1993) *Nature* **364,** 412–420.
15. Cirillo, L. A., McPherson, C. E., Bossard, P., Stevens, K., Cherian, S., Shim, E. Y., Clark, K. L., Burley, S. K. & Zaret, K. S. (1998) *EMBO J.* **17,** 244–254.
16. Gajiwala, K. S., Chen, H., Cornille, F., Roques, B. P., Reith, W., Mach, B. & Burley, S. K. (2000) *Nature* **403,** 916–921.
17. Roberts, V. A., Case, D. A. & Tsui, V. (2004) *Proteins* **57,** 172–187.
18. Bharath, M. M., Chandra, N. R. & Rao, M. R. (2003) *Nucleic Acids Res.* **31,** 4264–4274.
19. Goytisolo, F. A., Gerchman, S. E., Yu, X., Rees, C., Graziano, V., Ramakrishnan, V. & Thomas, J. O. (1996) *EMBO J.* **15,** 3421–3429.
20. Thomas, J. O. & Wilson, C. M. (1986) *EMBO J.* **5,** 3531–3537.
21. Buckle, R. S., Maman, J. D. & Allan, J. (1992) *J. Mol. Biol.* **223,** 651–659.
22. Duggan, M. M. & Thomas, J. O. (2000) *J. Mol. Biol.* **304,** 21–33.
23. Mirzabekov, A. D., Pruss, D. V. & Ebralidse, K. K. (1990) *J. Mol. Biol.* **211,** 479–491.
24. Varga-Weisz, P., Zlatanova, J., Leuba, S. H., Schroth, G. P. & van Holde, K. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 3525–3529.
25. Mandell, J. G., Roberts, V. A., Pique, M. E., Kotlovyi, V., Mitchell, J. C., Nelson, E., Tsigelny, I. & Ten Eyck, L. F. (2001) *Protein Eng.* **14,** 105–113.
26. Crane-Robinson, C. & Ptitsyn, O. B. (1989) *Protein Eng.* **2,** 577–582.
27. Clore, G. M., Gronenborn, A. M., Nilges, M., Sukumaran, D. K. & Zarbock, J. (1987) *EMBO J.* **6,** 1833–1842.
28. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. (1997) *Nature* **389,** 251–260.
29. Widom, J. (1998) *Curr. Biol.* **8,** R788–R791.
30. Butler, P. J. & Thomas, J. O. (1998) *J. Mol. Biol.* **281,** 401–407.
31. Schalch, T., Duda, S., Sargent, D. F. & Richmond, T. J. (2005) *Nature* **436,** 138–141.
32. Cerf, C., Lippens, G., Ramakrishnan, V., Muyldermans, S., Segers, A., Wyns, L., Wodak, S. J. & Hallenga, K. (1994) *Biochemistry* **33,** 11079–11086.
33. Sun, J. M., Ali, Z., Lurz, R. & Ruiz-Carrillo, A. (1990) *EMBO J.* **9,** 1651–1658.
34. Carter, G. J. & van Holde, K. (1998) *Biochemistry* **37,** 12477–12488.
35. Rydberg, B., Holley, W. R., Mian, I. S. & Chatterjee, A. (1998) *J. Mol. Biol.* **284,** 71–84.
36. Bednar, J., Horowitz, R. A., Grigoryev, S. A., Carruthers, L. M., Hansen, J. C., Koster, A. J. & Woodcock, C. L. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 14173–14178.
37. Carruthers, L. M., Bednar, J., Woodcock, C. L. & Hansen, J. C. (1998) *Biochemistry* **37,** 14776–14787.
38. Adesokan, A. A., Roberts, V. A., Lee, K. W., Lins, R. D. & Briggs, J. M. (2004) *J. Med. Chem.* **47,** 821–828.
39. Macke, T. & Case, D. A. (1998) in *Molecular Modeling of Nucleic Acids*, eds. Leontes, N. & Santa, L. J. (Am. Chem. Soc., Washington, DC), pp. 379–393.
40. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. (2002) *J. Mol. Biol.* **319,** 1097–1113.
41. Ten Eyck, L. F., Mandell, J. G., Roberts, V. A. & Pique, M. E. (1995) in *The 1995 ACM/IEEE Supercomputing Conference*, eds. Hayes, A. & Simmons, M. (IEEE Comp. Soc. Press, San Diego), pp. 1–12.
42. Connolly, M. L. (1983) *Science* **221,** 709–713.
43. Sanner, M. F., Olson, A. J. & Spehner, J. C. (1996) *Biopolymers* **38,** 305–320.
44. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, J. S. & Weiner, P. (1984) *J. Am. Chem. Soc.* **106,** 765–784.
45. Gilson, M., Davis, M., Luty, B. & McCammon, J. (1993) *J. Phys. Chem.* **97,** 3591–3600.
46. Upson, C., Faulhaber, T. J., Kamins, D., Laidlaw, D., Schlegel, D., Vroom, J., Gurwitz, R. & van Dam, A. (1989) *IEEE Comp. Graphics Appl.* **9,** 30–42.