



# THE ATOMIC STRUCTURE OF CHROMATIN'S NUCLEOSOME CORE PARTICLE AT 2 Å RESOLUTION

K. LUGER, A. MÄDER, D. F. SARGENT AND T.J. RICHMOND

INSTITUT FÜR MOLEKULARBIOLOGIE UND BIOPHYSIK, ETH ZÜRICH (SWITZERLAND)

*Protein crystallography is currently turning out novel atomic structures of biological macromolecules at the rate of 400-500 per year [1]. These structures range in size from proteins comprising under 100 amino acids to virus particles of several million molecular weight. An area of intense interest currently concerns macromolecular complexes containing multiple components that must periodically assemble and disassemble in order to carry out their biological function.*

**A** classic example in all higher cells is the nucleosome, which is the fundamental repeating unit of DNA organization in chromosomes and accounts for the two most fundamental levels of chromatin structure. Not only do nucleosomes efficiently package DNA, they are also intimately involved in the gene expression mechanisms that allow only selected regions of the vast store of genomic information to be read out as a consequence of signaling processes. These two functions require about 25 million nucleosomes in each human cell nucleus.

In 1984, the structure of the nucleosome core particle (NCP), the larger part of the nucleosome, was published at 7 Å resolution [2]. The x-ray data was collected before synchrotron radiation was generally available for protein crystallography by using a single detector diffractometer and a rotating anode source. The spatial resolution of this first structure was limited by the material itself as it was prepared from whole nuclear chromatin and was therefore heterogeneous in composition. Eventually, it became technically feasible to assemble NCP in the test tube from homogeneous components made individually in bacterial cells [3]. After a long period of experimentation for developing homogeneous preparations of NCP, crystals were obtained that diffracted to high resolution. Nevertheless, the Bragg intensities from these crystals are extremely weak. Fortunately, the ESRF had opened for business, and we were invited to collect full data sets at ID13 (C. Riek, scientist in charge) just after its construction. As a result, the structure of the nucleosome core



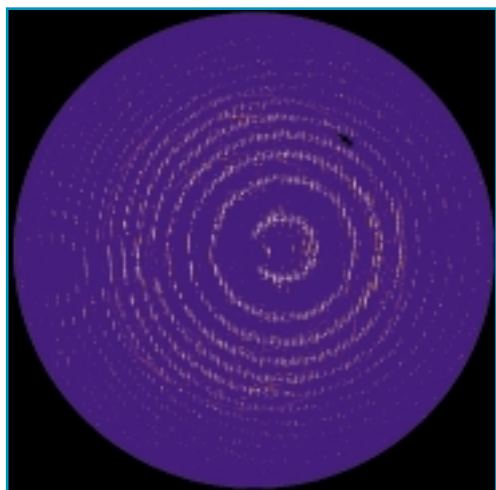
*Fig. 1: Crystal structure of the nucleosome core particle at 2.8 Å resolution. The DNA double helix (146 base pairs in two chains: turquoise and brown) is wound around the protein histone octamer (two copies each of H2A: yellow, H2B: red, H3: blue, and H4: green) in 1.65 left-handed superhelical turns. This is the form of DNA which predominates in higher living cells. The left view is down the superhelical axis. The right view is orthogonal to the superhelix and overall pseudo-twofold axis.*

particle was published in 1997 at 2.8 Å resolution (Figure 1) [4]. At 206 kDa, the NCP is the largest and most universal protein/DNA complex solved in atomic detail.

Our best nucleosome core particle crystals show Bragg intensities to 1.9 Å and have measurable data to 2.0-2.1 Å spacings (Figure 2). Most recently, 27 of these crystals were used to collect a data set to 2.0 Å resolution on ID9 (M. Wulff, scientist in charge). The high brightness of both ID9 and ID13 undulator beamlines enabled our crystallographic studies of the NCP at high resolution.

## AT 2.8 Å RESOLUTION (ID13)

The diffraction data was measured at ID13 (beam size 100 μm x 150 μm) over a period of 2 years (~75 shifts) and used to produce the first high resolution image of the NCP. In order to solve the phase problem by the multiple isomorphous replacement method (MIR), recombinant DNA technology was employed to create specific attachment sites for heavy atom derivatives in the protein portion of the complex [3]. A large number of heavy atom derivatives were screened to a



*Fig. 2: Diffraction from nucleosome core particle crystals to 2.0 Å spacings.*

*The view is nearly orthogonal to the  $a^*$  axis of the P212121 unit cell (108 Å x 186 Å x 111 Å). The X-ray wavelength is 0.842 Å and the crystal-to-detector distance is 320 mm.*

*The image has been median filtered to remove the high levels of diffuse scatter.*

resolution of 4.5 Å in successive beam times, solved by difference Patterson and difference Fourier techniques, and then ranked by phasing power. Finally, one native and four derivative data sets were collected at  $-170$  °C to a resolution of 2.8 Å. Crystal cryo-cooling procedures were also tested and developed during these sessions at the ESRF. Because each diffraction image required 1-3 min of crystal exposure to the x-ray beam, each final data set required 3-4 crystals to keep the overall effect of radiation damage on mean intensity below 20%.

## OVERVIEW OF THE STRUCTURE

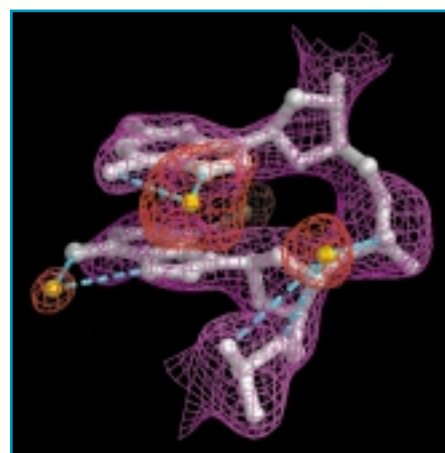
The NCP contains pairs of the four core histone protein molecules named H2A, H2B, H3, and H4, and a roughly equal mass of DNA in 147 nucleotide pairs (we used 146 bp). Compared to the nucleosome, the NCP is missing only the 'linker histone' H1 and the short stretches of DNA that connect the nucleosome cores to each other in chromatin. The core histones are arranged in an octameric unit around which the DNA is wrapped in 1.65 turns of a left-handed superhelix (Figure 1) [4]. This arrangement necessitates a substantial deformation of the DNA, bending the 22 Å diameter double helix to a mean radius of 42 Å in the nucleosomal superhelix.

The histone protein chains are divided into three types of structures: 1) rigid, folded alpha-helical domains named the histone-fold, 2) histone-fold extensions which interact with each other and the histone-folds, and 3) flexible 'histone tails'. The histone-fold domains are

structurally highly conserved between the four types of core histones and have also been discovered in an increasing number of other molecules involved in the regulation of gene read-out or transcription. They form crescent-shaped heterodimers which have extensive interaction interfaces in the pairings H3 with H4 and H2A with H2B. The histone-fold domains are responsible for organizing 121 base pairs (bp) of DNA in the superhelix, not the entire 147 bp. It is the responsibility of the extensions just prior to the H3 histone-folds to bind the first and last 13 bp of DNA. The flexible tails of the histones reach out between and around the gyres of the DNA superhelix to contact neighboring particles. About one-third of these flexible histone tails can be observed in the electron density map, the remainder are too disordered to be interpreted. The implication from the structure is that these flexible regions are meant to make inter-nucleosomal interactions, perhaps facilitating the formation of nucleosome higher-order structures (HOS).

*Fig. 3: Two consecutive bases of nucleosomal DNA (white) and solvent molecules (yellow).*

*The atomic positions are based on a 2Fo-Fc electron density map (DNA: magenta, solvent: red) calculated between 47 and 2 Å resolution (contoured at 1.1 sigma). Distances between DNA and solvent molecules greater than 2.2 Å and less than 4.4 Å are shown (cyan).*



## PROTEIN-DNA INTERACTIONS

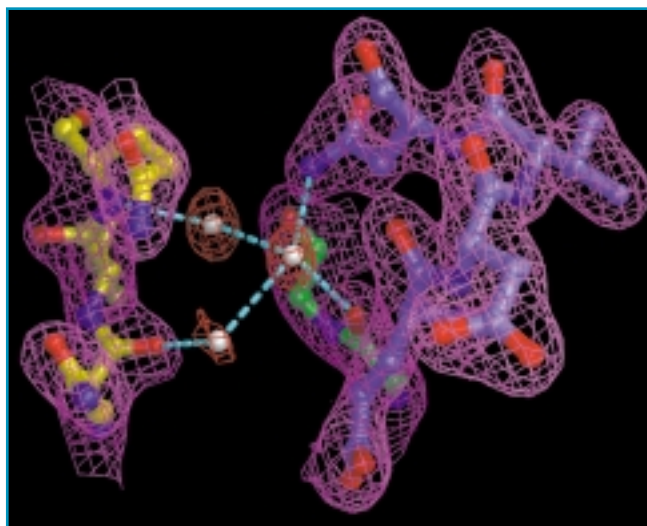
There are 14 regions of contact between the histone proteins and DNA: three by each of the four histone-fold dimers and two by histone-fold extensions. This construction allows the DNA molecule in a single nucleosome core to come loose over one-half of the superhelix while the histones maintain their grip on the other half, permitting the genetic information stored in the DNA to be read out without complete dissociation of the DNA from the histone octamer. The nucleosome core was previously thought to be held together simply by electrostatic attraction: the negatively charged DNA molecule wound as yarn around a positively charged histone spool. Although this type of interaction does occur, equally many interactions of other kinds, such as hydrogen bonds and hydrophobic interactions, are also important.

## DNA DEFORMATION

The path of the DNA around the histone octamer deviates from that of an ideal superhelix, displaying strong bends in some regions, while being nearly straight in others. This path is determined predominantly by the histone/DNA contacts and is probably largely independent of the DNA sequence of nucleotides. The close spatial proximity of the two turns of the DNA superhelix with a pitch of 24 Å, and the periodic variation of double helix parameters with a mean of 10.3 bp per turn, result in an alignment of major and minor grooves from one superhelical gyre to the next (Figure 1).



**Fig. 4: Ordered solvent molecules (white) at the interface between the protein chains of histone H2A (yellow), histone H3 (blue) and histone H4 (green). Oxygen atoms are shown in red, nitrogen atoms in dark blue. Distances between protein and solvent greater than 2.2 Å and less than 3.5 Å are shown in cyan. A 2Fo-Fc electron density map, calculated as described for Figure 3, and contoured at 1.7 sigma, is shown for the protein (magenta) and for solvent (red).**



The resulting narrow channels formed by the aligned minor grooves serve as the exit points for four of the eight basic histone tails, whereas the large pores formed by the aligned major grooves are, in principle, free to make base-specific contacts with other proteins. The Debye-Waller B-factors show that the mobility of the DNA backbone varies greatly, having low values when it is bound to the histone octamer and high values when it is facing away from it.

## AT 2.0 Å RESOLUTION (ID9)

X-ray data to 2.0 Å resolution was collected at ID9 (beam size 150 μm x 300 μm) using a monochromatic x-ray beam and a 30 cm Mar Image Plate detector (Figure 2). Crystals of an average size of 200 μm x 200 μm x 500 μm diffracted very weakly, requiring exposure times of 90 s for each 0.4° rotation picture to record the highest resolution data. The previous observations of radiation damage incurred by these crystals despite cryo-cooling indicated the need to collect a large number of partial data sets. A total of 27 partial data sets were collected, each one covering a rotation range of 8.0° and overlapping the neighboring sets by 4.0° to accommodate crystal misalignment. Two crystal orientations with respect to the rotation axis were used to ensure coverage of the blind zone. The substantial exposure times necessary to collect the high resolution data required collection of an additional 3.0 Å resolution data set using exposure times of 10 sec to recover 'overflows'. Ultimately, the program «denzo» was used to measure a total of 4,228,118 reflections from 570 exposures and to merge them to obtain our 2.0 Å data set.

The final completeness of the data is 98.3% (84.5% in the last shell: 2.04 - 1.97 Å) with an overall R-factor of 6.8% (reaching 30% in the 2.22 - 2.13 Å shell). The refinement of the structure at this resolution is currently in progress.

## WATER AND IONS

The 2.0 Å diffraction data have allowed us to locate a large number of well-ordered water molecules and ions. We expect to gain valuable general information on the role of water molecules in mediating protein-DNA and protein-protein interactions. The binding of divalent metal ions, such as manganese or magnesium, to the DNA appears to favor the distortion of the DNA seen in the NCP (Figures 3 and 4). The presence of ordered water molecules at the interface between the protein subunits may provide a means to favor their disassembly and thus could be important to the processes of DNA transcription and replication.

## OUTLOOK

The nucleosome core particle structure explains in atomic detail how DNA is kept untangled in the cell nucleus and clarifies the unique role of the nucleosome in maintaining and controlling the expression of genetic information. Nucleosomes do not exist as isolated particles in the cell, but are packed into arrays with an internal repeat of 157-240 bp. The dynamic

assembly and disassembly of the higher-order structures made from these arrays helps determine the functional state of DNA. HOS formation is apparently guided by interactions with non-histone nuclear proteins in addition to the DNA sequence preferences displayed by the histone octamer itself. The NCP structure has already provided information that has allowed us to construct a nucleofilament zigzag structure in crystals and add a major fragment of the H1 linker histone to it. Work on this structure is currently in progress at the ESRF. Beyond this, we will endeavor to elucidate the structures of higher-order arrangements of nucleosomes in the chromatin fiber and to relate this information to the way these assemblies participate in gene regulation. ■

## REFERENCES

- [1] W.A. Hendrickson and K. Wüthrich, eds., *Macromolecular Structures (Current Biology, London)* (1997).
- [2] T. J. Richmond, J. T. Finch, B. Rushton, D. Rhodes and A. Klug, *The structure of the Nucleosome Core Particle at 7 Å Resolution. Nature*, 311, 532-537 (1984).
- [3] K. Luger, T. Rechsteiner, A.J. Flaus, M. M. Y. Waye and T. J. Richmond, *Characterization of Nucleosome Core Particles Containing Histone Proteins Made in Bacteria. J. Mol. Biol.*, 272, 301-311 (1997).
- [4] K. Luger, A.W. Mäder, R.K. Richmond, D.F. Sargent and T.J. Richmond, *Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution. Nature*, 389, 251-260 (1997).