Crystal structure of the nucleosome core particle at 2.8 Å resolution

Karolin Luger, Armin W. Mäder, Robin K. Richmond, David F. Sargent & Timothy J. Richmond

Institut für Molekularbiologie und Biophysik ETHZ, ETH-Hönggerberg, CH-8093 Zürich, Switzerland

The X-ray crystal structure of the nucleosome core particle of chromatin shows in atomic detail how the histone protein octamer is assembled and how 146 base pairs of DNA are organized into a superhelix around it. Both histone/histone and histone/DNA interactions depend on the histone fold domains and additional, well ordered structure elements extending from this motif. Histone amino-terminal tails pass over and between the gyres of the DNA superhelix to contact neighbouring particles. The lack of uniformity between multiple histone/DNA-binding sites causes the DNA to deviate from ideal superhelix geometry.

DNA in chromatin is organized in arrays of nucleosomes[1]. Two copies of each histone protein, H2A, H2B, H3 and H4, are assembled into an octamer that has 145–147 base pairs (bp) of DNA wrapped around it to form a nucleosome core (of relative molecular mass 206K). This highly conserved nucleoprotein complex occurs essentially every 200 ± 40 bp throughout all eukaryotic genomes[2]. The repeating nucleosome cores further assemble into higher-order structures which are stabilized by the linker histone H1 and these compact linear DNA overall by a factor of 30–40. The nucleosome (nucleosome core, linker DNA and H1) thereby shapes the DNA molecule both at the atomic level through DNA bending, and on the much larger scale of genes by forming higher-order helices[3]. The nucleosome, in its role as the principal packaging element of DNA within the nucleus, is the primary determinant of DNA accessibility.

The physical properties of nucleosomes depend on solution conditions such as ionic strength, divalent-ion concentration, and on histone-modification state[4]. For example, the variability of nucleosome spacing within *in vitro*-assembled arrays demonstrates the underlying electrostatic component of the interaction of DNA with the basic histone proteins[5]. The characteristics of an individual nucleosome, however, depend on the actual DNA sequence incorporated, and this can have functional significance for specific gene promoters[6]. Positioning of nucleosomes along the DNA molecule is in part governed by DNA sequence preferences, but once formed, a nucleosome may not be entirely fixed in position[7]. The statistical preference observed for the DNA minor groove to face the histone octamer at (A + T)-rich sequences indicates that certain sequences will impart a rotational orientation to the DNA double helix when bound in a nucleosome[8]. The variation in this preference along the length of nucleosomal DNA and the intrinsic sequence-dependent bendability of DNA implies that strong translational settings can also occur. Both types of positioning have been detected with base-pair accuracy in the mouse mammary-tumour virus long terminal repeat[9].

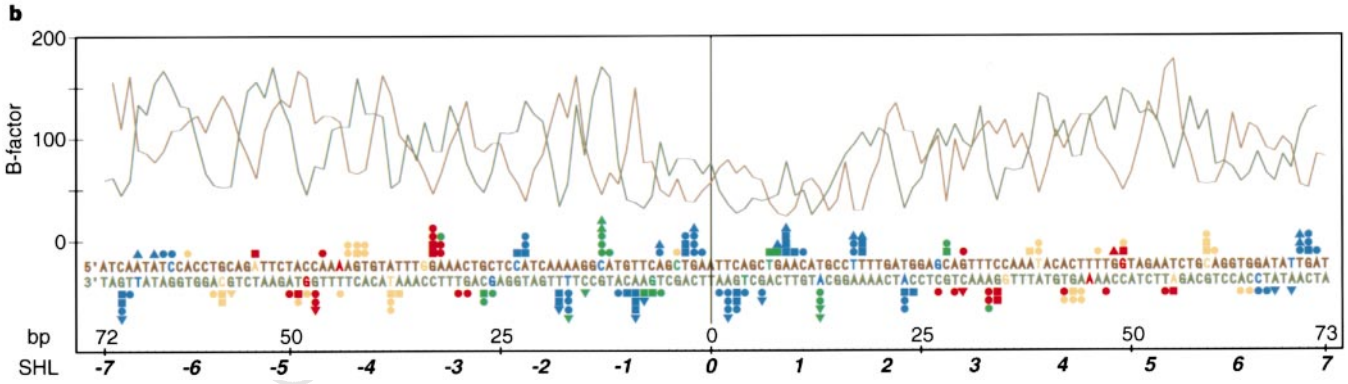
Recombination, replication, mitotic condensation and transcription involve the chromatin substrate and are thus affected by its structure. The generally repressive nature of chromatin structure has long been appreciated in transcription regulation[10]. Chromatin organization can facilitate the activation of specific genes[11,12]. Evidence supporting two themes is accumulating. First, histones collaborate with transcription factors to provide for their own removal or structural modification, resulting in gene derepression. This may occur by stepwise invasion of DNA-binding proteins entering at the terminal segments of the nucleosomal DNA[13]. This mechanism is apparently insufficient in general, however, because several ATP-dependent chromatin remodelling factors have been discovered that cause the core histones partially to lose or alter their grip on DNA[14]. Second, the bending and supercoiling of DNA on a nucleosome can promote binding of transcription factors and augment interactions between different factors[15,16].

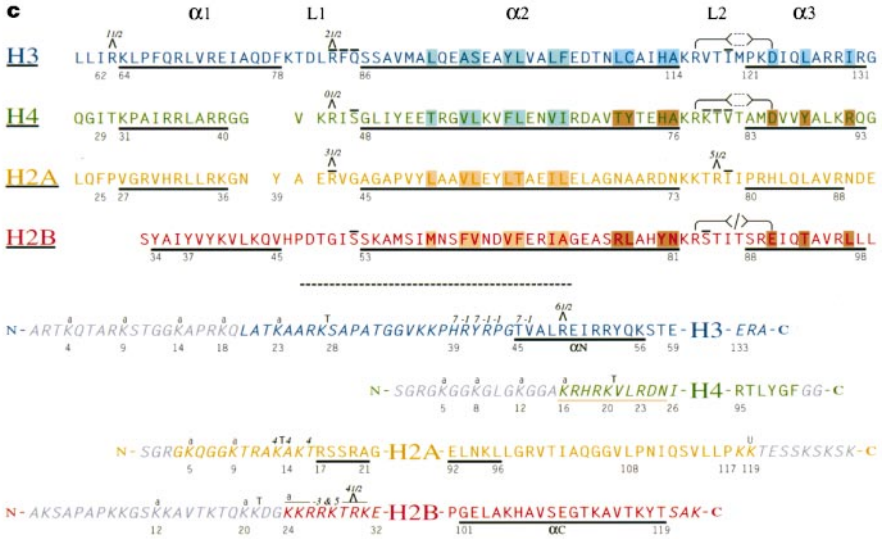
The structure of a nucleosome core particle, crystallized under physiologically relevant ionic conditions[17], was previously solved by X-ray crystallography at 7 Å resolution[18]. This low-resolution structure confirmed the disc-like shape of the particle deduced from earlier studies[19] which results from the flat, left-handed superhelical arrangement of the DNA. The path of the DNA superhelix was irregular, bending tightly at several positions and being nearly straight in others. Although detailed side-chain information or atomic coordinates are not available, the re-interpreted, high-salt, histone octamer crystal structure, for which the histone fold was first described[20], allowed more accurate assignment of the secondary structure of the histone proteins in the core particle at low resolution[21].

We have determined the high-resolution X-ray structure of the nucleosome core particle from new crystals which have all components individually made in bacteria and assembled after purification. These crystals, containing DNA of defined sequence and histones lacking post-translational modifications[22,23], diffract anisotropically to between 1.8–2.2 Å. Recombinant histone proteins enabled heavy-atom derivatives to be prepared through the substitution of cysteine at many sites. Use of an ESRF undulator beam line allowed accurate measurement of the weak diffraction intensities from these crystals and yielded a high-quality electron density map. The structure presented here is currently refined to 2.8 Å resolution, and includes an entire 146-bp DNA molecule and over 80% of the eight histone chains. Some portions of the N- and C-terminal histone-tail regions extending outside the confines of a single particle have not been included in the atomic model.

Organization of the nucleosome core particle

The 146 bp of DNA are wrapped around the histone octamer in 1.65 turns of a flat, left-handed superhelix (Fig. 1a). Although the human α-satellite DNA used was made palindromic to accommodate perfect twofold symmetry, the sequence in fact binds the histone octamer with a central base pair at the particle pseudo-twofold axis so that the DNA is divided into 73- and 72-bp halves, with one base pair falling on the dyad. We define the rotational orientation of the DNA double helix relative to the central base pair (superhelix location zero, or SHL0, where the major groove faces the octamer), and for each successive turn, the location number increases in the 73-bp half up to SHL + 7, and decreases in the 72-bp half down to SHL − 7 (Fig. 1d; only the 73-bp half of the DNA superhelix and associated histone proteins are shown for clarity). The superhelix





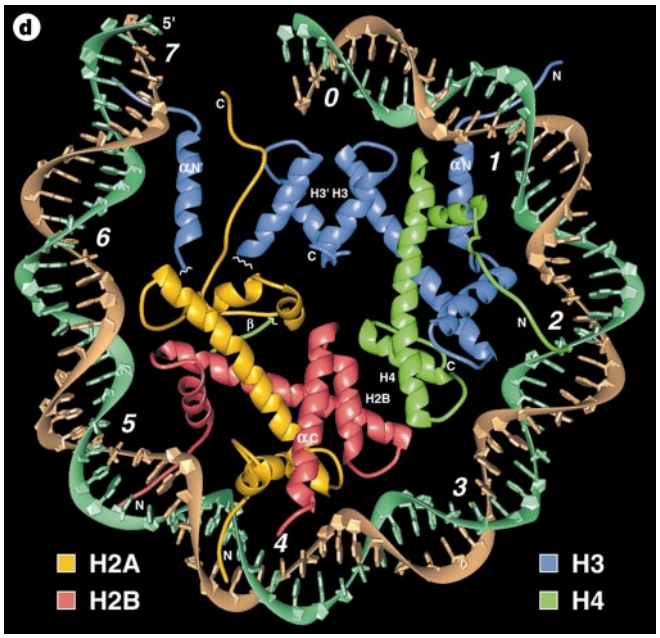


Figure 1 **a**, Nucleosome core particle: ribbon traces for the 146-bp DNA phosphodiester backbones (brown and turquoise) and eight histone protein main chains (blue: H3; green: H4; yellow: H2A; red: H2B. The views are down the DNA superhelix axis for the left particle and perpendicular to it for the right particle. For both particles, the pseudo-twofold axis is aligned vertically with the DNA centre at the top. **b**, DNA phosphate *B*-factors versus base pair. The sequence of the DNA used is shown with corresponding *B*-factors (Å^2) plotted for the 5′ phosphate group of each base. The contacts of the DNA phosphodiester chains with the histones are indicated: squares for main-chain hydrogen bonds; circles for side-chain hydrogen bonds, and triangles for hydrophobic bonds. The bases coloured blue, green, red, and yellow indicate close proximity to an arginine side chain inserted into the minor groove. The DNA phosphate groups have high mobility or are disordered when not contacted by the histones. **c**, Core histone sequences. The histone fold regions for H3, H4, H2A and H2B are aligned on the basis of their structures and labelled on the left (top). The α-helix (α1, α2, α3) and loop (L1, L2) secondary structural elements are labelled; α-helices are underlined. Λ designates an arginine side chain that is inserted into the DNA minor groove at the indicated SHL. The overlying lines indicate β-strand hydrogen bonds between L1 and L2 loops. The side chains indicate the positions of buried arginine–aspartate pairs for H3–H4 and the lack of a homologous pair for H2A–H2B. The histone-fold extensions and tail (italics) regions are shown with the core histone name representing the histone fold (bottom). Grey-shaded sequences were not included in the atomic model. Sites of *in vivo* acetylation are indicated with a letter 'a'. Trypsin cleavage sites that produce a 'tailless' core particle are denoted by T (ref. 50). The site of ubiquitination is marked U (ref. 51). SHL labels above the sequence indicate contacts with DNA at those points. The H4 sequence at amino acids 16–25 and underlined in orange makes a clear interparticle contact with a H2A–H2B dimer. Some interaction sites are highlighted: (1) H3–H4 α2-α2 (blue-green); (2) H2A–H2B α2-α2 (orange); (3) H3 4-helix bundle (blue); (4) H4–H2B 4-helix bundle (brown). **d**, Nucleosome core particle: 73-bp half. The view is down the superhelix axis with the pseudodyad axis aligned vertically. The central base pair through which the dyad passes is above the SHL0 label, *0* (SHL, superhelix axis location). Each SHL label represents one further DNA double helix turn from SHL0 (*1–7*). The complete histone proteins primarily associated with the 73-bp superhelix half are shown (interparticle tail regions are not shown). The two copies of each histone pair are distinguished as unprimed and primed copies, where the histone fold of the unprimed copy is primarily associated with the 73-bp DNA half and the primed copy with the 72-bp half. The 4-helix bundles are labelled as H3′ H3 and H2B H4; histone-fold extensions of H3 and H2B are labelled as αN′, αN and αC, respectively; the interface between the H2A docking domain and the H4 C terminus as β; and N- and C-terminal tail regions as N or C.

has an average diameter of 41.8 Å but is not uniformly bent, having maximum curvature and radius at SHL ± 1.5 and SHL ± 4–5. The difference in the relative mobility of the segments of the DNA phosphate chains facing the solvent compared with those bound to the histones is extreme compared with other protein/DNA complexes, as revealed by the 2–3-fold variation in crystallographic B-factors (Fig. 1b), and may explain the relatively weak diffraction intensity from the crystals.

The protein octamer is divided into four 'histone-fold' dimers defined by H3–H4 and H2A–H2B histone pairs (Fig. 1c). The two H3–H4 pairs interact through a 4-helix bundle formed only from H3 and H3′ histone folds to define the H3–H4 tetramer. Each H2A–H2B pair interacts with the tetramer through a second, homologous 4-helix bundle between H2B and H4 histone folds. The histone-fold regions of the H3–H4 tetramer bind to the centre of the DNA covering SHL − 3 to +3, whereas those of the H2A–H2B dimers bind from −6 to −3 and +3 to +6 (Fig. 1b, d). Further α-helices and coil elements (such as H3-αN and H2B-αC) extend from the histone-fold regions and are also an integral part of the core protein within the confines of the DNA superhelix. Beyond these histone-fold extensions, the histone N- and C-terminal tail sequences, which make up about 28% of the mass of the core histone proteins, are seen over about one-third of their total length in the electron density map. The tails of H3 and H2B pass through channels in the DNA superhelix created by two juxtaposed minor grooves. One H4 tail segment makes a strong interparticle connection, perhaps relevant to the higher-order structure of nucleosomes.

Histone-fold pairs

The central histone-fold domains of all four core histone proteins share a highly similar structural motif constructed from three α-helices connected by two loops, L1 and L2, denoted as α1-L1–α2-L2–α3 (Fig. 1c). These regions form crescent-shaped heterodimers in the pairings H3–H4 (Fig. 2a) and H2A–H2B (Fig. 2b) and bind 2.5 turns of DNA double helix, which arcs around them along their long axes to generate a 140° bend. The histone-fold motif is related to its counterpart in a pair by pseudo-twofold symmetry, with the symmetry axis passing between the two long, 8-turn α2 helices at their crossover. Their antiparallel arrangement places the L1 loop of one in juxtaposition with the L2 loop of the other. As the contact surfaces are offset towards the N terminus by one helical turn, the C terminus of each α2 helix extends further along the long axis than the adjacent N terminus of the paired histone (Figs 1d, 2a, b). This effectively tapers the ends of the pair to create a similar type of DNA-binding site, here called L1L2, at each end of the dimer. Within the histone fold, the N-terminal α1 helix and C-terminal α3 helix fold back on roughly the same side of the central α2 helix, each generating a left-handed spiral in the intervening loops. This arrangement gives a single histone a broad, skewed U-shaped appearance, with two Us slotting together to form a pair. The short, 3-turn α1 and α3 helices make contact with both α2 helices in the dimer, but because of the offset in the α2 to α2 packing, only the α1 helices and their short extensions make contact between opposite histones. The α1 surfaces complete the convex DNA-binding surface, providing its centre and second type, or α1α1, of DNA-binding site. On the opposite, concave side of the dimer, the α3 helices are not in contact.

The formation of only H3–H4 and H2A–H2B heterodimeric pairs instead of homodimers or other heterodimers results from complementary internal packing, as seen from alignment of the structures, although the packing is complicated by further requirements for histone pair–pair and histone–DNA interactions. Obvious differences do occur in the manner in which the α1 helices, their N-terminal extensions, and the L1, L2 loops interact within a pair. Alignment of the four histone types based on the α2 helices reveals that the L1 loop lengths for H3 and H2B are 8 amino acids, whereas for H4 and H2A they are 6 and 7 amino acids, respectively

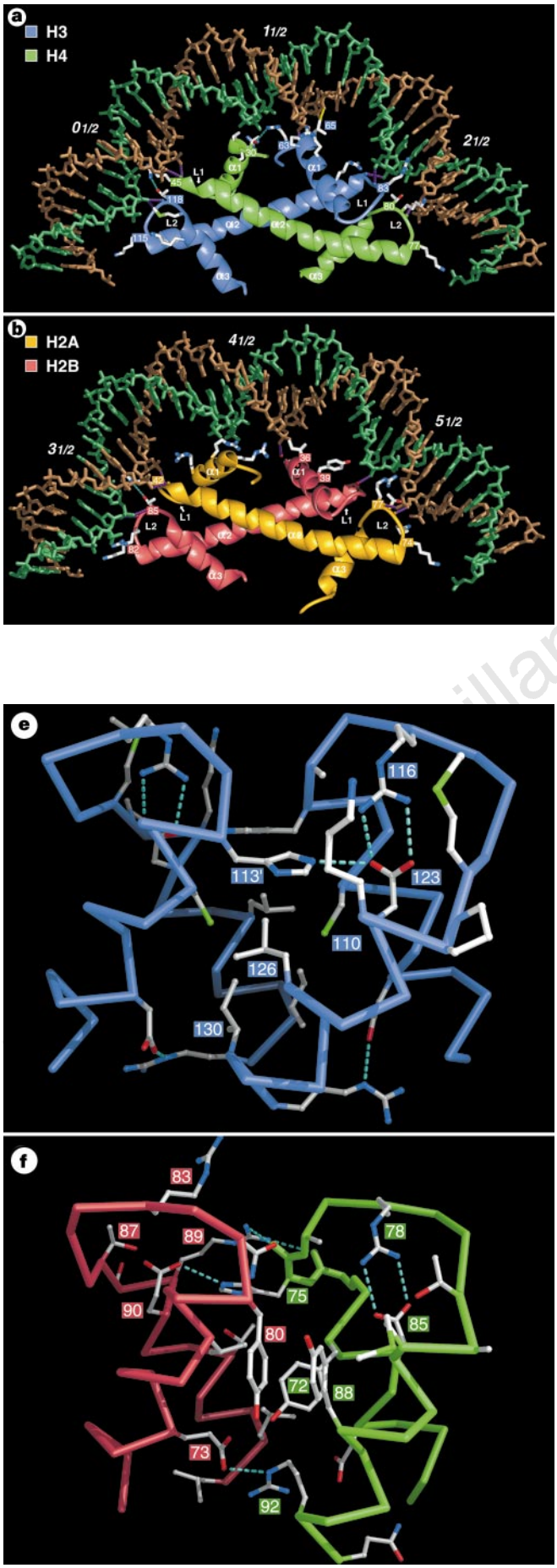


Figure 2 a, H3–H4 histone-fold pair. The α1-L1-α2-L2-α3 structural elements are shown. A pseudodyad axis of symmetry runs vertically through SHL1.5. Side chains that make hydrogen bonds or hydrophobic interactions with the DNA backbones, and arginines inserted in the minor groove are shown. Main-chain-to-DNA-phosphate hydrogen bonds are shown in magenta. Leucine 65 is in contact with a thymidine methyl group (yellow bond). Two of three arginine–threonine pairs are hydrogen bonded (cyan bond). **b**, H2A–H2B histone-fold pair. As in **a**, except that the pseudodyad axis runs through SHL4.5. **c**, H3–H4 L1L2 DNA-binding site. The H3 L1 and H4 L2 loops make 3 hydrogen bonds (cyan) with each other in a parallel β-structure. The L2 loop contains buried hydrogen bonds (cyan) between H4-R78 and H4-D85. The hydrogen-bonding interactions (magenta) between protein and DNA-phosphate groups involve both main-chain amide atoms and side-chain atoms. **d**, H2A–H2B L1L2 DNA-binding site. In contrast to the H3 L1 and H4 L2 loops, the other three L1L2 sites have only one well formed β-structure hydrogen bond, as for H2A L1 and H2B L2 shown here (cyan). An arginine (H2B-R83) hydrogen bonds to a phosphate group instead of an acidic side chain (H2B-E90), as occurs at the homologous sites in the other L1L2 loops. **e**, H3′–H3 4-helix bundle. The H3–H4 histone pairs form the tetramer through the interaction of the C-terminal halves of the α2 helices and the α3 helix of H3′ and H3 across the molecular dyad axis. Histidine 113′ makes a buried hydrogen bond to link the H3 molecules. In addition, several hydrophobic interactions occur. **f**, H2B–H4 4-helix bundle. The H2A–H2B dimer assembles into the histone octamer through the interaction of H2B with H4. The hydrogen bond of H4-H75 to H2B-E90 is analogous to that in H3′–H3, but is not buried here because of the orientation of H2B-R83. Across the pseudo-twofold axis, however, H2B-Y80 replaces the H4 histidine and, together with H4-Y72 and H4-Y88, forms a hydrophobic cluster.

(Fig. 1c). The loops run essentially antiparallel to the connected α1 helices, so that the shorter loops of H4 and H2A are compensated for by an earlier stop of the α1 helix with respect to the underlying contacts. The side chains I62 of H3 and I29 of H4 on the α1 helices of a H3–H4 pair pack together with pseudodyad symmetry (Fig. 2c), but the H2A–H2B dimer has an asymmetric interface here because of the insertion of Y39 of H2A into the H2A L1 loop, as compared to L1 for H4 (Figs 1c, 2d). Outside the histone folds, the αN extension of H3 and the αC extension of H2B contribute substantially to their respective pairings.

The extra length of the L1 loops and extended N terminus of the α1 helices for H3 and H2B take a path somewhat further from the overall long axis of a histone pair compared to the L1 loops of H4 and H2A. The adjacent H4 L2 and H2A L2 loops are thereby displaced towards this central axis so that the C termini of the H4 and H2A α2 helices are bent towards the L1 loops of the partner: the H4 and H2A α2 helices curve, whereas those of H3 and H2B are straight.

The L2 loops in the C-terminal half of the histone fold show a high degree of structural homology between histones. A three-amino-acid stretch in the L2 loop of one histone in a pair and the L1 loop of the other run roughly parallel to each other at each end of a dimer. In the case of H3 L1 with H4 L2, the maximum of three hydrogen bonds are made in parallel β-conformation (Fig. 2c), whereas for the other three L1–L2 interactions, only one hydrogen bond is evident. The conformation of these loops may depend on the DNA sequence bound, which can be tested by obtaining structural data from particles with different DNA fragments. A second important feature of the H3 L2 and H4 L2 loops is that although arginines H3-R116 and H4-R78 are in close proximity to the DNA backbone, they face away from it and make hydrogen bonds, respectively, with aspartate side chains H3-D123 and H4-D85 (first turn of the α3 helices). These acidic side chains are inaccessible to solvent contact. In contrast, the equivalent H2B-R83 does bind a DNA phosphate, despite the presence of H2B-E90 at the same position as the aspartate groups in H3 and H4 (Fig. 2d), whereas H2A lacks both the basic and acidic side chains at these sites.

The stable oligomeric aggregates of histones are the H3–H4 tetramer and the H2A–H2B dimer in solutions of physiologically relevant ionic strength, or the histone octamer in the presence of DNA or high salt concentration. The C-terminal halves of the α2 helices and the α3 helices are the primary determinants of the tetramer and octamer assemblies through the formation of three four-helix bundles—a single H3′–H3 association to form the tetramer (Fig. 2e) and two H2B–H4 associations to produce the octamer (Fig. 2f). Including the histone folds and extensions, further interactions occur within an octamer, between H3–H2A′, H4–H2A′, H4–H2B′, H2A–H2A′ and H2A–H2B′. The use of the H3, H4 and H2B α3 helices as the principal interaction sites for the assembly of tetramer and octamer suggests that the remaining H2A α3 helix might form an internucleosomal 4-helix bundle to create the higher-order structure of nucleosomes. However, the resulting complex seems unlikely to be formed, based on current models for the higher-order structure[3].

The H3′–H3 4-helix bundle contains buried charged groups which are hydrogen-bonded in the intermolecular interaction (Fig. 2e). These interactions involve H3-D123 again and, in addition, H3′-H113 on the adjacent copy of H3. The side chain of cysteine H3-C110, which has been the primary target for attachment of extrinsic probes of nucleosome structure, is in the centre of the bundle. Access to this site would require the H3-L126 side chain to adopt an alternative conformation, which is sterically possible. Core particles prepared in oxidizing conditions have the H3-C110 side chains linked in a disulphide, and these particles have somewhat different properties from those kept in fully reducing conditions[24]. The 6.2 Å separation of the sulphydryl groups is too large for disulphide bond formation without substantial distortion to the tetramer conformation. The H3-C110 amino acids were the target of the tetrakis(acetoxymercurimethane) (TAMM) multi-heavy-atom compound that we used to produce single-site derivative crystals for the structure determination (the H3-C110 sulphydryls are bound and the H3-H113 imidazoles are not). For other derivatives, the H3-C110A mutant was used to avoid disturbance of the structure and to avoid duplicate sites between derivatives.

Table 1 Crystallographic analysis

Diffraction data

	Native	1 × TAMM	2 × TAMM	4 × CH_3HgNO_3
Resolution	25–2.8 Å	25–2.8 Å	25–2.8 Å	25–2.8 Å
Observations	297, 397	235, 751	229, 513	186, 451
Unique *hkl*	52, 432	52, 345	52, 372	50, 986
Completion (%)	0.99	0.98	0.98	0.96
R_{merge}*	0.056	0.064	0.068	0.056
R_{merge}* 3.0–2.8 Å	0.132	0.135	0.127	0.129

MIR analysis

Heavy-atom sites	H3 C110	H4 T73	H3 E133, H4 S47
Phasing power† (MLPHARE)	0.84	0.92	0.97
Cullis *R*-factor‡	0.89	0.89	0.85

Refinement statistics

Observations ($F > 3 \times \sigma$), working/test	43,638/2,639
R-factor§/free *R*-factor (%)	22.4/30.2

Component	Number of atoms	R.m.s.d. bond length	R.m.s.d. bond angle	Mean *B*-factor
Protein	6,387	0.010 Å	1.92°	40.4
DNA	5,980	0.006 Å	1.13°	80.2
Total	12,367	0.009 Å	1.31°	59.7

* $\sum_{hkl}\sum_i \left|I_{hkl,i} - \bar{I}_{hkl}\right| \Big/ \sum_{hkl}\sum_i I_{hkl,i}$, where $\bar{I}_{hkl}$ is the mean of measurements for a single *hkl*.

† $\left\{\sum_{hkl} F_H^2 \Big/ \sum_{hkl}\left\{F_{PH,obs} - F_{PH,calc}\right\}^2\right\}^{1/2}$.

‡ $\sum_{hkl}\left|\left|F_{PH} \pm F_P\right| - F_{H,calc}\right| \Big/ \sum_{hkl}\left|F_{PH} \pm F_P\right|$, centric only.

§ $\sum_{hkl}\left|F_{obs} - F_{calc}\right| \Big/ \sum_{hkl} F_{obs}$.

The H2B–H4 4-helix bundle is closely related in its overall conformation to that for H3′–H3 (r.m.s.d. on α-carbon of 1.85 Å). Histidine H4-H75 makes a buried hydrogen bond with glutamate H2B-E90, but the twofold related interaction is completely different (Fig. 2e, f). A tyrosine side chain, H2A-Y72, replaces H3-C110 and lies flat against the tyrosine side chain H2B-Y80, where one of the buried histidine residues is found in H3. Both H3′–H3 and H2B–H4 bundles contain several other hydrophobic and hydrogen bonds. The increase in hydrophobic contribution to this H2B–H4 interface compared to that in H3′–H3 explains, at least in part, the instability of the histone octamer in contrast to the tetramer at low salt concentration.

The remaining interactions between the histone-fold domains occur between H2A, H2B and H2A′. Glutamate H2A-E41 and histidine H2B-H79 are organized each to make a hydrogen bond with the side chain of glutamine H2A′-N38. A second pair of identical bonds is made nearby across the particle pseudo-twofold axis. These interactions suggest that there is cooperativity in dimer binding.

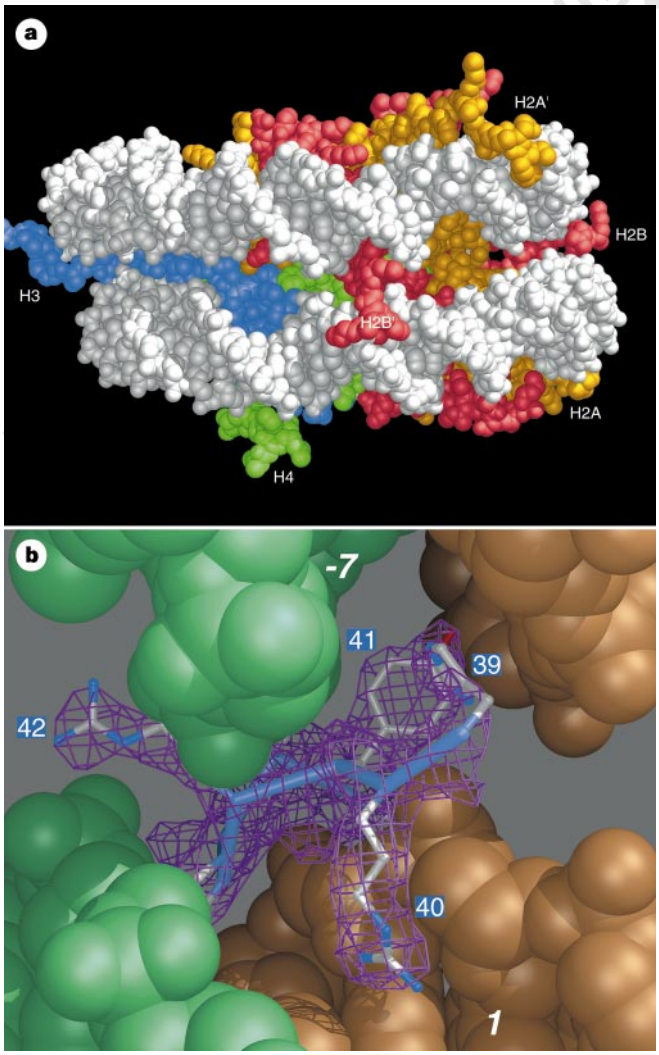
Histone-fold/DNA interactions

The histone-fold domains account directly for the organization of 121 bp of DNA. Each histone-fold pair, whether half of the H3–H4 tetramer or a H2A–H2B dimer, is associated with 27–28 bp of DNA, leaving 4-bp linkers between these units (Fig. 1d, 2a, b). Binding is primarily to the DNA phosphodiester backbones as they face the protein over 2.5 turns of the double helix. Each helical turn of a chain commits a segment of only two consecutive phosphate groups to hydrogen-bonding interactions with the histone-fold pairs. The length of the DNA bound is extended by an additional backbone segment in both directions, eight in total (34–36 bp), if

the H3-K115, H4-K77, H2A-K74 and H2B-K82 conserved lysines that begin the L2 loops and make salt links to phosphate groups are included. The extended DNA sites overlap from one histone-fold pair to the next.

The histone-fold DNA-binding sites can be divided into two types. The first, or α1α1 site, uses both α1 helices to bind the two DNA backbone segments at the centre of the bound DNA stretch. The second, or L1L2, sites are formed from L1 and L2 loops and termini of the α2 helices at each end of the histone-pair crescent. Each binds to the two DNA backbone segments at the edge of the whole DNA site. In general, five predominant features of the histone-fold/DNA interaction are observed each time the phosphodiester chain faces the histone octamer (Fig. 2a–d). (1) The N termini of the α1 helices of H3, H4 and H2B, and the α2-helices of all four core histones, are each used to fix the position of an individual phosphate group through the positive charge generated by the helix dipole. (2) Hydrogen bonds to phosphates are made from main-chain amide nitrogen atoms of amino acids either near or in the last turn of the α1 and α2 helices. (3) An arginine side chain from a histone fold enters the minor groove 10 of the 14 times it faces the histone octamer. The other four occurrences have arginine side chains from tail regions penetrating the minor groove. (4) Extensive nonpolar contacts are made with the deoxyribose groups. (5) Hydrogen bonds and salt links occur frequently between DNA-phosphate oxygen atoms and protein basic and hydroxyl side chain groups. These side chains approach the inward-facing DNA backbone segments from both the major and minor groove sides.

In the α1α1 DNA-binding site, H3, H4 and H2B all have the N terminus of their α1 helix pointing at an individual phosphate group. For H3 and H4 at SHL1.5, proline side chains in the first turn of α1 contact deoxyribose groups (H3-P66, H4-P32) and H3 main-chain amide nitrogen atoms (H3-K64, H3-L65) hydrogen-bond to phosphate oxygen atoms. In the case of H2B at SHL4.5, the α1 helix begins one turn later in the structure superposition of the histone folds (Fig. 1c), and isoleucine H2B-I36 now contacts a deoxyribose while the main-chain amide nitrogen of H2B-S33 hydrogen-bonds to phosphate. As a consequence of the orientation of the H3 and H4 α1 helices relative to those of H2A and H2B, the DNA is 3.5 Å displaced from the underlying α2 helices of a H3–H4 pair compared to a H2A–H2B pair. On the side of the α1 helix that faces the DNA major groove, the side chains (of arginine, lysine, histidine and tyrosine) generally hydrogen-bond to phosphate groups (H3-R69, H4-R35, H2A-R32) or create a complementary, positive electrostatic charge density (Fig. 2a, b). In one case, H3 leucine 65 reaches into the major groove and makes a hydrophobic interaction with the 5-methyl group of thymidine (Fig. 2c). The minor groove at the centre of the binding site for the H3–H4 pair has H3-R63 inserted between phosphate chains (Fig. 2a), whereas the structurally analogous site on H4 contains a threonine group in contact with deoxyribose. These threonine and arginine side chains hydrogen-bond, restraining the arginine guanidinium group from penetrating more deeply into the DNA groove. The H2A α1 helix is oriented differently with respect to the DNA because of the insertion of Tyr 39 in the L1 loop. In this case, the phosphodiester chain is



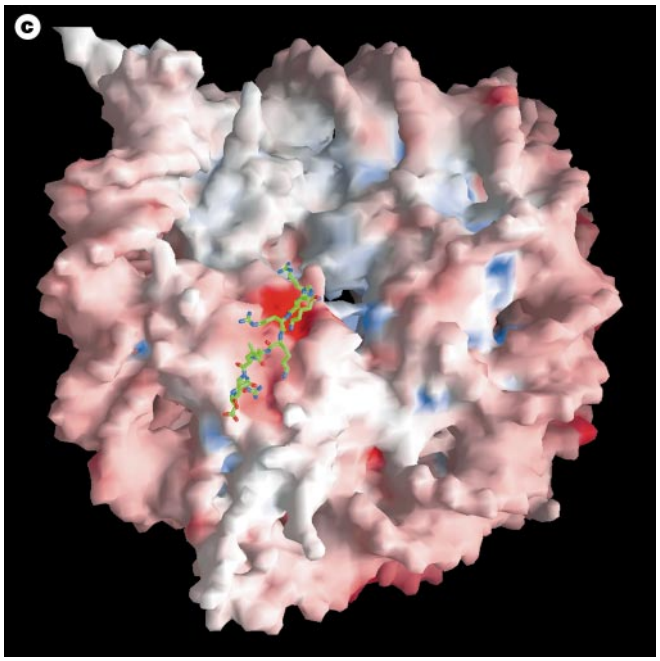


Figure 3 **a**, Histone tails between DNA gyres. The H2B (red) and H3 (blue) N-terminal tails pass through channels in the DNA superhelix (white) formed by aligned minor grooves. **b**, H3 tail segment in minor-groove channel. The H3 tail sequence HRYRP passes through the juxtaposed minor grooves of SHL –7 (DNA terminus) and SHL1 (DNA centre). The electron density (magenta) clearly identifies the conformation of these amino acids. **c**, H4 tail bound to the H2A–H2B dimer. The electrostatic surface of the nucleosome core particle shows an intensely negatively charged region (bright red) near the centre of its exposed protein face. Amino acids 16–25 of the H4 tail lie on this region of the H2A–H2B dimer, which contains 7 clustered acidic side chains. The orientation of the particle is as for Fig. 1d.

held by H2A arginines (R29, R32) extending from the α1 helix. The N termini of the α1 helices of the H2A–H2B dimers are the takeoff points for N-terminal tails that interact with the DNA. The tail of H2A clasps the DNA at the minor groove on the outside of the particle, and the tail of H2B passes through the minor groove channel between superhelix gyres. At SHL4.5, the guanidinium group of H2B-R30 in the tail penetrates most deeply into the minor groove.

In an L1L2 DNA-binding site, the N terminus of an α2 helix points towards a phosphate group and, for H3, H4 and H2A, the arginine side chain in the preceding L1 loop is inserted into the minor groove (Fig. 2c, d). The main-chain amide nitrogen atoms in the L1-α2 junction hydrogen-bond to the DNA phosphates, probably specifying their positions precisely. The H4 and H2A-L1 loops are nearest to the pseudodyad axis of the core particle at SHL0.5 and 3.5, respectively, whereas the longer H3 and H2B-L1 loops are more distant at SHL2.5 and 5.5, respectively. The more highly conserved L2 loops at the α2 C termini bind the outermost DNA strands of the bound 26-28-bp DNA segment and hold the phosphates in place with hydrogen bonds from main-chain amide nitrogen atoms as well as side chains. The single arginine in the L1 loop of each H3, H4 and H2A (H3-R83, H4-R45, H2A-R42), which extends into the DNA minor groove in each case, is hydrogen-bonded to a threonine hydroxyl group in the adjacent L2 loops of H4, H3 and H2B (H4-T80, H3-T118, H2B-T85), respectively. This interaction appears to restrain the arginine side-chain conformation so that it is not free to hydrogen-bond with the nearest O2 atom of cystosine or thymidine, or the N3 of adenine or guanine. In one case, however, H4-R45 hydrogen-bonds to the O2 of a thymidine base at SHL0.5 while maintaining the threonine interaction. In the H2B sequence, the L1-loop arginine is replaced by glycine (H2B-G50) and the position of the restraining threonine is replaced by H2A-R77, which penetrates the DNA minor groove at SHL5.5. As the DNA straightens at this location, probably because there are fewer histone contacts compared with other binding sites, H2A-R77 does not extend further into the minor groove than the arginine–threonine pairs.

Tails and extensions

The N-terminal tails of both H3 and H2B have random-coil segments that pass between the gyres of the DNA superhelix (Figs 1c, 3a). For H3, five amino acids (39–43) pass through a channel in the superhelix formed by the minor grooves of the DNA terminus (SHL6.5 to 7) and central turn (SHL −0.5 to −1) (Fig. 3b). An extremely basic, 8-amino-acid span of H2B also passes through the superhelix (SHL4.5 and SHL −2.5 to −3), which creates, with H3 and the twofold molecular symmetry, a 20-bp periodicity of tails in minor-groove channels. At a location (SHL4 to 4.5) adjacent to the H2B interaction, four amino acids of the H2A N-terminal tail are bound to the minor groove on the outside of the superhelix (Fig. 3a). The electron density for tail sequences further N-terminal is weak in most cases and therefore not interpreted. These regions, which are highly basic and contain the sites of acetylation for H3 and H2B and border the ubiquitination sites for H2A, are probably involved in formation of the chromatin fibre through interactions with neighbouring nucleosomes.

The H3 αN helix extension of the histone-fold and preceding tail region in the minor grooves are responsible for binding the 13 bp of DNA at each terminus of the superhelix (Fig. 1b, d). The αN helix lies on the H4 α2 N terminus and H4 L1 loop on the opposite face of the particle from the α1 helix to which it is connected. A section of the H2A′ C-terminal tail (amino acids 105–117) runs antiparallel to αN and further links it to the underlying H3–H4 histone-fold domains.

Preceding and overlapping the region of the H2A C terminus that contacts H3′ αN, the H2A tail (amino acids 92–108) forms a folded 'docking' domain with the H2A α3 helix. This buried domain contains a short α-helix (amino acids 92–96), and makes β-sheet interactions with the short H4′ C-terminal tail (amino acids 95–102) folded back over its α3 helix (Fig. 1d). The H2A tail and docking domain (amino acids 80–119) account for a loss of 1,914 Å^2 in accessible surface area on binding of a H2A–H2B dimer to the H3–H4 tetramer, as compared to 1,074 Å^2 for the H2B–H4 4-helix bundle interaction. These extensions of the histone-fold motif appear to be necessary to stabilize the helical ramp formed by the histone-fold domains, maintaining its pitch accurately.

A striking feature of the protein surface is the V-shaped pair of antiparallel H2B αC and α3 helices (Fig. 1d; side chains not shown). The αC helix defines the outer limit of the flat face of the core particle and is held in place by a 4-amino-acid linker to the preceding H2B α3 helix and by extensive hydrophobic interaction with the H2A α2 helix, including H2A-Y50, which is inserted between the two H2B α-helices. The tyrosine H2B-Y118 at the C terminus of the αC helix interacts with a single-turn α-helix within the H2A N terminus (amino acids 17–20) before this tail passes over the DNA minor groove at SHL4 to 4.5 (Fig. 1d). The N-terminal tail of H2B passes through the minor groove channel just preceding SLH −3 and +5, with only H2B-E32 linking this DNA-binding region to the histone-fold domain. The requirement of immediate entry of the H2B tail into the adjacent minor groove is accommodated by the limited interaction between the H2B and H2A α1 helices.

The two H4 N-terminal tails up to amino acid 25 have divergent structures. Only one is well localized and then only in the stretch from K20 to I26, although diffuse electron density exists additionally for K16 to R19. The observed part makes extensive contact with an H2A–H2B dimer of an adjacent particle. This H4 region, K16 to N25, makes multiple hydrogen bonds and salt bridges between its basic side chains (K16, R19, K20, R23) and acidic side chains of H2A (E56, E61, E64, D90, E91, E92) and H2B (E110). Because of the positive charge of the H4 tail and negative charge on the H2A–H2B dimer surface (Fig. 3c), it is uncertain whether the observed interactions are unique. The importance to successful crystallization of one of these interactions was confirmed as the mutant with K20C cannot crystallize under our conditions, but can be overcome through reintroduction of the ϵ-amino group by making the H4-C20–cystamine adduct (K.L. and T.J.R., unpublished results).

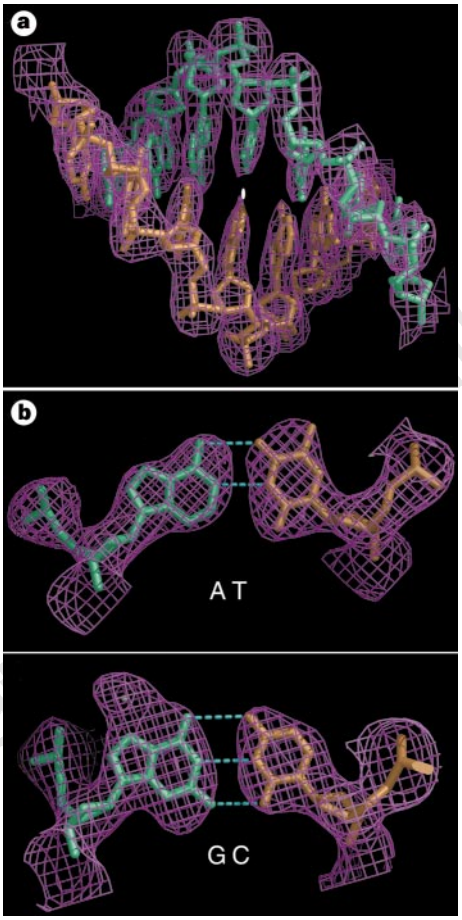
DNA superhelix

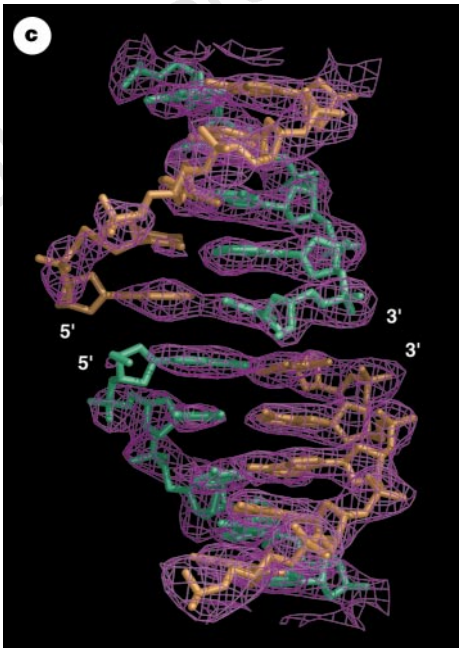
The diffraction quality of nucleosome core particle crystals depends largely on the DNA sequence that is incorporated. Through trials of many sequences, we have found that two different halves of human α-satellite repeats yield diffraction to 2.0 Å resolution when used as an inverted repeat. The potential for structure determination of a mutant of one of the α-satellite repeats has been noted[25]. We were able rapidly to check the position(s) of octamer on a variety of different DNA sequences by using our site-specific, hydroxyl-radical cleavage method[7]. The X-ray structure verifies our discovery made by using this technique, namely that the molecular two-fold axis of the particle passes through a single base pair and not between two, despite the use of a 146-bp palindromic sequence (Figs 1b, 4a). We can identify directly from the electron density the type of base pair for ~50% of the sequence (Fig. 4b, c), and thereby distinguish the 72-bp from the 73-bp half extending from the central base pair. Termini of the 72- and 73-bp halves from adjacent particles pack end-to-end in a pseudo-continuous double helix, but with a twist of 132° at the junction (Fig. 4d). Superposition of the two DNA halves shows that they have closely similar structures apart from a 12-bp section of the 72-bp half (base pairs 12–23) which covers a 13-bp region in the 73-bp half (Fig. 4d). We attribute the location of this stretched region either to strong, specific DNA-positioning requirements in the flanking sequences, or to the strength of the inter-particle contact between DNA termini. The latter explanation is favoured because a second 146-bp DNA sequence that yields

high-quality crystals displays a similar distortion, but it occurs only two turns away from a DNA terminus (data not shown).

The best ideal superhelix fit to the double helix axis over base pairs from 62 in the 72-bp half to 63 in the 73-bp half has a 41.8 Å radius and a 23.9 Å pitch. As the 10-bp segment at each terminus is essentially straight, the effective number of superhelical turns is 1.65. The DNA locations most distant from the superhelix axis, SHL ±1.5 (44 Å) and ±4 to 5 (45 Å), also show the greatest bending. The underlying histone structures are the α1α1 DNA-binding sites of the H3–H4 pairs for SHL ±1.5 (Fig. 2a), which cause an outward bulge in the DNA, and the adjacent L1L2 and α1α1 DNA-binding sites of the H2A–H2B dimers, which buckle the DNA outwards between them. In general, the superhelix path is significantly distorted owing to the local structure of the histone DNA-binding surface.

The parameters of the double helix also show substantial variation over the length of the superhelix. The overall twist, for example, is 10.2 bp per turn in the local reference frame (independent of superhelical pitch). Dividing the DNA into segments delimited by the arginine side chains inserted into the minor groove, the minimum value is obtained for the overwound stretch in the 72-bp half that compensates for one less base pair: 9.4 bp per turn versus 10.6 for the same segment in the 73-bp half. The maximum value of 10.9 for unwinding occurs at the SHL-5 segment, where an H2B N-terminal tail passes through the minor-groove channel. When applied to the left-handed superhelix, the overall overwinding of nucleosomal DNA by 0.3 bp per turn compared to free DNA creates an alignment of minor grooves between superhelix gyres that allows an essentially straight passage for the H3 and H2B tails through the channels formed. Further details of the nucleosome DNA parameters





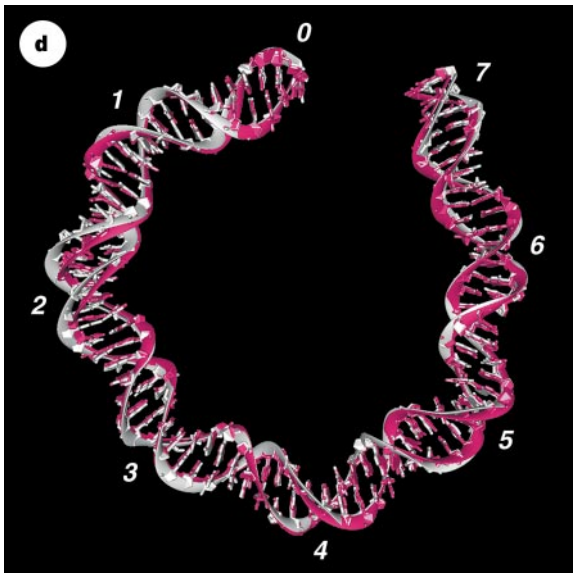


Figure 4 a, Central DNA base pairs. The electron density of the 7 central base pairs shows that the molecular twofold (white mark) axis passes through a base pair. For the 146-bp DNA used, the two halves of the superhelix have lengths of 72 bp and 73 bp. **b**, A·T and G·C base pairs. The A·T base pair is adjacent to and the G·C base pair is 3 bp removed from the central base pair in the 73-bp half. Base-pair sequencing from the electron density map identifies the 72-bp and 73-bp halves of the superhelix (the density above the guanosine N7 atom is a Mn^{2+} ion). **c**, Stacked DNA termini. The 72-bp (top) and 73-bp (bottom) DNA ends from adjacent particles abut each other and are partially base stacked. The discontinuity in the junction because of approximate 5′–5′ and 3′–3′ pairings permits identification of the last base pair for both termini. **d**, Superposition of 72- and 73-bp DNA superhelix halves. The 72-bp (red) and 73-bp (white) superhelix halves, excluding the regions from 12–23, superimpose with a r.m.s.d. of 1.13 Å (all atoms). The 12-bp stretch excluded is overwound in the 72-bp half and covers the same region of the H3–H4 tetramer as does the 73-bp half in 13 bp.

will be reported in a comparison of two nucleosome structures containing different DNA sequences (A.M., K.L. and T.J.R., manuscript in preparation).

Discussion

The significantly non-uniform shape of the DNA superhelix and semiregular variation in DNA parameters suggests that nucleosome positioning results in part by DNA-sequence-dependent flexibility and histone-induced bendability. The double-helix twist value of 10.2 bp per turn observed for the structure overall matches the sequence periodicity of DNA taken from nucleosome cores[26] but deviates from the 10.5-bp-per-turn twist value measured for free DNA[27], indicating that eukaryotic DNA has adapted to fit the nucleosome. Conversely, DNA sequence should affect nucleosome formation and stability. The observation that G·C base pairs are preferred when the major groove faces inwards is simply explained by their tendency to bend into the major groove[28]. Furthermore, the single arginine side chains that are inserted into the minor groove at every turn of the double helix impart a sequence preference to those base pairs in contact with the histones. Sequences that can most easily adopt a narrow minor groove (correlated with bending into the minor groove and a local helix overwinding, as observed in the structure) will be preferred because stronger salt links can be made between phosphates across the groove, resulting in the preference for A·T base pairs in these locations[29]. In addition to sequence-dependent properties that can affect nucleosome positioning, specific and semispecific interactions can occur, such as that between a leucine residue and a thymidine 5-methyl group or an arginine hydrogen-bonding to a pyrimidine O2 atom. The generation[26] and characterization[7] of a nucleosomal DNA sequence database, when analysed in terms of the actual nucleosome core particle DNA structural template, will enable rules to be drawn up for predicting nucleosome position.

The packaging of DNA in nucleosomes generally inhibits the binding of non-histone, DNA-binding proteins such as transcription factors. The restrictions imposed by bound histones may be as much due to distortion of the DNA as to the direct blocking of binding sites. The static accessibility of the DNA molecule bound in the core particle to a probe of 4 Å diameter is 73% of the total value for the bare superhelix. In principle, a well positioned nucleosome could provide factor-binding sites that are distinct from any available in free DNA. The HIV integrase, for example, preferentially inserts the HIV genome into nucleosomes with two apparent hot spots, where the major groove faces outwards at SHL 1.5 and SHL 3.5 to 4 (refs 30, 31). These two sites have the highest curvature of locations in which the major groove faces outwards. The histone tail segments passing through the minor-groove channels every 20 bp may reinforce the periodicity of the integration pattern.

Most DNA-binding factors need DNA to dissociate from the nucleosome in the region of their binding site to allow access. Because of the multiple DNA-binding sites of the octamer (9–10 pairs of chemically distinct structures: four L1L2, two α1α1, 3–4 tails), dissociation would occur in semi-cooperative stages (non-contacted base pairs in one stretch are maximally four), beginning at the entry and exit points of the DNA held by the H3 N-terminal tail and αN extension, and then proceed into the H2A–H2B dimer, and eventually into the H3–H4 tetramer regions. This scheme was confirmed by measurement of the accessibility of restriction enzyme sites in nucleosome cores, which showed an increasing resistance to digestion of the order of 10^2 for sites at DNA termini, to 10^5 for sites at the DNA centre, compared to free DNA[32]. When a DNA-binding factor approaches nucleosomal DNA, the reduction of cooperativity in binding to the histones appears as cooperativity in the binding of successive protein factors, even without their direct interaction[13]. The affinity of one to a few histone sites for DNA is not insignificant, however, even in the region of SHL6–7, because transcription by a bacteriophage RNA polymerase is impeded first when it encounters a nucleosome and then again when it reaches SHL5, the location just before the site of interaction with the H2B N-terminal tail. Apparently the polymerase pauses when the DNA behind it binds the histone octamer at SHL6–7 and creates a constraining loop[33]. For many genes, however, ATP-dependent nucleosome-modification complexes, such as the yeast SWI/SNF complex, are required to permit transcription in the presence of nucleosomes[34]. Remodelling factors may simply aid the dissociation of DNA from the histone octamer, for example, by destabilizing the H2A–H2B dimer/H3–H4 tetramer interactions. Mutations in yeast histones H3 and H4 that partially overcome the loss of the SWI/SNF complex are found predominantly in the H4 L1 loop and the adjoining H3 L2 loop at SHL0.5 (ref. 35). The specific residues affected are the buried H3-R116, H4-R45 inserted in the minor groove, and H3-T118, which is hydrogen-bonded to both H4-R45 and a DNA phosphate group. Additional residues with the same phenotype are H4-V43, which contributes to the shape of the loops by contacting an analogous H3 isoleucine, and H3 E105, for which a direct effect is difficult to envisage at the mononucleosome level.

A model for the higher-order structure of chromatin can be built using our structure of the nucleosome core particle that is highly consistent with the 'standard' model of Finch and Klug[36–39] (T.J.R., unpublished observation). A remarkable feature suggested by this model is that the highly basic, H4 N-terminal tail sequence from amino acids 1–25 binds as an extended chain to a region of extreme acidity on the exposed face of the H2A–H2B dimer (Fig. 3c). This region clearly binds the H4 amino acids 16–24 in the interparticle contacts of the crystals. Although in our model of higher-order structure, the H4 tail is incorporated in the opposite directional sense compared to that in the crystal, the acidic surface and basic tail should permit an equally extensive interaction in either orientation. In the crystal, the H2A–H2B dimer binding site for the H4 tail includes seven clustered acidic amino acids. The H4 amino-acid stretch 16–24, which is bound between particles, binds to the yeast silencing protein SIR3 and participates in the formation of yeast 'heterochromatin'[40]. Our model suggests that an internucleosome connection in the higher-order structure is replaced by an interaction with SIR3. In the model, the four lysine acetylation sites at the H4 N terminus are near the H2B αC-helix. The affinity of the crossbridge formed by the H4 tail to the open face of the adjacent nucleosome in the model of the higher-order helix would be reduced by successive acetylation of the H4 lysines 5, 8, 12 and 16. The role of acetylation in releasing DNA bound in chromatin is more likely to be a destabilization of the nucleosome higher-order helix than of the nucleosome itself.

The histone-fold motif is highly conserved, as seen from structures obtained from organisms as diverse as archaeal bacteria[41], insects[42], birds[20] and amphibians (this study), presumably because of its unique DNA-binding properties. Nevertheless, the core histones occur in four types which may have diverged to impart particular characteristics to different segments of the DNA superhelix, and to accommodate requirements of the higher-order structure and interaction with assembly, transcription and remodelling factors. Histone sequences are more variable outside the histone fold; these extensions and tail regions may be absent, as in the case of the archaeal proteins[41], or replaced by entirely different domains, as in the *Drosophila* factors $TAF_{II}42$ and $TAF_{II}62$ and their homologues[42]. For these TAFs and TAF30α, which is the most similar to H2B, the arginine side chains of the core histones that insert into the minor groove are replaced by other side chains, suggesting that if these TAFs do bind DNA, then the mode of interaction is likely to be different from that in the nucleosome. □

Methods

Crystal preparation. Nucleosome core particles were prepared from a 146-bp palindromic DNA fragment derived from human α-satellite DNA[43] and recombinant histone proteins (Fig. 1b, c)[23]. Crystals were grown by vapour

diffusion in 8–20 days at 20 °C using a droplet containing 4 mg ml^{-1} core particle, 50 mM KCl, 70–75 mM $MnCl_2$, and 20 mM potassium cacodylate, pH 6.0, surrounded by silicon oil DC200 (110 mPa s; Fluka) and equilibrated against 40–46 mM $MnCl_2$, 35–40 mM KCl and 20 mM potassium cacodylate, pH 6.0. Crystals were transferred to 37 mM $MnCl_2$, 40 mM KCl, 20 mM cacodylate, pH 6.0, and 24% 2-methyl-2,4-pentanediol for stabilization and improvement of diffraction resolution. Three heavy-atom derivatives, selected from 11 crystal types each with a different mutant particle, were made by soaking crystals with TAMM[44] (1 × TAMM), or by solution-labelling core particles containing the mutants H4-T73C with TAMM (2 × TAMM), or H3-E133C and H4-S47C (4 × CH_3HgNO_3) in 20 mM Tris-acetate, pH 7.5, with methyl-mercury nitrate before crystallization.

Crystallographic analysis. Data are summarized in Table 1. Crystals belong to space group $P2_12_12_1$, with one nucleosome core particle per asymmetric unit and unit cell parameters: $a = 106.04$, $b = 181.78$, $c = 110.12$. For data collection, crystals were mounted in tapered glass capillaries, transferred to 4 °C, cooled over 5 min to −18 °C, flash-cooled in liquid propane at −120 °C, and transferred into a N_2 gas stream at -170 ± 5 °C. X-ray intensities were collected using a 30-cm MAR image plate scanner at a wavelength of 0.97 Å (ID13, ESRF, Grenoble). Data (25–2.8 Å) were processed using local programs (T.J.R., unpublished) and heavy-atom sites were identified from difference Patterson maps. MIR phases and model combination were calculated using local programs (T.J.R., unpublished) and yielded an overall figure of merit of 73.5 (46.7 with MLPHARE).

Model building and refinement. The initial MIR map was used to trace 120 bp of DNA and 745 amino acids. Iterative rounds of model building with O[45], simulated annealing refinement with XPLOR[46], and phase combination with sigma-a[49] weighting (T.J.R., unpublished) resulted in a model containing the entire DNA and histone octamer minus some of the tail regions (H3: amino acids 38–135; H3′: 20–135; H4; 20–102; H4′: 16–102; H2A: 4–118; H2A′: 12–118; H2B and H2B′: 24–122). After group temperature-factor refinement and application of a solvent mask, an *R*-factor of 22.4 (R_{free} of 30.2) was obtained. All amino acids included in the model were within the allowed regions of the Ramachandran plot. Tail regions generally had high temperature factors (>60), unless engaged in crystal contacts. Graphic figures were prepared with MidasPlus[47] and GRASP[48]. The electron density displayed is at 1.3× sigma of the map.

Received 8 May; accepted 9 July 1997.

1. Kornberg, R. D. Structure of chromatin. *Annu. Rev. Biochem.* **46,** 931–954 (1977).
2. McGhee, J. D. & Felsenfeld, G. Nucleosome structure. *Annu. Rev. Biochem.* **49,** 1115–1156 (1980).
3. Widom, J. Toward a unified model of chromatin folding. *Annu. Rev. Biophys. Biophys. Chem.* **18,** 365–395 (1989).
4. van Holde, K. E. *Chromatin* (Springer, New York, 1988).
5. Blank, T. A. & Becker, P. B. The effect of nucleosome phasing sequences and DNA topology on nucleosome spacing. *J. Mol. Biol.* **260,** 1–8 (1996).
6. Wallrath, L. L., Lu, Q., Granok, H. & Elgin, S. C. R. Architectural variations of inducible eukaryotic promoters: Preset and remodeling chromatin structures. *BioEssays* **16,** 165–170 (1994).
7. Flaus, A., Luger, K., Tan, S. & Richmond, T. J. Mapping nucleosome position at single base-pair resolution by using site-directed hydroxyl radicals. *Proc. Natl Acad. Sci. USA* **93,** 1370–1375 (1996).
8. Travers, A. A. DNA bending and nucleosome positioning. *Trends Biochem. Sci.* **12,** 108–112 (1987).
9. Flaus, A. & Richmond, T. J. Positioning and stability of nucleosomes on MMTV 3′ LTR sequences. *J. Mol. Biol.* (in the press).
10. Wasylyk, B. & Chambon, P. Transcription by eukaryotic RNA polymerases A and B of chromatin assembled *in vitro*. *Eur. J. Biochem.* **98,** 317–327 (1979).
11. Grunstein, M. Histone function in transcription. *Annu. Rev. Cell Biol.* **6,** 643–678 (1990).
12. Paranjape, S. M., Kamakaka, R. T. & Kadonaga, J. T. Role of chromatin structure in the regulation of transcription by RNA polymerase II. *Annu. Rev. Biochem.* **63,** 265–297 (1994).
13. Polach, K. J. & Widom, J. A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. *J. Mol. Biol.* **258,** 800–812 (1996).
14. Felsenfeld, G. Chromatin unfolds. *Cell* **86,** 13–19 (1996).
15. Schild, C., Claret, F. X., Wahli, W. & Wolffe, A. P. A nucleosome-dependent static loop potentiates estrogen-regulated transcription from the *Xenopus* vitellogenin-B1 promoter *in vitro*. *EMBO J.* **12,** 423–433 (1993).
16. Truss, M. *et al.* Hormone induces binding of receptors and transcription factors to a rearranged nucleosome on the MMTV promoter *in vivo*. *EMBO J.* **14,** 1737–1751 (1995).
17. Rhodes, D., Brown, R. S. & Klug, A. *Meth. Enzymol.* 420–428 (Academic, San Diego, 1989).
18. Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D. & Klug, A. Structure of the nucleosome core particle at 7 Å resolution. *Nature* **311,** 532–537 (1984).
19. Finch, J. T. *et al.* X-ray and electron microscope studies on the nucleosome structure. *FEBS Lett.* **51,** 193–197 (1979).
20. Arents, G., Burlingame, R. W., Wang, B.-C., Love, W. E. & Moudrianakis, E. N. The nucleosomal core histone octamer at 3.1 Å resolution: A tripartite protein assembly and a left-handed superhelix. *Proc. Natl Acad. Sci. USA* **88,** 10148–10152 (1991).
21. Richmond, T. J., Rechsteiner, T. & Luger, K. Studies of nucleosome structure. *Cold Spring Harbor Symp. Quant. Biol.* **LVIII,** 265–272 (1993).
22. Richmond, T. J., Searles, M. A. & Simpson, R. T. Crystals of a nucleosome core particle containing defined sequence DNA. *J. Mol. Biol.* **199,** 161–170 (1988).
23. Luger, K., Rechsteiner, T. J., Flaus, A., Waye, M. M. Y. & Richmond, T. J. Characterization of nucleosome core particles containing histone proteins made in bacteria. *J. Mol. Biol.* (in the press).
24. Camerini-Otero, R. D. & Felsenfeld, G. Sulfhydryl modificaiton of nucleosome. *Proc. Natl Acad. Sci. USA* **74,** 5519–5523 (1977).
25. Harp, J. M. *et al.* X-ray diffraction analysis of crystals containing twofold symmetric nucleosome core particles. *Acta Crystallogr. D* **52,** 283–288 (1996).
26. Satchwell, S. C., Drew, H. R. & Travers, A. A. Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **191,** 659–675 (1986).
27. Rhodes, D. & Klug, A. Sequence-dependent helical periodicity of DNA. *Nature* **292,** 378–380 (1981).
28. Dickerson, R. E., Goodsell, D. S. & Neidle, S. "... The tyranny of the lattice...". *Proc. Natl Acad. Sci. USA* **91,** 3579–3583 (1994).
29. Travers, A. A. & Klug, A. in *DNA Topology and its Biological Effects* (eds Cozzarelli, N. R. & Wang, J. C.) 57–106 (Cold Spring Harbor Press, Cold Spring Harbor, New York, 1990).
30. Pryciak, P. M. & Varmus, H. E. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell* **69,** 769–780 (1992).
31. Pruss, D., Bushmann, F. D. & Wolffe, A. P. Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. *Proc. Natl Acad. Sci. USA* **91,** 5913–5917 (1994).
32. Polach, K. J. & Widom, J. Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. *J. Mol. Biol.* **254,** 130–149 (1995).
33. Studitsky, V. M., Clark, D. J. & Felsenfeld, G. Overcoming a nucleosomal barrier to transcription. *Cell* **83,** 19–27 (1995).
34. Hirschhorn, J. N., Brown, S. A., Clark, C. D. & Winston, F. Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. *Genes Dev.* **6,** 2288–2298 (1992).
35. Kruger, W. *et al.* Amino acid substitutions in the structured domains of histones H3 and H4 partially relieve the requirement of the yeast SWI/SNF complex for transcription. *Genes Dev.* **9,** 2770–2779 (1995).
36. Finch, J. T. & Klug, A. Solenoidal model for superstructure in chromatin. *Proc. Natl Acad. Sci. USA* **73,** 1897–1901 (1976).
37. Thoma, F., Koller, T. & Klug, A. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J. Cell Biol.* **83,** 403–427 (1979).
38. Widom, J. & Klug, A. Structure of the 300 Å chromatin filament: X-ray diffraction from oriented samples. *Cell* **43,** 207–213 (1985).
39. Graziano, V., Gerchman, S. E., Schneider, D. K. & Ramakrishnan, V. Histone H1 is located in the interior of the chromatin 30-nm filament. *Nature* **368,** 351–354 (1994).
40. Hecht, A., Laroche, T., Strahl-Bolsinger, S., Gasser, S. M. & Grunstein, M. Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins: A molecular model for the formation of heterochromatin in yeast. *Cell* **80,** 583–592 (1995).
41. Starich, M. R., Sandman, K., Reeve, J. N. & Summers, M. F. NMR structure of HMfB from the hyperthermophile, *Methanothermus fervidus*, confirms that this archaeal protein is a histone. *J. Mol. Biol.* **255,** 187–203 (1996).
42. Xie, X. *et al.* Structural similarity between TAFs and the heterotetrameric core of the histone octamer. *Nature* **380,** 316–322 (1996).
43. Yang, T. P., Hansen, S. K., Oishi, K. K., Ryder, O. A. & Hamkalo, B. A. Characterization of a cloned repetitive DNA sequence concentrated on the human X chromosome. *Proc. Natl Acad. Sci. USA* **79,** 6593–6597 (1982).
44. O'Halloran, T. V., Lippard, S. J., Richmond, T. J. & Klug, A. Multiple heavy-atom reagents for macromolecular X-ray structure determination. Application to the nucleosome core particle. *J. Mol. Biol.* **194,** 705–712 (1987).
45. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
46. Brünger, A. *X-PLOR v3.1 Manual* (Yale Univ. Press, New Haven, 1992).
47. Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. The MIDAS display system. *J. Mol. Graph.* **6,** 13–27 (1988).
48. Nicholls, A., Sharp, K. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11,** 281–296 (1991).
49. Read, R. J. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* **42,** 140–149 (1986).
50. Böhm, L. & Crane-Robinson, C. Proteases as structural probes for chromatin: the domain structure of histones. *Biosci. Rep.* **4,** 365–386 (1984).
51. Goldknopf, I. L. & Busch, H. Isopeptide linkage between nonhistone and histone 2A polypeptides of chromosomal conjugated-protein A24. *Proc. Natl Acad. Sci. USA* **74,** 864–868 (1977).

Acknowledgements. We thank S. Halford for the EcoRV enzyme; A. Flaus, T. Rechsteiner and S. Tan for support and discussion; and C. Riekel and his co-workers at ID13 of the ESRF, Grenoble for support and cooperation. This research was supported in part by the Swiss National Fond.

Correspondence and requests for materials should be addressed to T.J.R. The atomic coordinates have been deposited in the Brookhaven Protein Databank under code 1aoi.