ITERATIVE HELICAL REAL-SPACE RECONSTRUCTION OF HISTONE OCTAMER TUBULAR CRYSTALS AND IMPLICATIONS FOR THE 30 NM CHROMATIN FIBER

Timothy Duncan Frouws

A minithesis submitted in partial fulfillment of the requirements of the degree of Magister Scientiae (Structural Biology) at the University of the Western Cape.

UNIVERSITY *of the*
WESTERN CAPE

Supervisors:

Associate Professor Trevor Sewell
Associate Professor Hugh Patterton

June 2006

KEYWORDS

Histone octamer

Tubular crystals

Iterative Helical Real Space Reconstruction (IHRSR)

Chromatin

Nucleosome

3D reconstruction

Negative stain

Electron microscopy

Docking

Protein structure

ABSTRACT

ITERATIVE HELICAL REAL-SPACE RECONSTRUCTION OF HISTONE OCTAMER TUBULAR CRYSTALS AND IMPLICATIONS FOR THE 30 NM CHROMATIN FIBER

T. D. Frouws

MSc minithesis, Faculty of Natural Sciences, University of the Western Cape and Faculty of Science, University of Cape Town.

The histone octamer is a quaternary complex comprising 2 copies of each of the 4 core histone proteins H2A, H2B, H3 and H4. This represents the protein core around which 1¾ turns of DNA is wrapped to form the nucleosome core particle – the first level of organization in chromatin. Nucleosome arrays are compacted into a 30nm fiber and ultimately chromosomes. The organization of the chromatin fiber has been a matter of debate for over 25 years, but some controversies can now be resolved because the atomic structure of the nucleosome core particle is known (Luger et al., 1997). Here we investigate the helical structure of core histone octamers to discover interacting surfaces and their relevance to the compaction of nucleosome arrays into the chromatin fiber.

Histone proteins were isolated from soluble chromatin prepared from chicken red blood cell nuclei. DNA was separated from histones via hydroxyapatite chromatography and core histones assembled into octamers by incubation with 2M NaCl. Octamers were isolated from free histones by gel exclusion chromatography and encouraged to form helices by prolonged incubation in 40% $(NH_4)_2SO_4$. Precipitate was stabilized by the addition of 0.2% uranyl acetate and applied to glow-discharged carbon-coated grids. Micrographs were taken at 50K magnification using minimum dose microscopy (≤ 10 e$^-$/nm^2) on a Leo 912 TEM operating at 120kV with an in-column energy filter set for zero-loss imaging. Three-dimensional reconstruction was carried out using the SPIDER (Frank et al., 1996) software suite and the iterative helical real-space reconstruction method (Egelman, 2000) using 3500 boxed helical images. The final resolution of this core histone octamer reconstruction was judged by Fourier shell correlation between two independently generated structures, to be 16 Å.

The final structure comprises stacked rings of octamers with 11-fold symmetry. The helical parameters converged to an axial rotation of -7.5° and rise of 65Å. Both the symmetry and electron density differ from a previous reconstruction performed by Fourier-Bessel methods (Klug et al., 1980), because our helix represents a different helical form of the same 2D lattice at improved resolution. Docking of the atomic structure of the octamer (Arents et al., 1991) into the electron density was done using the helical parameters, with the dyad and superhelical axes acting as constraints to reduce the docking to a 1-dimensional search. Our reconstruction is validated by the excellent fit of the atomic model and the ability to rationalize the stabilizing interactions between octamers within the helix.

Intra-ring contacts are formed between 2-fold related histones H2B and H2B', by a 4-helix bundle formed between their α3 and αC accessory helices. The relevance of this interaction in stabilizing adjacently stacked nucleosomes within the chromatin fiber is discussed. Successive octamer rings are coupled via helix dipoles formed between H3 and H2A histones. These interactions are not physiologically relevant as DNA would normally obscure this face, but the extended N-terminal tails of H3 and H2B are poised to carry out chromatin fiber interactions between successive pitches.

June 2006

DECLARATION

I declare that *Iterative Helical Real-Space Reconstruction of Histone Octamer Tubular Crystals and Implications for the 30 nm Chromatin Fiber* is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Timothy Duncan Frouws   19 June 2006

Signed..

ACKNOWLEDGEMENTS

UNIVERSITY *of the*
WESTERN CAPE

ABBREVIATIONS

3D – Three-Dimensional

601 – Widom 601 reconstitution sequence

Å – Angstrom

CAP – Chromatin Associated Protein

CCD – Charge-Coupled Device

cm^3 – cubic centimeter

FEG – Field Emission Gun

FSC – Fourier Shell Correlation

g – gram

h-bond – hydrogen bond

H1/H5 – Linker Histones

HA – Hydroxyapatite

IHRSR – Iterative Helical Real-Space Reconstruction

kDa – Kilo-Dalton

mGI – minor groove inwards

MNase – Micrococcal Nuclease

MSA – Multivariate Statistical Analysis

OD – Optical Density

PDB – Protein Data Bank

SELEX – Selected Evolution of Ligands by EXponential Enrichment

SHL – Super Helical Location

TEM – Transmission Electron Microscope

vdw – van der Waals

TABLE OF CONTENTS

CHAPTER 3: RESULTS

CHAPTER 4: DISCUSSION

CHAPTER 1: LITERATURE REVIEW

1.1 Introduction

One of the major features of eukaryotes is that DNA is not naked, but occurs as a nucleoprotein complex called chromatin. As such, chromatin forms the substrate for the nuclear functions of transcription, replication, recombination and repair. Thus understanding the structure of chromatin in context of these processes is vital to understanding their functioning and regulation.

DNA is packaged as chromatin via a folding hierarchy beginning with the fundamental unit, the nucleosome, and ending with the metaphase chromosome. Nucleosomes occur approximately every 200±40 bp throughout all eukaryotic genomes, where each repeating nucleosome consists of a nucleosome core and linker DNA (Wolffe, 1999). The nucleosome core is composed of two copies of each of the core histone proteins H2A, H2B, H3 and H4 to form an octamer around which 147 bp of DNA is wound by 1.67 turns of a negative superhelix (Richmond and Davey, 2003). The linker DNA consists of a variable stretch (10-90 bp) of DNA which joins adjacent nucleosome cores.

These repeating nucleosomes are further assembled into a 30 nm fiber, a compaction process facilitated by the linker histone H1. The linker histone binds on the outside of the nucleosome and to the linker DNA, aiding the association of the nucleosomes into a quasi-helical array. This is achieved by the partial neutralization of linker DNA charge, allowing the close approach of the linker DNA within the fiber structure.

The last decade has seen a paradigm shift from viewing chromatin as an inert packaging of DNA, to that of a dynamic structure which modulates gene expression by regulating the access and recruitment of transcription factors. The nucleosome effect on transcription can be two-fold. Firstly, the precise positioning of a nucleosome on a

promoter can provide a direct binding site for transcription factors via the unique structure of DNA presented (Cirillo et al., 1998). Secondly, the histone tails are subject to a variety of post-translational modifications which in combination form an epigenetic code that can be read off by various factors (Strahl and Allis, 2000). This is a code hierarchy above that of DNA and labels the underlying DNA which is directly unreadable. Based on the code, various factors can collaborate to switch between transcriptionally-inactive heterochromatin, and active decondensed euchromatin.

To begin to understand these biological roles, structural details of the nucleosome and chromatin fiber are prerequisites. Particularly in the context of chromatin, direct structural evidence has often resolved longstanding biochemical ambiguities. This literature review serves to describe the fundamental structure of chromatin as solved by the various structural methods. On one end of the scale, crystallography provides atomic details of the nucleosome core particle (NCP), while on the other, EM provides details how these building blocks are further assembled into higher order structures. Consolidation of these methods results in a unified view of the structure of chromatin.

1.2 Nucleosome Structure

1.2.1 *Historical Overview*

It was long known that bulk chromatin consisted of a nucleoprotein complex formed by the five histone proteins and DNA (Kossel, 1928) - some 50 years before DNA was even known to be the hereditary molecule, and both histones and DNA were still contenders for this title. The idea that these histones were arranged with DNA as regular repeating units termed nucleosomes arose from a combination of biochemical, electron microscopic, and X-ray diffraction studies - for references see (Kornberg, 1977; Kornberg and Lorch, 1999; Oudet et al., 1975; Olins and Olins, 2003).

Endonuclease digestions of native chromatin resulted in a characteristic ladder of fragments which differed by increments of ±200 bp. These repeats were confirmed by X-

ray diffraction and electron microscopy and revealed the regular repeat was composed of a roughly spherical 100 Å particle. Furthermore, reconstitutions with equimolar amounts of each histone to naked DNA could restore a similar X-ray diffraction pattern or EM observation - without a requirement for the linker histone. Chemical cross-linking of native chromatin produced predominantly octamer species, showing that the nucleosome contained two copies of each of the four core histones. Extended endonuclease digestion resulted in a ~147 bp metastable limit, indicating that this was the length of DNA associated with the nucleosomal core. DNase I digests further confirmed this length, and was used to determine the rotational position of DNA on the nucleosome. Because the nucleosome core particle (NCP) did not contain the variable linker DNA stretches, it represented a more distinct biochemical species than the nucleosome, and therefore became the target for ensuing structural studies.

1.2.2 *Histone Octamer Tubes*
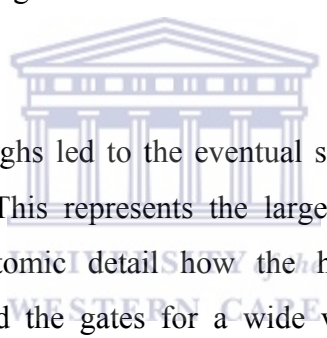
In the absence of DNA, the histone octamer is unstable and disassembles into a H3-H4 tetramer and two H2A-H2B dimers. The discovery that the octamer is a stable unit in high salt was the basis for crystallography trials thereof. But, an early attempt to crystallize the octamer resulted in helical tubes being formed instead - suitable for electron microscopy and 3D reconstruction by Fourier-Bessel methods (Klug et al., 1980). Although the 3D map produced in this manner was of low quality (22 Å), it allowed some important deductions to be made. The octamer dimensions were discovered and shown to contain a 2-fold axis. Corroborating this with previous cross-linking experiments (Mirzabekov et al., 1978), it allowed the rough assignment of the core histone positions within the map. The H3-H4 tetramer was centrally located and straddled the 2-fold axis, while the H2A-H2B dimers occupied the periphery. Because the linear order of histones along DNA was known to be H2A-H2B-H4-H3-H3'-H4'-H2B'-H2A' (Mirzabekov et al., 1978), the potential surface for the wrapping of DNA around the octamer could also be predicted. This represented an early implementation of a bioinformatics approach to interpreting EM maps, which is now common practice.

1.2.3 *The Crystallographic Challenge*

Along with the ribosome, the nucleosome had long been considered the holy grail of crystallography. The fame of these structures is due to their ubiquitous nature, and obvious importance within the cell. The challenge posed for crystallography by the nucleosome was huge and it took some 20 years from the nucleosome hypothesis to atomic structure (Kornberg, 1977; Luger et al., 1997a). The NCP is a large 206 kDa complex making it difficult to phase by isomorphous replacement. Additionally, 30% of the core histone sequences occur in unstructured tail domains, and NCPs produced by nuclease digestion contain bound mixed sequence DNA. This led to disorder in NCP crystals, resulting in weak diffraction. A low-resolution structure (7 Å) from NCPs prepared in this way provided a view of how the DNA was organized around the histone octamer in 1.75 left-handed superhelical turns (Richmond et al., 1984). The wrapping was non-uniform and possessed sharp kinks where the minor groove contacted the octamer surface. Crystallization of the histone octamer alone at 3.1 Å was a major breakthrough, and first described the architecture including the histone fold motif of the core histones and how they assemble via the handshake and 4-helix bundle motifs (Arents et al., 1991). And it was through the reinterpretation of the 7 Å NCP structure, that the rod-like densities could then be assigned to the individual alpha helices of the core histones (Richmond et al., 1993). This revealed how repetitive DNA binding elements formed by the histone dimers (Moudrianakis and Arents, 1993) are positioned to contact the minor groove of DNA as it faces the octamer every 10 bp, thereby causing those local distortions.

Achieving a high-resolution structure required engineering NCPs to a higher level of structural homogeneity to improve diffraction - reviewed by (Harp et al., 2000; Luger et al., 1998). Using a defined sequence of DNA would help, but the problem was how to incorporate this into a structure which in nature has evolved to package random DNA. Sequence rules were later identified which predisposed them to package favorably into nucleosomes and used to successfully identify candidate sequences in human α-satellite DNA (Uberbacher et al., 1988). Using one of these sequences and cloning it as a palindrome of two 73 bp segments, was rationalized as a way of extending the 2-fold axis

of the octamer to the entire NCP (Harp et al., 1996). The crystal asymmetric unit would then be halved, simplifying solution of the structure. Despite this effort, the NCP still crystallized as an asymmetric unit, because a single base pair from one palindrome section straddled the would-be 2-fold axis of the NCP (Harp et al., 2000; Luger et al., 1997a) – a result predicted by earlier hydroxyl radial mapping (Flaus et al., 1996), but not heeded to. This splits the NCP by a pseudo-dyad axis of symmetry, separating one half of the NCP to contain 72 bp and the other 73 bp. Fortuitously, the use of a palindrome did at least work in reducing the 2-fold packing disorder between the two faces of the NCP, leading to improved diffraction. The use of recombinant histones free from post-translational modifications further reduced heterogeneity when reconstituted with the palindrome, and also made it possible to introduce cysteines - aiding the incorporation of heavy atoms (Luger et al., 1997b). Synchrotron radiation, cryo-crystallography, and macromolecular crystal annealing were then essential to maximize the diffraction from weakly diffracting crystals.

These technological breakthroughs led to the eventual solution of the NCP structure at 2.8 Å (Luger et al., 1997a). This represents the largest single stretch of DNA ever crystallized, and shows in atomic detail how the histones organize it into 1.65 superhelical turns. This opened the gates for a wide variety of high-resolution NCP structures from various organisms - including frog (Luger et al., 1997a), chicken (Harp et al., 2000), yeast (White et al., 2001), and human (Tsunaka et al., 2005); with different sequences (Davey et al., 2002); containing histone variants (Suto et al., 2000); containing histone mutants (Muthurajan et al., 2004); and even bound minor-groove ligands (Suto et al., 2003). The highest resolution structure to date at 1.9 Å resulted by including a central base pair between the 73 bp segments (Luger et al., 1998). Over 3000 water molecules (Davey et al., 2002); 18 ions (Davey and Richmond, 2002); and the exact DNA base-step parameters (Richmond and Davey, 2003) could subsequently be identified and detailed how these contribute to histone-histone/DNA interactions. The histone octamer in isolation has also been extended to both 2.15 Å and 1.9 Å resolutions (Chantalat et al., 2003; Wood et al., 2005). The following sections will draw upon these structures to detail the architecture of the nucleosome core particle.

1.2.4 *The Core Histones and Heterodimers*

The four core histones comprise a central structured domain (figure 1.1A), flanked by various disordered N- and C-terminal tail domains and ordered extensions (figure 1.1B). In the conserved central domain they all share a common fold termed the histone fold motif, first discovered and described by Arents et al. (1991,1995) with their 3.1 Å octamer structure. The elements of the histone fold motif (figure 1.2A) are 3 α-helices connected by 2 loops, abbreviated as α1-L1-α2-L2-α3 using the nomenclature of Luger et al. (1997). The long central alpha helix α2 is flanked on either end by two short terminal helices α1 and α3, which rotate and fold back over approximately the same side of the central helix - in this way the motif possesses a rough two-fold axis of symmetry about the middle of α2. As will be seen, symmetry, pseudo-symmetry and repetitive motifs are frequently used construction principles in the NCP.
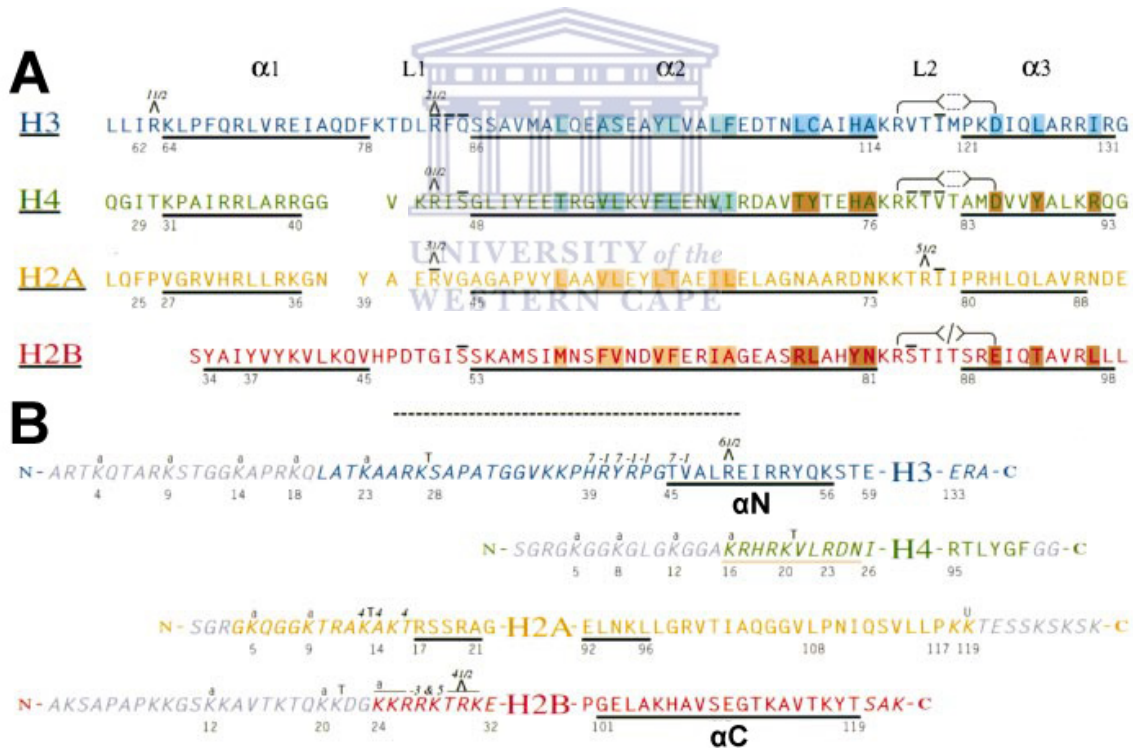
Figure 1.1 Core histone sequences. **(A)** The histone fold regions are aligned based on their conserved central histone fold motifs and labeled structural elements. **(B)** Tails and extensions external to the region shown in (A). Helices are underlined. Colors shown here for core histones are conserved in all future figures. Figure reproduced from Luger et al. (1997).
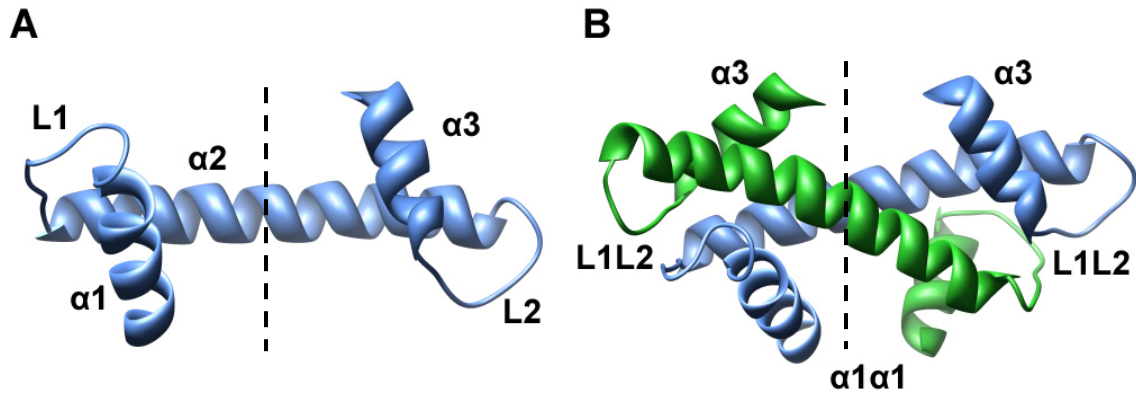
Figure 1.2 (A) The histone fold motif of H3 and labeled elements. **(B)** The H3-H4 heterodimer pair and created L1L2 and α1α1 sites. Pseudo 2-fold axes shown as dashed lines and tails are deleted for clarity. PDB coordinates are from 2HIO (Arents et al., 1991).

Core histones assemble as the heterodimer pairs H2A-H2B and H3-H4, via the handshake motif (Arents et al., 1991). Histones pair across their central α2 helices in an antiparallel manner over a pseudo two-fold axis, with their terminal helices facing inward toward the interface (figure 1.2B). This clasping shape creates a large interaction surface which is stabilized by hydrophobic forces. Because the center of interaction is offset from the histone fold's own pseudo two-fold axis by one helical turn of α2 toward the N-terminus, the heterodimers form a crescent shape. This arrangement places the L1 loop parallel to the L2 loop of the other histone in the pair to form a L1L2 ß-bridge on both ends, and strengthens dimer stability via hydrogen bonds. The L1L2 loops form a left-handed spiral and are one of two types of DNA-binding sites created in the dimer. The offset also creates the other site called α1α1 by pairing both α1 helices on the convex side of the dimer crescent, and these helix dipoles point their basic ends outward. Conversely because of the offset, on the opposite concave side of the dimer, the α3 helices are not in close proximity.

1.2.5 *The Histone Octamer*

The histone octamer is assembled by associating the heterodimers together via three 4-helix bundles formed between their C-terminal ends (Arents et al., 1991). These 4-helix bundles are stabilized by hydrophobic interactions and hydrogen bonds formed between buried charged groups. At the center of the octamer, two H3-H4 dimers interact via the C-terminal halves of the α2 and α3 helices of H3 and the symmetrically related molecule H3', to form a 4-helix bundle.
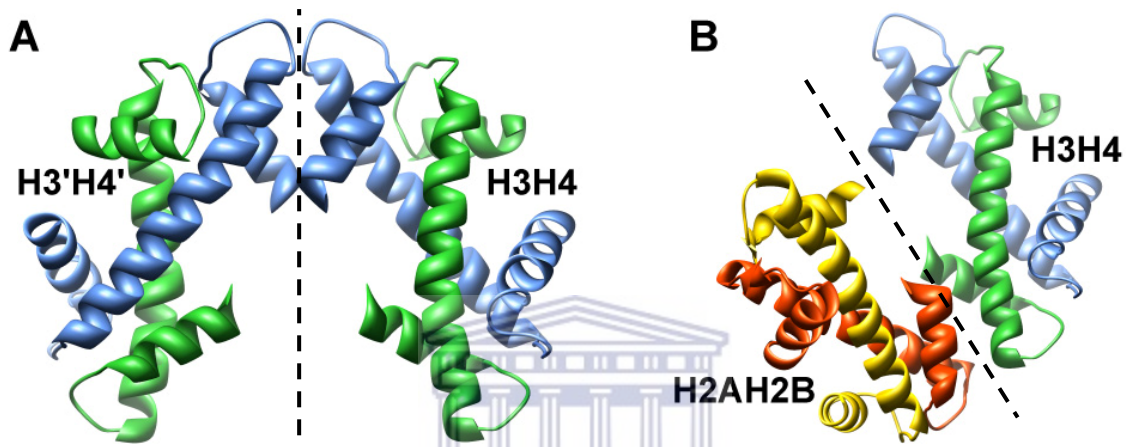
Figure 1.2 4-Helix Bundles. **(A)** 4-Helix bundle formed across the crystallographic 2-fold axis between H3 and H3' (B) 4-Helix bundle formed between H4 and H2A across a pseudo 2-fold axis. PDB coordinates are from 2HIO (Arents et al., 1991).

The H3-H3' interaction occurs across the crystallographic 2-fold axis of the octamer, and forms the H3-H4 tetramer which is the central core of the octamer (figure 1.3A). Finally, the flanking units of H2A-H2B heterodimers are added to either end of the tetramer via two 4-helix bundles - created by the C-terminal ends of H2B and H4; and symmetrically related H2B' and H4' (figure 1.3B). It is suggestive to think the remaining C-terminal element in H2A might be used to form a further internucleosomal 4-helix bundle, but this is unlikely since the octamer face is obscured by the more terminal αC accessory helix of H2B. Apart from the 4-helix bundles formed, the octamer is stabilized by the formation of a docking domain between H2A'-H2B' and H3-H4 dimers across the 2-fold axis. The C-terminal docking sequence of H2A' serves to position the H3-αN helix for DNA binding, and also forms a parallel ß-sheet with the H4 docking sequence (Chantalat et al., 2003; Luger et al., 1997a).

Figure 1.4 The complete histone octamer. Crystallographic 2-fold axis is shown as a dashed line. PDB coordinates are from 2HIO (Arents et al., 1991)

Because each heterodimer pair is crescent-shaped and an angle joins adjacent heterodimer pairs through the 4-helix bundle, these domains spiral inward to complete the superhelical architecture of the octamer (figure 1.4). This overall shape of the octamer generates three characteristic views from orthogonal perspectives, as described by Arents *et al.* (1991). When viewed directly down the dyad axis the octamer resembles a left-handed propeller, referred to as the front view (figure 1.5A). This view most reflects the tripartite nature of the H3-H4 tetramer and flanking H2A-H2B dimers. From the top view (figure 1.5B) the octamer appears as a bipartite wedge, with the narrow end formed by the tetramer. Finally, when seen from the side, down the superhelical axis (figure 1.5C), it is shaped like a disc with a diameter of 65 Å.
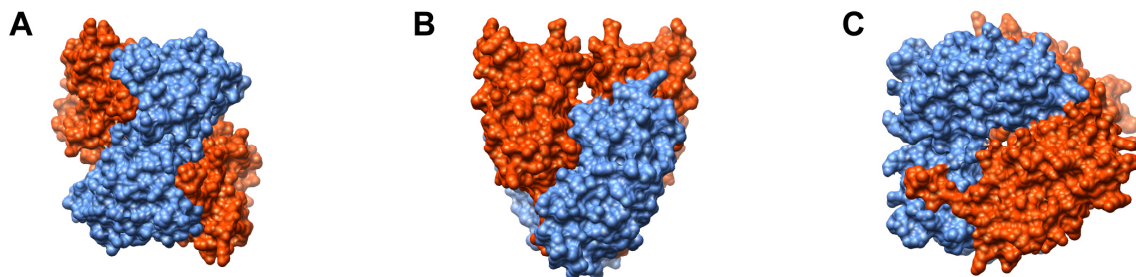
Figure 1.5 The three distinctive orthogonal views of the histone octamer surface. The central H3H4-H3'H4' is shown in blue, while the two peripheral H2A-H2B pairs are shown in orange-red. **(A)** Front propeller view **(B)** Top wedge view. **(C)** Side disc view – seen down the superhelical axis. PDB coordinates are from 2HIO (Arents et al., 1991)

These shapes were seen by Klug et al. (1980) at 22 Å, and furthermore they correctly assigned the superhelical surface around which DNA was potentially wrapped in the NCP. This was later confirmed by Arents *et al.* (1991) at 3.1 Å, who additionally observed the ridges and grooves which traversed this surface, representing the repetitive DNA binding elements (Moudrianakis and Arents, 1993).

1.2.6 *Histone-DNA Interactions*

The core of the nucleosome comprises 147bp of DNA coiled around the histone octamer in a 1.67 turn left-handed superhelix (Richmond and Davey, 2003). The pseudo-dyad axis divides the NCP into two half particles, each containing identical 73 bp segments of the palindrome used for crystallization (figure 1.6A). A central base pair straddles this axis and is referred to as superhelix location 0 (SHL), using the notation of Luger et al. (1997). Proceeding along the DNA into either half-particle therefore increases the SHL number till +7 or -7. This is a convenient notation, because as the helical repeat of DNA in the NCP is close to 10 bp, the SHL therefore defines the rotational orientation of DNA relative to the octamer surface. There are 14 distinct binding sites in the NCP – at every minor groove-inward (mGI) position, as expected for non-sequence specific DNA binding. Because at SHL0 the minor groove faces outward, these 14 mGI sites are located from SHL 0.5 through to SHL ±6.5. The central 12 of these binding sites organize 121 bp and are created directly by the histone fold motifs through eight L1L2 and four α1α1 binding sites. Each heterodimer is responsible for binding a 2.5 turn segment of DNA and inducing a 140º bend as it conforms to the convex surface of its crescent shape (figure 1.7). Bending is not smooth, and kinks are induced into the minor groove by associated histone binding.
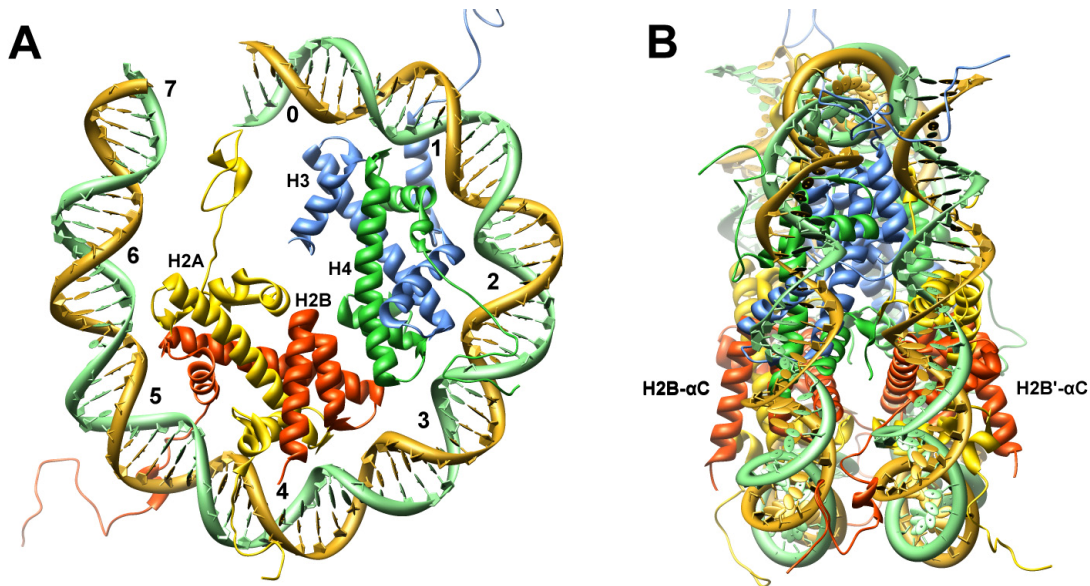
Figure 1.6 The nucleosome core particle. **(A)** The half-particle showing a 73bp segment of the DNA palindrome and four core histones. Super helical locations (SHL) are numbered, and a central base-pair straddles he 2-fold axis at SHL0. **(B)** View of the full NCP perpendicular to the superhelical axis. PDB coordinates are from 1KX5 (Richmond and Davey 2003).

The following generalities can be made about the interactions that take place upon DNA binding (Luger et al., 1997a): Firstly at every mGI position, an arginine residue is inserted into the minor groove, whose guanidinium group simultaneously binds to two phosphate groups. These two phosphate groups are located on opposite strands 5 bp apart and the binding results in a pinching together of the minor groove as this phosphate-to-phosphate distance is decreased. An outcome is unusual base step parameters and kinking at the mGI positions, while in contrast the major groove-in positions are more smoothly bend inwards. The arginine residue is in most cases restrained from penetrating the minor groove too deeply by being hydrogen-bonded to a neighboring threonine. Secondly, the negative dipole moments produced at the N-terminus of the α1 helices in the α1α1 site and α2 helices within the L1L2 site, are each used to fix the position of a single phosphate moiety (Except in the case of the deviant H2A-α1 which possesses an additional αN helix to aid binding). Thirdly, this helix dipole is conveyed to the phosphate via direct hydrogen-bonds by main-chain amide groups located in the first turn of the helix. Fourthly, the other half of direct hydrogen bonds and salt-links are made by basic and hydroxyl amino acid side chains, like from lysine and threonine residues. Fifthly, numerous hydrophobic contacts form between non-polar amino acids and

deoxyribose rings. Finally, all these interactions mentioned can be enhanced by water molecules, only seen later once the resolution of the NCP structure was extended to 1.9 Å (Davey et al., 2002). Within the NCP they serve as adaptor molecules to either extend the reach of protein side chains to link with distant phosphate groups, or to facilitate existing bonds by forming hydrogen-bond networks. In extreme cases they even link glutamate carboxyl or main-chain carbonyl groups to DNA phosphodiesters. In this way, as adaptors, they would facilitate the accommodation of variable DNA sequences, as encountered in nature. Overall, water molecules are thought to lower the entropic contribution of binding, lowering the thermodynamic barrier to octamer sliding – important in chromatin remodeling.

Architecturally, the α1α1 element uses both helices as the primary determinate for DNA binding. While in the L1L2 site, the main mode of binding is through the adjacent loops which spiral together in a left-handed sense to match the curvature of phosphate backbone of the minor groove. In this way each loop follows along opposite strands of DNA, creating a large interaction surface. The long α2 helix also contributes a dipole moment to this motif.
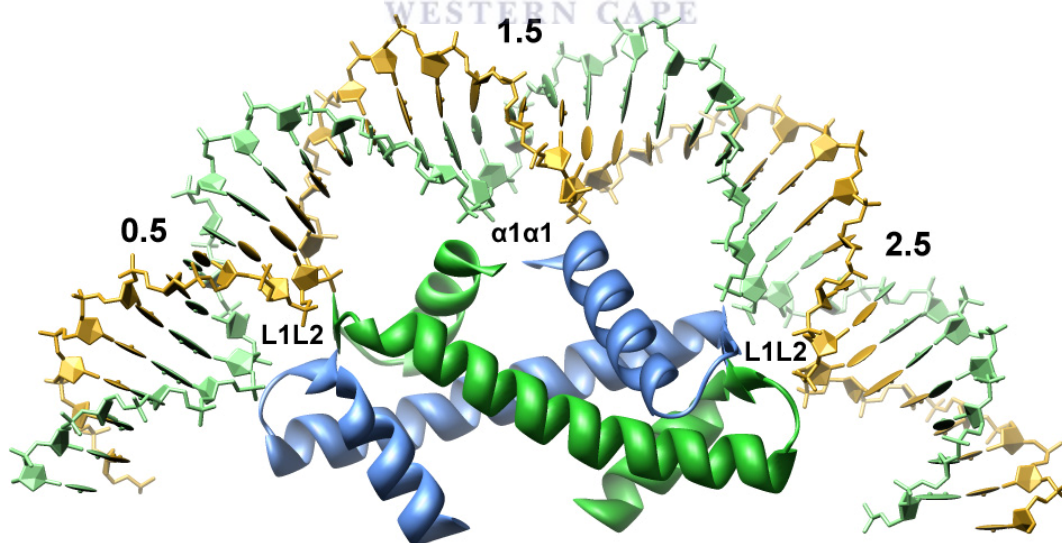
Figure 1.7 The H3H4 heterodimer pair bound to roughly 2.5 turns of DNA helix. The three binding motifs and SHLs are labeled. Tails are deleted for clarity. PDB coordinates are from 1KX5 (Richmond and Davey 2003)

Outside of the histone fold, the terminal 13 bp segments of DNA (SHL±6.5) are bound by the H3-N accessory helix and preceding tail region. As a consequence of the different binding mode, these terminal segments are relatively straight, contributing little to the 1.67 turns of the DNA superhelix (figure 1.6A).

1.2.7 *Nucleosome Positioning*

Nucleosomes are naturally evolved to package the variety of DNA sequences present in the genome, and are mobile under physiological conditions existing in equilibrium at a variety of positions (Meersseman et al., 1992). However, 5S rRNA gene sequences were discovered to contain a precisely phased nucleosome, specifically correlated with gene regulation (Rhodes, 1985; Simpson and Stafford, 1983) − thus becoming the first reconstitution sequences for *in vitro* studies.

As nucleosomes do not make base-specific interactions, positioning effects should result from indirect DNA sequence-dependent flexibility lowering the free energy of DNA distortion into the 1.67 superhelical turns needed to conform to the histone octamer surface. For this anisotropic flexibility, G-C and A-T tracks are distributed with the helical repeat of DNA, because of their tendency to bend into the major and minor grooves respectively (Ramakrishnan, 1997). These simple rules however, lead only to a strong rotational positioning which can repeat every ~10 bp.

Translational positioning results from more subtle effects regarding the distribution of specific dinucleotide steps within a sequence, which are more poorly understood (Widom, 1998). Ultimately, this relates to the free-energy cost of local sequence-dependent DNA distortions at the various mGI positions and concomitant ability to form compensating bonds with the underlying histones. This in turn is related to the specifics of the two types of DNA-binding sites α1α1 and L1L2, and distribution of local curvature throughout the nucleosome − with maximal curvature occurring at SHL ±1.5 and ±4.5

(Richmond and Davey, 2003). Nonetheless, full understanding will require more NCP structures at atomic resolution on a variety of sequences.

This is where the power of a selectional approach has been useful in generating a reconstitution sequence with 150 times more affinity than the previously used 5S sequences (Thastrom et al., 2004). This Widom '601' sequence was isolated from a random pool of oligonucleotides by SELEX methods, which involves successive rounds of selection and enrichment in an *in vitro* evolution setup (Lowary and Widom, 1998).

1.2.8 *Tails, Extensions and Internucleosomal Contacts*

The tails of the nucleosome beyond the central histone fold motif and extensions make up 28% of the total amino acids and are highly basic, because over a third of their residues are lysines and arginines (figure 1.1) (Luger et al., 1997a). Combined with the observation that chromatin compaction displayed a salt-dependency (Thoma et al., 1979), this led to the idea that the compaction of chromatin was mediated by the neutralization of DNA charge by the tail regions (Clark and Kimura, 1990). The crystal structure of the octamer revealed the tails to be too disordered to be visualized (Arents et al., 1991), but it was expected that they would order upon DNA binding in the NCP structure (Dutnall and Ramakrishnan, 1997). However, this turned out only to be true for the short segments bound to minor grooves as they exit the NCP.

Here random-coiled N-terminal tails of H3 and H2B pass between gaps created by adjacent minor grooves in the superhelix (Luger et al., 1997a). Basic amino acids coordinate with the minor groove as they pass through it, and out of the fourteen arginines residues inserted into the minor groove, four are donated by these segments. The remaining tails of H2A and H4 also exit along the minor groove, but on the outer turn of the superhelix by the histone octamer face. All the histone tails then extend away from the NCP roughly perpendicular to the superhelical axis, and mostly form crystal contacts with neighboring NCPs – chiefly to their DNA minor grooves. In the same way

in vivo, this exit of the tail regions is likely important in mediating distant internucleosomal contacts and quenching linker or nucleosomal DNA through an otherwise obscured surface of the NCP (Hansen, 2002). This could stabilize the chromatin fiber across the groove formed between successive helical turns. Thus, their function seems separate from the core of the nucleosome and they confer little in the way of stability to the 1.67 turns bound (Bohm and Crane-Robinson, 1984).

The exact way in which histone tails mediate their effects mechanistically however, is difficult to interpret from the x-ray structures, and care should always be exercised in extrapolating higher-order interactions from crystal contacts. Although the 1.9 Å structure extended the assignment of the tail regions to an almost complete view, it showed they possessed no secondary structure and are mobile elements with concomitantly high B-factors (Davey et al., 2002). Of the areas which are more ordered, where the tails contact DNA, they do not display any clear binding motif. Furthermore, the interactions that are made by each tail are different in each half-particle, showing that they are polymorphic – imposed by the 2-fold packing disorder in the crystal. Another reason would be the loss of the natural substrate for some of the tails in the NCP structure. The H3 and H2A tails have been shown to preferentially interact with linker DNA (Angelov et al., 2001), and due to its absence are forced to rearrange upon the DNA of the NCP instead – producing artefactual interactions.

The tail of H4 is an exception, since in the crystal structure of the *Xenopus laevis* NCP, it forms a distinct protein-protein interaction with the H2A-H2B dimer on the face of an adjacent NCP (Luger et al., 1997a). Numerous hydrogen bonds are made between the basic side chains of H4 (K16, R19, K20, K23) and an acidic charge patch formed by H2A (E56, E61, E64, D90, E91, E92) and H2B (E110) (figure 1.8A).
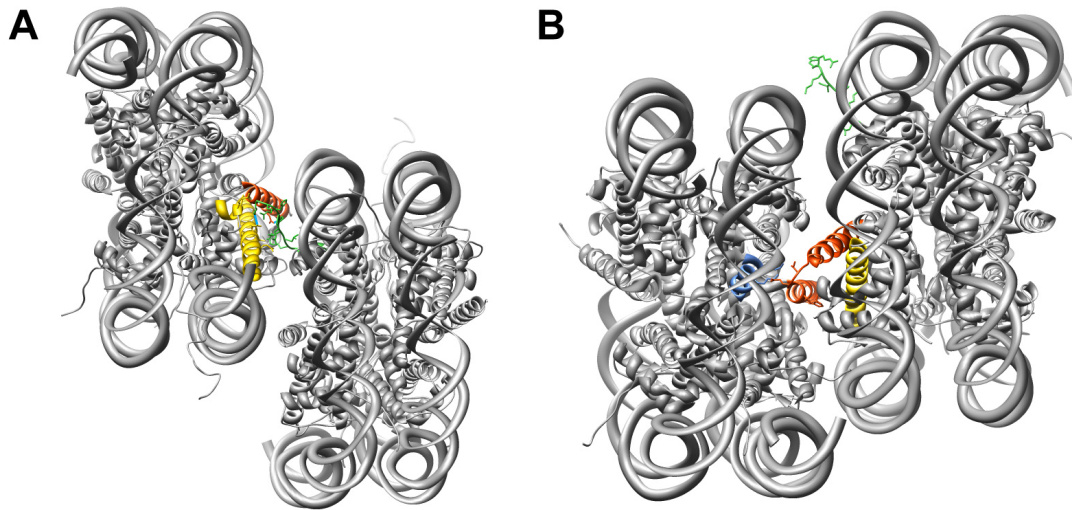
Figure 1.8 Differences in stacking interactions within the crystal lattice. **(A)** *Xenopus laevis* NCP. **(B)** Yeast NCP. The N-terminal tail of H4 is colored green, while selected helices are colored as follows: H2A – yellow, H2B – red-orange, H3 – blue. PDB coordinates are taken from 1AOI (Luger et al., 1997a) and 1ID3 (White et al., 2001).

The spatial arrangement of NCPs due to the stacking interactions is not compatible with any models of the chromatin fiber, due the superhelical axes being largely offset and the pseudo-dyad axes rotated by 180° (figure 1.8A). Regardless of this, the H4 tail interaction could still be maintained due to the flexibility of the tail, and two experiments lend credence to this idea. Firstly, H4-V21C/H2A-E64C cross-links stabilize *in vitro* assembled nucleosome arrays in their compact form (Dorigo et al., 2004). Secondly, the deletion of the amino acid stretch 14-19 in H4 critically retards the ability of defined nucleosome arrays to fully compact (Dorigo et al., 2003).

H4 may be the most important tail, but deletion of any single histone tail is generally compensated for by the others and the effects are not simply additive and each tail is partially redundant *in vitro* (Dorigo et al., 2003). This implies that because of their basic charge they function as tethered polycations, non-specifically neutralizing DNA charge. Even tailless nucleosome arrays can compact to near wild-type levels at elevated divalent cation concentrations. This cation dependence illustrates the primary mechanism of the tails is electrostatic in nature, and allows the close approach of DNA within a compact fiber (Clark and Kimura, 1990). And due to the length and flexibility of the tails, these

effects can be mediated from a distance and accommodate the variability inherit within the chromatin fiber. In fact, a problem inherent in x-ray crystallography is the anthropic selection for ordered structures – thus leading to the belief they are physiologically ordered. Thus a technique like NMR is better suited to study these tails, as variability is incorporated as a feature and their dynamic nature can be truly appreciated (Bang et al., 2001).

Besides the tails, other short-range interactions may come into play as the nucleosomes closely approach, and the histone octamer faces dock upon full compaction. A distinctive feature of the open disc face of the NCP is the αC accessory helix of histone H2B. This is a likely candidate to mediate such internucleosomal interactions due to its exterior position and the fact that it is the only extension not involved in octamer assembly or DNA binding. This would be a *cis* interaction - between the faces of adjacent nucleosomes within the same turn of the chromatin fiber. In addition, three lysines (K105, K113, and K117) on this helix are known targets for acetylation (Cosgrove et al., 2004), likely used to disrupt higher order structure during chromatin remodeling.

The H4 charge patch contact is not seen in the crystal structures of either the yeast NCP or the tetranucleosome (Schalch et al., 2005; White et al., 2001) – and H4 is positioned to interact with DNA instead. Yeast NCPs stack alternatively with their superhelical axes nearly superimposed (figure 1.8B) – more compatible with current models for the 30 nm fiber. Their pseudo-dyad axes are still reversed however, as wedge-shaped octamer faces are expected to pack more efficiently within the crystal in an alternating fashion. In the chromatin fiber, it is reasonable to postulate that this wedge-shape with its pseudo-dyads aligned will cause the inward spiral necessary to generate a helical structure.

In the yeast NCP (White et al., 2001), this alternate packing is a direct result of two new crystal contacts formed using the H2B-αC accessory helix. Two residues (T128, N129) at the C-terminus of H2B-αC hydrogen bond to lysines on H3-α3 (K121, K125) (figure

1.8B). In the other interaction E108 and H112 in the N-terminus of H2B'-αC of an adjoining NCP coordinate a manganese ion together with H2A-α2-E65 and H2B-L1-H52 (figure 1.8B). Although, this interaction mode may seem more physiologically-relevant than the H4 charge patch, as yet there is no evidence to support this.

The various octamer crystal structures also display numerous contacts involving H2B-αC, although never between the disc faces (Chantalat et al., 2003). With the absence of DNA, the constraint to limit protein-protein interactions to the octamer faces is gone.

1.2.9 *The Histone Code*

Up till now the histone tail domains have only been considered in a purely mechanistic sense, but are also subject to a variety of post-translational modifications. The best studied include acetylation, methylation and phosphorylation and are set to modulate chromatin function through two distinct mechanisms. Immediately obvious is that modifications which either neutralize charged lysine and arginine residues, or introduce negative charge, are predicted to directly alter the charge of the tail domains and thereby inhibit the electrostatic chromatin folding mechanism (Clark and Kimura, 1990).

However, the situation is more complex than this, and tail modifications can also mediate their effects indirectly through the recruitment of various chromatin associated proteins. In this sense, the extended tails serve as ideal signaling platforms, and act as tags for labeling the potential of underlying DNA, and thus represent an epigenetic code above the level of the canonical DNA code (Strahl and Allis, 2000). The specific order, spacing and combination of these chemical markers means proteins can communicate through a common substrate of the histone tails in a language of histone modifications, specifying distinct downstream enzymatic events including: the addition or removal of covalent modifications; various forms of ATP-dependent chromatin remodeling; or the induction of hierarchal levels of chromatin folding stabilized by specific chromatin architectural proteins (Luger and Hansen, 2005; Woodcock, 2006). Furthermore, the discovery of discrete chromatin modification and recognition domains, means that these activities can

be combined into multi-subunit complexes in order to achieve modular catalytic effects (Langst and Becker, 2004).

More recently, proteomic techniques have also discovered numerous covalent modifications occurring on the structured portion of the octamer (Cosgrove et al., 2004). When occurring on the lateral surface, these modifications are postulated to act directly through the alteration of DNA-binding affinity within the nucleosome and thereby alter nucleosome mobility and DNA breathing rates. *In vivo*, this can promote the maintenance of a transcriptional state or transient exposure of regulatory elements on the packed DNA respectively. However if similar modifications are instead located on the open octamer face of the nucleosome, they can possibly be responsible for the disruption of internucleosomal interactions, and chromatin higher-order structure in preparation for transcription.

1.3 Higher Order Structure of Chromatin
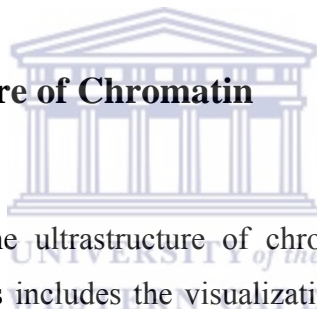
Most information regarding the ultrastructure of chromatin has come from electron microscopic (EM) studies. This includes the visualization of the nucleosomal repeat of chromatin (Olins and Olins, 1974); the 30nm fiber (Finch and Klug, 1976); and the role of salt and linker histones its compaction (Thoma et al., 1979). Additionally, EM has been a good complement to the wealth of biochemical data regarding chromatin and has helped visually confirm results.

The power of EM comes with its ability to view individual chromatin fibers individually – this in contrast to biochemical experiments which usually measure the ensemble average (Woodcock and Horowitz, 1997). This gives it the ability to deal with extreme sample heterogeneity inherent to chromatin, and as such rivals other structural techniques such as X-ray crystallography and NMR - X-ray crystallography requiring the growth of crystals and both techniques suffering from an upper size limit.

EM is not without its own drawbacks however, and it suffers from low resolution and noisy individual images. Image averaging techniques are thus used to improve resolution to usable levels, but this requires multiple images of the same molecule. Additionally, use of the single-particle technique (Jiang and Ludtke, 2005) for the 3D reconstruction of an object requires multiple views from different angles. Therefore either a homogenous sample is required, or otherwise the inherent symmetry of the object supplies the additional angular views (e.g. viruses and helices). Because chromatin displays heterogeneity above the level of the nucleosome, these techniques cannot be used. Tomography is one solution to the problem, however the sensitivity of biological samples to electron beam exposure, limits the amount of usable views that can be taken (Woodcock et al., 1991).

There is renewed interest in chromatin ultrastructure (Adkins et al., 2004), following the observation that covalent modification of core histone tails alters chromatin fiber dynamics and recruits other chromatin associated proteins (CAPs) via the histone code (Allis 2001). EM has already started to deliver structural information concerning these transitions through the newer techniques of electron cryo-microscopy (ECM) and electron spectroscopic imaging (ESI). New physical techniques like atomic force microscopy (AFM) and scanning force microscopy (SFM) and force measurements are also complementary and are likely to play a major role unraveling the structure of chromatin (Zlatanova and Leuba, 2003).

1.3.1 *Chromatin-Folding Nomenclature*

Prior to a discussion of chromatin ultrastructure, the hierarchical levels of chromatin folding need to be defined, as laid out by Woodcock and Dimitrov (2001). The primary level represents the linear array of nucleosomes on DNA, with a 10 nm beads-on-a-string appearance. These arrays are folded under increasing salt concentration via the intermediate zigzag chromatin fiber, into the maximally compacted 30 nm fiber - forming the secondary structure which involves nucleosome-nucleosome interactions. Tertiary structures involve intercalation and folding back of the 30 nm fibers, which can be

assisted by CAPs such as MENT, HP1, MeCP2 and Sir3 (Adkins et al., 2004). Beyond this level evidence is contradictory and it is difficult to assign an exact hierarchy, but the final levels likely consist of a helically coiled 30 nm fiber further organized into radial loops anchored to a scaffold core (Horowitz-Scherer and Woodcock, 2005). Topological supercoiling in each radial loop has been implicated in chromosome condensation with the identification of the condensin complex (Woodcock and Dimitrov, 2001).

1.3.2 *Historical Overview*

Using a low-salt spreading procedure (Miller, Jr. and Beatty, 1969), the first EM images of chromatin revealed the ubiquitous beads-on-a-string architecture (Olins and Olins, 1974). Taken together with the corroborating biochemical evidence, the nucleosome was born (Kornberg, 1977). How this fundamental repeating unit of chromatin was further packaged into the nucleus and chromosomes however, was unknown.

Finch and Klug (1976) showed that at 0.2 mM Mg^{2+} and above, the nucleosomes are folded into a maximally compacted fiber with a diameter of 30 nm. The presence of short regular regions containing diagonal striations in an otherwise irregular fiber - was seen as evidence for a helical packaging of nucleosomes. Thus a solenoid model was proposed that consisted of 6 nucleosomes per turn with a pitch of 110 Å, and linker DNA was envisioned to continue the superhelical trajectory of inside the nucleosomal core particle. This seemed to agree well with x-ray chromatin fiber diffraction experiments, which showed a 110 Å layer line oriented with the fiber (Pardon and Wilkins, 1972). Removal of H1 resulted in the inability to form a solenoid and instead it assumed a more open fiber (Finch and Klug, 1976).

The 30 nm fiber structure became the main pursuit of EM and numerous papers appeared describing chromatin conformations under different preparative conditions (Adkins et al., 2004). The systematic study of Thoma *et al.* (1979) served to resolve the determinants of chromatin folding. In low ionic strength medium, chromatin appears as relaxed filaments

with linker DNA forming a zigzag shape between successive nucleosomes. On increase in salt, the nucleosomes are seen to approach each other more closely as the fiber condenses and the zigzags close up. The 30 nm maximally compacted state is reached at either 60 mM Na^+ or 0.3 mM Mg^{2+}. In H1-containing chromatin linker DNA is seen to enter and exit the nucleosome at the same site, whereas in H1-stripped chromatin the two sites are on opposite sides, removing the zigzag appearance. H1-stipped chromatin is also less capable of compaction and requires higher ionic strengths. Even then, the fiber formed contains no distinct regular 30 nm structure, but instead irregular clumps.

1.3.3 *30 nm Fiber Models*

In the years following, numerous models were proposed for the 30 nm chromatin fiber, with the two major contenders being the solenoid and zigzag models (figure 1.9). Whereas the solenoid is a one-start helix with bent linker DNA, the zigzag models are two-start helices with straight linker DNA which connects separate stacks of nucleosomes. How these stacks are helically arranged divides the zigzag class into the helical ribbon model (Worcel et al., 1981; Woodcock et al., 1984) and the crossed-linker model (Williams et al., 1986). In the crossed-linker model, adjacent stacks of nucleosomes are twisted about the fiber axis which the straight linker DNA intersects, placing successive nucleosomes on opposite sides of the fiber (figure 1.9C). The helical ribbon instead supercoils the stacks around the fiber axis without intersecting it, and consecutive nucleosomes are separated vertically with the straight linker DNA now being parallel to the fiber axis (figure 1.9B). These two models are not mutually exclusive, but instead represent purely twisted or supercoiled versions of zigzag connected nucleosome-stacks. As twist and writhe can be topologically interconverted, so the zigzag fiber could exist in intermediate geometries (Richmond 2005).
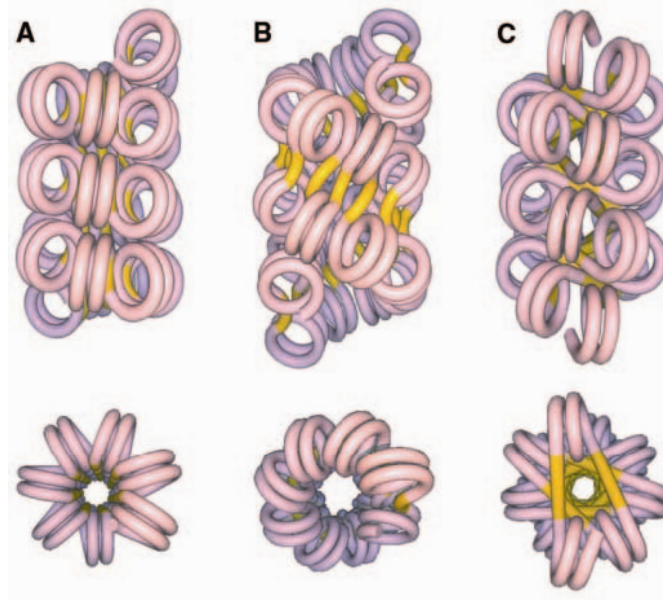
Figure 2.9 Models for the 30 nm chromatin fiber. **(A)** one-start solenoid (Finch and Klug, 1976). **(B)** two-start helical ribbon (Woodcock et al., 1984). **(C)** two-start crossed linker (Williams et al., 1986). Figure reproduced from Dorigo et al. (2004).

1.3.4 *Reevaluation of the 30 nm Fiber*

The solenoid model is still the most widely accepted model (Finch & Klug 1976), but has not gone unchallenged and the old evidence has been extensively reviewed (Woodcock and Horowitz, 1995; Zlatanova et al., 1999; Adkins et al., 2004). New evidence in the form of cryo-EM and other physical methods are clearly in support of a zigzag model (Zlatanova and Leuba, 2003). In fact there is no new direct structural evidence in support of solenoidal structures using these techniques.

A retrospective look at the original EM images shows that the majority of fibers are irregular, except for rare regular striated sections (Woodcock et al., 1976; Olins and Olins, 1974; Finch and Klug, 1976; Oudet et al., 1975; Thoma et al., 1979). It seems the predominant feature of chromatin in these images is the actual variability of the fibers. In fact, if the fiber was a regular helix, it would have been possible to produce a 3D reconstruction like was done for the T4 bacteriophage tail assembly (Amos and Klug, 1975). These studies (Finch and Klug, 1976) do not show unequivocal evidence of a solenoid as the resolution is unable to distinguish individual nucleosomes from within the

compact 30 nm state – they are only recognizable in looser structures where they exist in a zigzag arrangement. This is a problem facing all imaging techniques aimed at solution of the 30 nm fiber structure.

Early EM work had an added disadvantage: For samples to survive the high vacuum environment of the electron microscope, they were usually fixed, dehydrated and absorbed onto or embedded into a supporting substrate and stained. Any of these procedures can lead to preparation artifacts producing distortion of the native hydrated structure (Woodcock and Horowitz, 1997). The surface tension forces experienced during drying can stretch and shear, and absorption onto the viewing substrate flattens the fibers. In fact, these artifacts are used as an explanation for the sparse appearance of "helical" sections, in that the native solenoidal structure has been disrupted and the zigzag structure is seen as a consequence of fiber flattening (Thoma et al., 1979).

The solenoid "mindset" comes from the observation of nucleosomes being a basic repeating unit, can be conveniently packaged into a helix like most biological fibers made of a monomer repeat. However, nucleosomes are not identical, due to the contained DNA sequence. The linker length between adjacent NCPs is variable and so are the entry/exit angles of linker DNA off them. Furthermore, the variable linker length results in the relative rotation of adjacent NCPs, due to the twist of DNA - these features are not easily reconcilable with a solenoid model. Conversely, the zigzag model actually uses these features as the primary determinants of the variable spatial arrangement of nucleosomes in formation of the chromatin fiber (Woodcock et al., 1993). It also predicts the formation of regular helical structures when the underlying sequence is repetitive - such as with satellite DNA.

These observations and controversy led to the need for an improvement of imaging techniques for the viewing of isolated chromatin, *in situ* chromatin and reconstituted nucleosomal arrays.

1.3.5 *Linker DNA Bending*

Linker DNA bending is a major discriminating feature between the solenoid and zigzag models. In the solenoid model, the linker DNA is forced to bend in order to maintain the superhelical curvature initiated by the NCP and to continue into the adjacent NCP. The linker histone was thought to aid this, through its interaction with linker DNA, proving its essential role in 30 nm fiber formation (Thoma et al., 1979). However, more modern evidence of a linker histone-induced stem structure on the nucleosome nullifies this hypothesis (Hamiche et al., 1996). This result is more consistent with a zigzag model in which entry/exit DNA is poised by H1 to enable straight linker DNA to cross the fiber to an opposite nucleosome.

The high resolution x-ray structure of the DNA in the NCP (Richmond and Davey, 2003) shows the 9bp terminal extensions are straight, and cryo-EM images of mono-nucleosomes show that DNA bends away from the octamer surface as it leaves the NCP (Furrer et al., 1995). Both demonstrate that continuation of a superhelical curvature is unlikely.
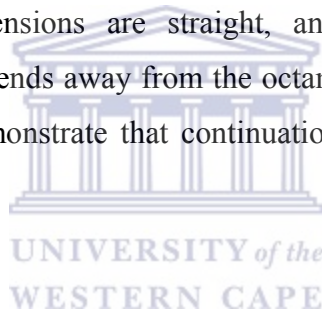
The same research group who proposed the solenoid model, observed straight linkers and a zigzag motif in low ionic strength (Thoma et al., 1979). Consequently through an increase in salt, linker DNA bending must then cause an unlikely and unobservable transition for formation of a solenoid structure. And during the compaction process, nucleosome-nucleosome distance is expected to increase; however a systematic cryo-EM study witnesses no change over a 0-40 mM NaCl range (Bednar et al., 1995). This mechanism of chromatin compaction is expected to cause a decrease in average fiber diameter, and again no changes through the salt range are observed by cryo-EM. The bending of DNA requires charge neutralization applied only to one side of the helix (as is achieved in the NCP itself), and in absence of an obvious protein complement throughout the full length of linker DNA this seems unlikely. Anyway, a protein-induced ionic bending mechanism would be expected to be maximal at a low ionic strength. Here the linker histone itself provides evidence for the opposite effect. A bioinformatics model

built for the chromatosome particle requires straight linker DNA (Bharath et al., 2003) - where the C-terminal tail mediates the close approach of the two adjacent linker DNA strands. This incorporates the evidence of 'lollipop' type structures observed via EM of chromatosomes (Hamiche et al., 1996).

Mechanistically, chromatin compaction in the zigzag model occurs through an entirely different mechanism and is mediated through the change in entry/exit angles of linker DNA. In turn ionic strength modulates the charge repulsion between the two linker strands. Of course, there is a degree of linker DNA bending for this effect to occur, but is localized close to the surface of the histone octamer and the majority is in an extended conformation. Together with the variability in linker length, these parameters are used in computer simulations to create a fiber which resembles the variable structures seen in so many EM micrographs (Woodcock et al., 1993). A Monte Carlo simulation also based on the zigzag model, generated fibers with bulk properties which agree well with experimental data and fold into a compact 30 nm fiber (Wedemann and Langowski, 2002). What is more, the fiber resembles the appearance of a solenoid with nucleosomes packed into an imperfect helix – even though this is achieved with straight linker DNA and next-nearest neighbor packing of nucleosomes.

Due to the differences in the mode of compaction of the two models, di- and tri-nucleosomes are expected to behave differently. The solenoid model predicts the sedimentation velocity of dinucleosomes should increase with increasing strength. The zigzag model, however, predicts compaction of only of tri-nucleosomes and above. Unfortunately, numerous labs report contradicting results and this issue remains controversial (Butler and Thomas, 1998; Butler and Thomas, 1980; Yao et al., 1991; Furrer et al., 1995). It is important to note that direct visualization of dinucleosome center-to-center distance via both cryo-EM and SFM has shown it to be constant with increasing salt (Bustamante et al., 1997; Bednar et al., 1995).

1.3.6 *Strong Evidence for a 2-start Crossed-Linker Model*

Two more recent experiments by the Richmond group provide strong evidence for a crossed-linker model for the chromatin fiber. Using the information of the H4-charge patch interaction derived form the NCP structure (Luger et al., 1997a), dual cysteine-replacement mutants were designed that might stabilize adjacent nucleosomes in a compact *in vitro* assembled nucleosome array (Dorigo et al., 2004). One such resultant mutant H4-V21C/H2A-E64C was successful in forming disulphide crosslinks, and once formed prevented the unfolding of compact arrays. Subsequent cleavage of the linker DNA always results in a maximum species containing half the number of nucleosomes as the full array. This demonstrates that compact arrays have a two-start helical geometry - as on the contrary a one-start solenoid would be expected to crosslink all nucleosomes, as they are arranged in a linear topology. Subsequent EM confirmed that crosslinked arrays were indeed 2-start, and appeared as two parallel stacks of nucleosomes linked by DNA. Each stack can be interpreted as nucleosomes conjoined by their histone octamer faces. Although this experiment clearly demonstrates the two-startedness of nucleosome fibers, it does not distinguish between a two-start helical-ribbon and a two-start crossed-linker model (figure 1.9BC).
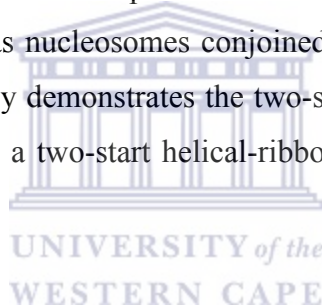
This ambiguity is resolved by the 9 Å crystal structure of a compacted tetranucleosome (Schalch et al., 2005). The tetranucleosome is assembled from recombinant octamers with a DNA segment containing four repeats of the 'Widom 601' strong positioning sequence (Lowary and Widom, 1998), and crystallized under conditions which result in maximum array compaction. The structure represents a truncated left-handed two-start helix with two stacks of two nucleosomes connected by straight stretches of linker DNA. (figure 1.9C). Adjacent nucleosomes within a single stack are seen to stack through a pseudo 2-fold contact between the H2A-α2, H2B-α3, and H2B-αC helices - which make up the periphery of the octamer face. Exact interactions cannot be deduced at this resolution, but density indicates the H2A-H2B dimer could be shifted out the nucleosome by up to 5 Å. Building a fiber model extrapolating this interfacial angle formed to a repeating unit of the tetranucleosome results in severe clashes, which are relieved in an

'idealized model' by shifting the internucleosomal interface to favor the H4-charge patch interaction instead and then using a dinucleosome as the repeat unit. However, it is conceivable that the almost perpendicular angle formed between the two adjacent stacks in the tetranucleosome is an artifact of crystal packing, rather than the interfacial angles – due to DNA being elastic. Still, both direct and idealized models provide good evidence that the 30 nm chromatin fiber corresponds to a left-handed crossed-linker model (Williams et al., 1986).

CHAPTER 2: MATERIALS AND METHODS

2.1 Protein Purification

All procedures were carried out at 4°C unless otherwise stated. Chemicals used were of molecular biology grade and obtained from either Sigma or Saarchem. Methods which are not elaborated upon were performed by standard molecular biology protocols (Ausubel et al., 2002). SDS-PAGE gels were run according to the discontinuous method introduced by Laemmli (Laemmli, 1970). Protein was concentrated via ultrafiltration using a 50 ml Amicon stirred cell (Millipore), fitted with a YM-10 membrane and operated under a nitrogen pressure of 40 psi.

2.1.1 *Isolation of chicken erythrocyte nuclei*

500 ml fresh chicken blood was collected into a beaker containing 70 ml of ACD buffer (16 mM citrate, 89 mM sodium citrate, 16 mM NaH_2PO_4, and 129 mM glucose). The blood was then filtered through 3 layers of cheesecloth into a 500ml Schott bottle and immediately placed on ice. 8 × JA20 centrifuge tubes were filled with blood and cells were pelleted by centrifugation at 1,000 g for 10 minutes and subsequently washed 3 times with 1 × SSC (10 mM sodium citrate, 150 mM NaCl). Each time the supernatant was aspirated together with the top white layer of leukocytes. Erythrocytes were then washed in a similar manner, but now using SSC containing 0.1% Triton X-100 This is repeated until only cream-colored nuclei remained in the pellet and the supernatant is clear of hemoglobin. Nuclei were further washed in buffer A (15 mM Tris-Cl pH 7.5, 65 mM NaCl, 60 mM KC1, 0.15mM spermine, 0.5 mM spermidine, 0.2 mM EDTA, 0.2 mM EGTA, 5 mM ß- mercaptoethanol, 0.2 mM PMSF) and finally stored at -20°C in buffer A containing 50% glycerol.

2.1.2 *Trial MNase*

Stored nuclei were washed 3 times in buffer A to remove the glycerol. The OD (optical density) of the nuclei suspension was then measured at 260 nm in 0.1 M NaOH and the concentration of nuclei adjusted to 100 OD/ml (~ 10 mg/ml chromatin) in buffer A.

A trial MNase (Micrococcal nuclease) digest was performed to determine the ideal time of digestion to produce chromatin of a desired length. A 1 ml aliquot of nuclei was adjusted to 1 mM $CaCl_2$ and equilibrated at 37°C for 3 minutes. 40 Worthington Units of MNase (Sigma) was added and 200 µl aliquots removed at the following time points: 5, 10, 15, and 20 minutes. These aliquots were pipetted into eppendorf tubes containing 1 µl of 500 mM EDTA for a final concentration of 2.5 mM and placed on ice to stop the reaction. Samples were spun in a desktop Eppendorf centrifuge at 5,000 × g for 10 minutes. Supernatant was poured off and pellets resuspended in 1 ml chromatin buffer (10 mM Tris-Cl pH 7.4, 0.2 mM PMSF, 0.2 mM EDTA). Samples were now shaken gently for 30 minutes to release the soluble chromatin (setting 1 on a Vortex Genie). Samples were centrifuged at 5,000 × g for 10 minutes and 500 µl aliquots removed to clean eppendorf tubes. SDS was added to a concentration of 0.1%, Proteinase K to 50 µg/ml and digested at 37°C in order to release the histones. Protein was subsequently removed by double phenol extraction, a single chloroform extraction and ethanol precipitation by standard procedures. Samples were then dried in a rotary evaporator and resuspended in 15 µl of 1 × TE (10 mM Tris-C1 pH 7.5, 0.2 mM EDTA). Samples were analyzed by electrophoresis in an ethidium bromide-stained (0.5 µg/ml) 1% (w/v) agarose gel in 1 × TAE (40 mM Tris-acetate, 2 mM EDTA) at 120 V for 40 minutes.

2.1.3 *Bulk Digest*

The nuclei suspension (100 OD/ml) was adjusted to 1 mM $CaCl_2$ and equilibrated at 37°C for 10 min. MNase was added to 40 U/ml and allowed to digest for 7 minutes (as determined from the trail MNase results), after which EDTA was added to a concentration of 5 mM and the reaction was placed on ice. Sample was placed in a JA-20 centrifuge tube and centrifuged at 5,000 × g for 10 min. The supernatant was discarded and pellet resuspended in half the original volume with chromatin buffer. This solution was gently shaken a Vortex Genie on setting 1 for 30 minutes. The sample was then centrifuged at 500 × g for 10 minutes, after which the soluble chromatin-containing supernatant was kept and volume determined using a sterile 25 ml pipette.

2.1.4 *Hydroxyapatite Chromatography*

This procedure was done according to the original method (Simon and Felsenfeld, 1979) with some modifications (von Holt et al., 1989). Soluble chromatin (~100 mg) was loaded onto a 4.5 × 10 cm hydroxyapatite column (BioRad Bio-gel HTP) pre-equilibrated with 3 column volumes of equilibration buffer (10 mM $NaPO_4$ pH 7.4, 0.2 mM PMSF), and unbound material was allowed to wash out with a further column volume. Histones were selectively eluted with a step gradient of increasing salt concentrations and manually collected fractions. Linker histones were eluted using equilibration buffers containing 0.65 M and 0.8 M NaCl, followed by the core histones using 3 M NaCl. Finally, the column was regenerated by washing with 0.5 M KPO_4 (pH 6.4) to remove bound DNA. The length of the DNA can be gauged by electrophoresis of these fractions on a 1 % agarose gel in 1 × TAE. SDS-PAGE of the other fractions are used to judge their histone composition and resultant core histones fractions were pooled and concentrated via ultrafiltration to 7 OD_{230nm}/ml (~2 mg/ml). Core histones can be stored at -20°C at this point by adding glycerol to 50%.

2.1.5 *Octamer Reconstitution and Purification*

This method was performed essentially as previously described (von Holt et al., 1989). Free core histones were assembled into complete histone octamers by overnight dialysis in high salt reconstitution buffer (20 mM Tris-Cl pH 7.4, 2 M NaCl, 5 mM ß-Mercaptoethanol, 0.2 mM EDTA, 0.2 mM EDTA). Subsequent separation of assembled octamers and excess free H2A-H2B heterodimers is done via the method of gel-filtration chromatography. A column packed with sephadex G-100 (Pharmacia) with a bed volume of 120 ml was pre-equilibrated with 10 column volumes of reconstitution buffer. The entire sample was then loaded onto the column and fractions were collected using a Gilson low-pressure chromatography workstation while recording a continuous absorption trace at 230 nm. Selected fractions of peaks were further analyzed on a 15 % SDS-PAGE gel to confirm their histone content, and appropriate histone octamer fractions were pooled and concentrated to 7 OD_{230nm} /ml (~2 mg/ml).

2.1.4 *Octamer Crystallization*

Histone octamer tubes were prepared essentially as by the previous methodology (Klug et al., 1980; Greyling, 1987). Purified histone octamers at ~2 mg/ml are slowly brought to 20 % ammonium sulphate by dialysis for 16 hours at room temperature against pre-crystallization buffer (20 % Ammonium Sulphate, 100 mM Tris-Cl pH 7.4, 2 M NaCl). The sample was then centrifuged at 2000 × g for 2 minutes to remove protein aggregates. The supernatant was then double-dialyzed at 4° C again with pre-crystallization buffer, but now with the slow complete exchange with crystallization buffer (45 % Ammonium Sulphate, 100 mM Tris-Cl pH 7.4, 2 M NaCl) over the course of two weeks via a peristaltic pump. The formation of histone octamer tubes could be judged by the appearance of a furry precipitate.

2.2 Electron Microscopy

2.2.1 *Negative Staining*

Sample was negatively stained by adding 5 µl of the precipitate to 5 µl of a 0.4 % uranyl acetate solution on top of a sheet of Parafilm. A freshly glow-discharged carbon grid was then floated on the 10 µl drop for a minute, before being removed and blotted with filter paper. The grid was then floated on a 10 µl drop of 0.2 % uranyl acetate for another minute and blotted again. Finally, the grid was again floated on 0.2 % uranyl acetate and blotted dry. Grids were placed in an oven at 50 °C for at least 30 minutes for complete drying before viewing in the microscope.

2.2.2 *Data Collection*

All electron micrographs were recorded at 50 000 × magnification using a Leo 912 TEM operating at 120 kV, with an in-column omega energy filter set for zero-loss imaging. A minimum-dose focusing (MDF) procedure (Williams and Fisher, 1970) was used in a semi-automated manner using the EM-Menu programme (Tietz, 1986) to ensure minimal specimen beam damage. In focusing mode a real-time computed power spectrum allowed for focusing and astigmatism correction. Two separate data sets of 100 and 50 micrographs each were captured using either a TVIPS SlowScan 2048 × 2048 format

CCD camera or Kodak SO-163 film respectively. The CCD dataset was captured into 16 bit MRC format with a pixel size of 14 μm corresponding to sampling of 2.8 Å at the specimen level. To ensure a grey-level histogram which optimally covered the available 16 bits and was within the linear dynamic range of the CCD, an electron dose of 1000 e⁻/nm² was needed. Film negatives were developed for 12 minutes at 20°C using full strength Kodak D-19 developer and digitized on an Ilford Leafscan 45 at 10 μm/pixel which equals a sampling of 2 Å/pixel. The electron microscope was previously calibrated to produce developed SO-163 negatives with an optical density of 1.0, and corresponded to a dose rate of 100 e⁻/nm² at the specimen level - therefore this electron exposure places the film within its linear dynamic range. Similarly, the scanning times of the negatives into 16 bit images was optimized for maximal grey levels. The film dataset also captures a much larger viewing area than the CCD camera - typically an area of 9000 × 5000 pixels.

2.2.3 *Helical Boxing*

Individual helices were manually selected and segmented into boxes (figure 2.1) with the use of Boxer, a programme included in the EMAN suite (Ludtke et al., 1999). The CCD data set was picked using a box size of 192 × 192 pixels and an overlap between boxes of 10 pixels (too low in retrospect). Boxed segments are rotated in-plane within Boxer to follow the helical axis, based on the angle between the user-selected beginning and end points of the helix. This resulted in a dataset of 697 boxed helices, which due to its small size and higher dose rate, was used for a preliminary 3D reconstruction. The film data was picked using a box size of 256 × 256 pixels with an overlap of 230 pixels, and led to an image stack of 3549 boxes. The box size was chosen as a compromise between including enough the helical symmetry in a single view while avoiding the inclusion of excess helical curvature. The large overlap was chosen as the box size minus the expected rise per subunit (Δz) of the helix. This was to avoid the identical alignment of consecutive boxes within the image stack – i.e. each consecutive box should therefore represent a view of a new rise of the helix and should align to a different reference.
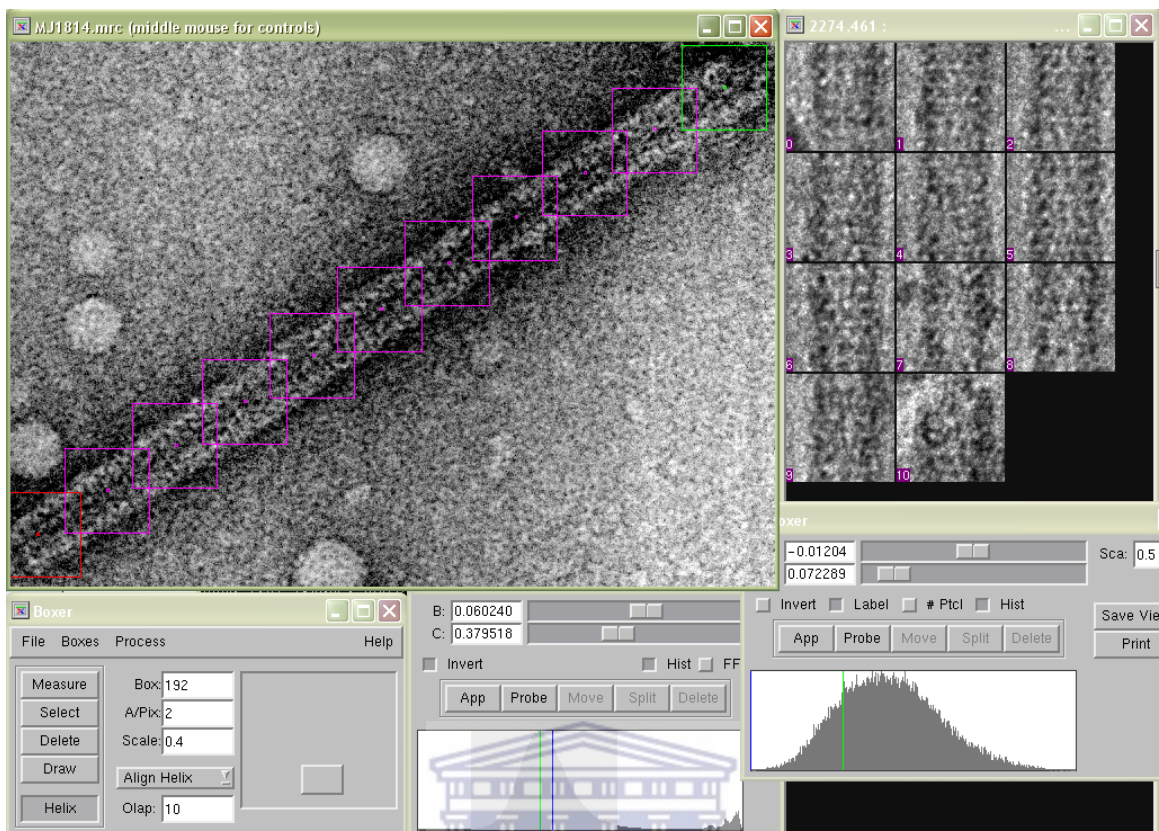
Figure 2.1 Segmentation of helices into boxes using the EMAN program Boxer (Ludtke et al., 1999), with the CCD dataset as an example.

2.2.4 *Preprocessing*

All subsequent processing operations were carried out on the image stacks within the image processing package SPIDER (Frank et al., 1996) unless otherwise mentioned. SPIDER was run under the Linux distribution Fedora Core 4, on a dual 3.2 GHz Intel Xeon server with 4 GB of RAM.

Binning – Both datasets were down-interpolated by a factor of 2, to reduce the sampling to 5.6 Å/pixel and 4 Å/pixel for the CCD and film image stacks respectively, and serves to speed up subsequent data processing. For the film data this was acceptable as the new sampling still represents an oversampling factor of 3.75 for a highest expected frequency of 15 Å – far above the limit set by the Shannon sampling criteria (Shannon, 1949). Oversampling is necessary though to compensate for the low-pass filtration effect that interpolation causes during future alignment steps.

Filtering – Images were band-pass filtered in Fourier space by the multiplication with a Gaussian-edged filter function at 300 Å and 15 Å. The high-pass filter was calculated to exclude low frequency terms such as stain variations and density ramps, not arising from the structure as the helix diameter was measured to be 300 Å. The low-pass filter is designed to exclude noise beyond the expected resolution of the structure, and improve alignments. The low-pass filter was eventually omitted for the film images when it was figured not to affect the image alignments.

Normalization – Image grey-level histograms were adjusted to a mean of zero and a standard deviation of one. As each image essentially represents a projection of the same structure, their mean density should be set equal. Correlation-based alignment also requires that the experimental and reference images have a mean of zero, for a least-squares approximation.

Pre-centering – Images were centered in the x-dimension via a reference-free approach (Penczek et al., 1992), wherein images are correlated with a moving global average to obtain optimal x- and y-shifts. The global average itself must then be centered via correlation with its mirror image - reflected through the y-axis. Thus the x-value needed to center the average must be added to the previous reference-free x-value, and the combined x-shift applied to center each image. If the x-shift calculated was greater than half the helix radius, then that image was deleted. In this way 49 film images were rejected leaving 3500 images in the stack. Pre-centering is important in reducing the search range needed later for multi-reference alignment and in producing more meaningful alignment statistics thereof.

Multivariate Statistical Analysis – MSA was performed on the pre-centered film image stack using IMAGIC (van Heel et al., 1996). This is done to identify any size variations in the helix diameter - reflected by the eigenimages which represent the principal components of variation in the data set.

2.3 3D Reconstruction

3D reconstruction was undertaken via the iterative helical real-space reconstruction (IHRSR) method (Egelman, 2000) in SPIDER (Frank et al., 1996). The IHRSR method (figure 2.1) represents the application of single particle techniques to the reconstruction of helical filaments, and supersedes the older Fourier-Bessel methods (Crowther and Klug, 1975). Briefly the steps involve: multi-reference alignment and angle assignment of images against projections of a 3D model; back projection of class averages into a 3D volume; helical symmetry search and imposition; and projection of a new set of references from the symmetrized volume. This process is iterated till no further improvement of the 3D volume is achieved as judged by the various parameters which measure convergence. The validity of the final reconstruction can finally be confirmed by Fourier-shell correlation, agreement of power spectra, and docking.
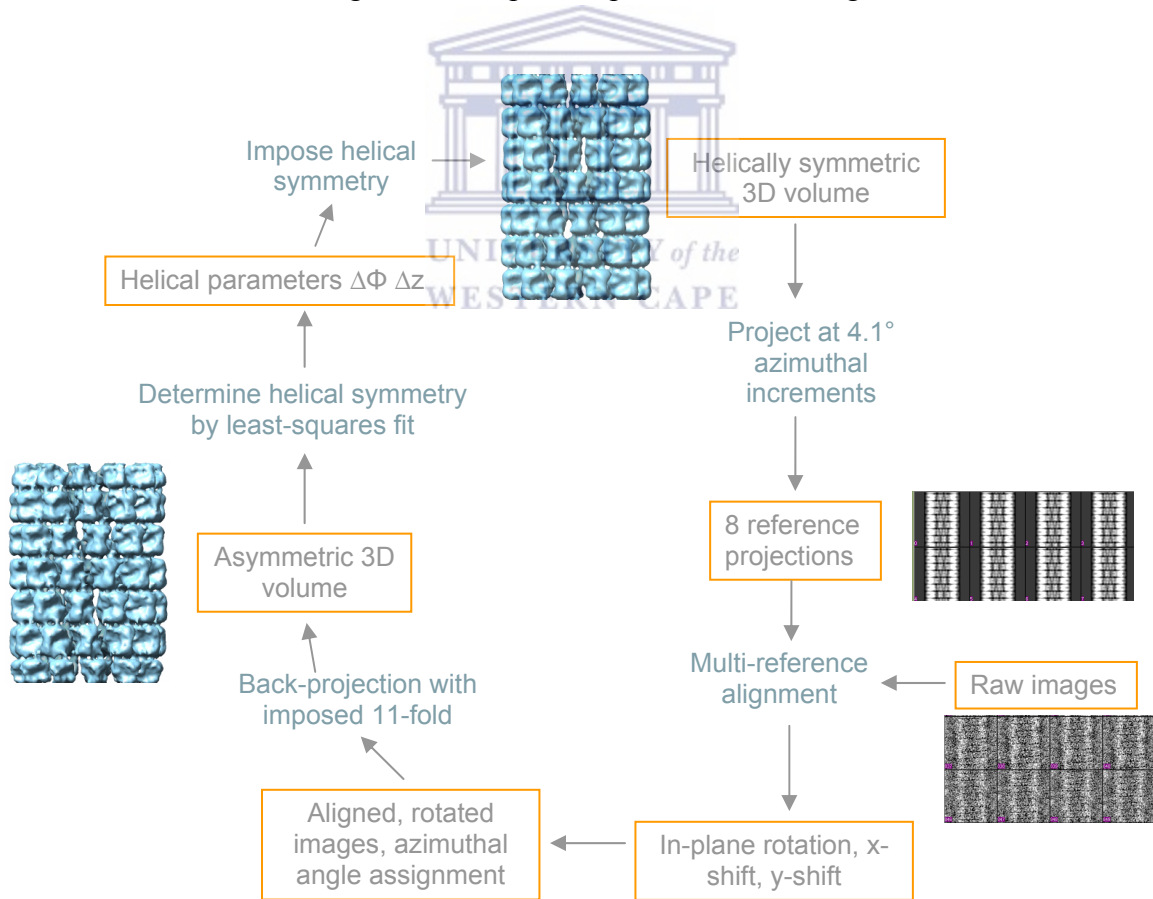

Figure 2.2 The IHRSR iterative refinement cycle demonstrated with the histone octamer tube as an example. Details are provided in text.

Start-Models – For the preliminary reconstruction using the CCD data, a start model was created using the crystal structure of the histone octamer (Arents et al., 1991) generated into a helix using the parameters of Klug et al. (1980), and low-pass filtered to 15 Å. For the film dataset, an *ab initio* start model of a featureless cylinder (created in SPIDER using 'mo 3' and low-pass filtered to 50 Å) was used to ensure no model bias.

Multi-Reference Alignment – Projections of a start model, or a back-projected volume during the iteration cycle, form the templates for multi-reference alignments. As helices are absorbed onto a carbon support, captured images thereof are constrained to be views perpendicular to the fiber axis. This sets the reconstruction geometry to be that of a single axis/tomographic tilt series - therefore projections of a model volume need only be about the azimuth. The number of projections needed is determined by the sampling of Fourier space and by the given point-group symmetry of the structure. Because of the reconstruction geometry, the overlap of central sections in Fourier space (and thus the achievable resolution) is only limited away from the fiber axis. The amount of overlap is dictated by the thickness of central sections which therefore follows a π proportionality as described by the equation (Crowther et al., 1970):

Number of templates = (π × Diameter of helix) / (desired resolution × symmetry)

Thus for the histone octamer helix, with a point group symmetry of 11 and a helix diameter of 300 Å, only 8 references were sufficient for a resolution of better than 15 Å. The high point group symmetry means views of the helix repeat every 360/11°, and the 8 projections were made at 4.1° increments to evenly cover this asymmetric wedge.

The produced references together with the raw images are then used by the SPIDER multi-reference alignment command 'AP NQ' (Joyeux and Penczek, 2002). Alignment is essentially a 4-dimensional search to bring each image into common register and separate them according to the projected view they represent. Each raw image is aligned to the references via cross-correlation to determine the x-shift, y-shift and in-plane rotation angle. The azimuthal angular assignment, of each raw image is then based on to which

reference overall it produces the highest correlation coefficient, and furthermore which 2D alignment parameters are henceforth applied (because the projection angles of the references are known). Images which represent the same view are averaged together to form 2D class averages with a better signal-to-noise ratio than the individual images. Images are rejected during alignment if their in-plane rotation deviated by more than 10° from either 0° or 180° (since helices were roughly oriented during boxing) - although once the iteration cycle had converged, rejections no longer occurred.

Back-projection – Back-projection is the reverse of projection, except that it occurs via interpolation in Fourier space, where there can be clearer derivation of filter functions. It is performed by the same geometry as the projections, thus by a single-axis tilt series with equally spaced angles. Because central-sections fan out from the origin in Fourier space, low frequency terms are overrepresented. An R-weighted filter combats this by individually filtering each central section by a factor inverse to the radius and number of central sections. 11-fold symmetry is imposed simply by replicating the central sections into symmetry-related positions. The various filtered central sections are described by cylindrical-polar coordinates which must be interpolated to the Cartesian grid of the 3D Fourier transform (Crowther et al., 1970). Inverse Fourier transformation of the thus-sampled 3D Fourier transform, results in a 3D volume in real-space.

Helical Search and Imposition – This is the most important part of the IHRSR cycle and takes advantage of the underlying helical symmetry to average all subunits in the backprojected 3D volume (Egelman, 2000). The helical symmetry can be completely described by a screw operation involving a coupled azimuthal rotation ($\Delta\Phi$) and axial translation (Δz) which describes the position of each sequential subunit. No assumption is made about the helical symmetry, and the volume is searched by a least-squares quadratic fit of helically-related voxels after conversion to cylindrical-polar coordinates. The process is iterated between the quadratic fit of $\Delta\Phi$ and Δz, because they are interdependent. A starting guess of $\Delta\Phi$ and Δz must be made on the first iteration round, but afterwards, the found symmetry from the previous round can simply be used as the starting point for the new search. During each iteration round, the found helical symmetry

is subsequently imposed on the back-projected volume to vastly improve the appearance and signal-to-noise ratio of the volume. The programs to perform the helical search and imposition were kindly provided by Prof. E. Egelman.

Angular and Helical Refinement – The helical search and imposition completes one iteration cycle and projections of the helically-averaged volume serve as new templates for multi-reference alignment. A progressive refinement of alignment and angular assignment occurs during the iteration cycling, as the projections of the reconstruction begin to resemble the raw images more closely (Penczek et al., 1994). Coupled to this, the helical symmetry estimate also improves and helical averaging increases the quality of produced projections, meaning that the IHRSR algorithm can converge quicker than regular single particle methods.

Visualization – 2D images were visualized by the EMAN command 'v2' (Ludtke et al., 1999), while 3D volumes were viewed within Chimera (Pettersen et al., 2004). The molecular density threshold for enclosing the 3D maps was calculated using an average protein density value, octamer molecular weight (109.1 kDa) and number of subunits in the helix. The volume expected used the accepted protein density value of 1.35 g/cm^3 (more conveniently expressed as 813 Å^3/kDa) (Quillin and Matthews, 2000) and with the number of octamer subunits in the final reconstruction judged to be 86.65 (box length / Δz * point group). The tail regions of the octamer which account for 28% of its molecular weight must be excluded from the calculation as they are expected to be disordered. Finally, dividing by the cubic sampling (4^3 voxel/Å^3), the final volume was thus calculated to enclose 86,464 voxels or a threshold value of 0.021 in Chimera. This value was confirmed later by comparison with the calculated van der Waals surface of the crystal structure of the histone octamer (Arents et al., 1991).

Convergence – The rate of convergence of the IHRSR process is dependant on the starting model is used and the initial guess of the helical symmetry – since this determines where the model starts relative to the radius of convergence. Convergence is ultimately achieved when the helical parameters of $\Delta \Phi$ and Δz stabilize and the average

cross-correlation score reaches a plateau – reflecting no further improvement. Any instability or lack of convergence in these parameters indicates a non-stable solution, and is therefore artefactual. In the reconstruction of the histone octamer, although convergence was generally achieved after less than 10 rounds, it was allowed to proceed for a full 50 rounds as a precaution and to gather statistics on the stability of the iteration parameters.

Validation – The alignment statistics are a good indicator of the validity of a reconstruction. Angular assignment of raw images to references should show a relatively flat distribution, because theoretically all views of a helix should be equally likely. Also due to the images having been pre-aligned in x, when aligning to the references, they should show a narrow histogram symmetric about zero. And if the reconstructed volume is indeed representative of the true structure, then the raw images and subsequent class-averages should agree visually with the projections. In a similar way, the averaged power spectrum from the images should also agree with the computed spectrum from a projection of the final volume, and that the helical symmetry indicated therein is consistent.

The reproducibility is confirmed by performing the entire IHRSR process twice using two independent start models which agree to less than the resolution of the reconstructed volumes. In this case the same *ab initio* start volume was used, but this is valid because it is filtered to 50 Å and the starting guess of the helical symmetry differed by 2° in ΔΦ. The resolution of the thus-generated two volumes can be calculated after alignment, by their comparison in Fourier space via a Fourier Shell Correlation (FSC) (van Heel and Harauz, 1986). To avoid a spurious boost to resolution that helical averaging would bring, the asymmetric volumes are compared instead. A conservative threshold of 0.5 was used to determine the FSC cutoff resolution.

Rationalization of the 3D volume provides a final validation. All density should be continuous and the connectivities which stabilize the helix visible – there should be no floating density. Dependant on the resolution of the EM map, known structural features

such as symmetry elements should be readily identifiable. Both can then be directly confirmed by the docking of an atomic structure if available.

2.4 Docking

The crystal structure of the histone octamer at 3.1 Å (Arents et al., 1991) (PDB code: 2HIO) was used for all docking procedures, because it lacks the disordered tail domains and also is the only octamer structure where the dyad axis is a true 2-fold crystallographic axis. Thus the two halves of the octamer are identical, and each half comprises an asymmetric unit.

Rigid-body Docking - was performed using a correlation-based search in the program Situs (Wriggers et al., 1999). A full 6-dimensional search was carried out with an angular increment of 15° with the x-ray density filtered to 16 Å (as determined from the FSC). The symmetry operators of $\Delta\Phi$ and Δz were then applied to the best fitting solution to generate the entire helix – implemented by a self-written Fortran 90 program.

2D Constrained Docking - The crystal structure (2HIO) was aligned so that both the dyad and superhelical axes were in the x-y plane (of the helix), and so a projection of the dyad along y-axis intersected the z-axis (figure 2.3). In this way the atomic coordinates could be easily rotated about the z-axis to form the C_{11} cyclic point group symmetry, leaving the only variables of the starting y-position - corresponding to the helical diameter, and a rotation about the dyad axis. The best values were found by empirically generating a variety of helices of different diameters and dyad rotations (written within Fortran 90) which were converted to electron density, filtered to 16 Å and cross-correlated to the EM map (within SPIDER). This process was automated by use of a Linux bash shell script to generate a 2D cross-correlation surface as a function of helical diameter and in-plane dyad axis rotation.
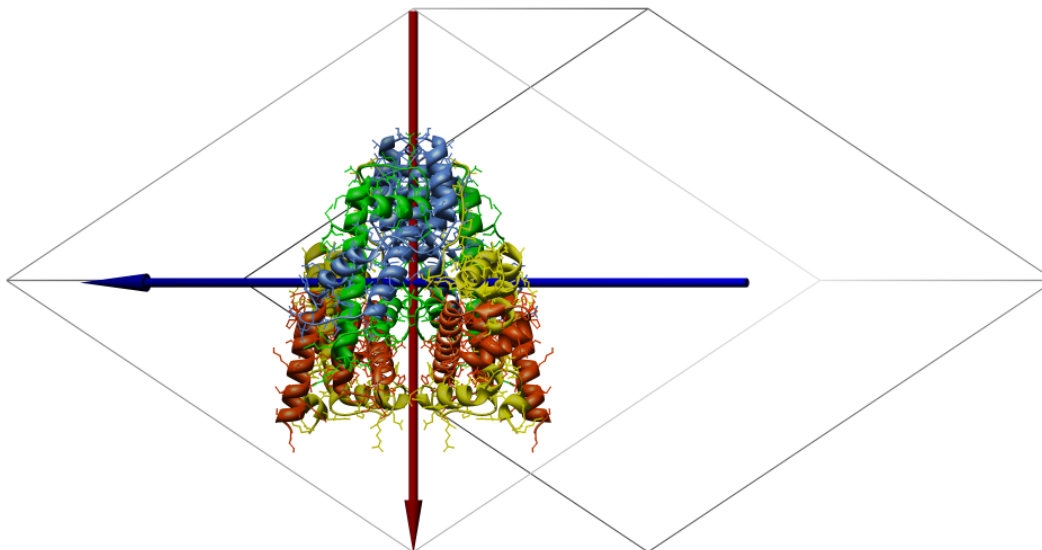
Figure 2.3 Histone octamer 2HIO unit cell orientated in preparation for constrained docking. The dyad axis is labeled by the red arrow, and relates each half particle via a crystallographic 2-fold. The superhelical axis is indicated by the blue arrow and occurs at 60° to the x-y plane of the unit cell.

Van der Waals Docking – Using the superhelical axis as an additional constraint, and limiting it to the x-y plane, the search is reduced to one dimension. Models of decreasing diameters were empirically generated by a Fortran programme and manually evaluated in Chimera (Pettersen et al., 2004) for steric clashes between neighboring octamers in their van der Waal's surfaces.
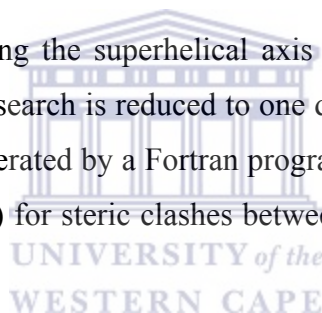
2.5 Interface Analysis

Electrostatics – Electrostatic potentials were calculated using the program APBS (Baker et al., 2001), which uses an adaptive finite element algorithm (Holst, 2001) to solve the linear Poisson-Boltzmann equation within the molecular graphics program PMV (Sanner, 1999). Potentials were calculated using a protein dielectric of 3, solvent dielectric of 78.5, ion-exclusion radius of 1.4 Å, and monovalent salt added to a concentration of 150 mM (Na^+ - 1.16 Å ion radius, Cl^- - 1.67 Å ion radius). Solvent accessible surfaces were calculated using the MSMS program (Sanner et al., 1996) within Chimera (Pettersen et al., 2004). Electrostatic surfaces were ultimately created by mapping the electrostatic potentials to the MSMS surface, and colored with a range of 7.0 (red) to +7.0 (blue) $kcal.mol^{-1}.e^{-1}$.

Protein-Protein Interactions – Interface surfaces were calculated using the InterSurf tool within Chimera (Pettersen et al., 2004), and colored by the attribute of separation distance. Interfaces were further analyzed by submitting the docked structure to the protein-protein interaction server (http://www.biochem.ucl.ac.uk/bsm/PP/server/). The interface surface is characterized as the region inaccessible to solvent (1.4 Å sphere). Details of the procedures used are further described in the literature (Jones and Thornton, 1995; Jones and Thornton, 1996).

Energy Minimization – Iterative energy minimization was carried out on the H2A-α3/αC interaction surfaces (N84 – K125), by the Polak-Ribiére conjugate gradient method (Polak and Ribiére, 1969) within the program HyperChem v7.5 (Froimowitz, 1993). Minimization was performed using the AMBER99 force field (Case et al., 2005) *in vacuo* with electrostatic and van der Waals scale factors of 0.5. Convergence was achieved after 1346 cycles. Newly formed hydrogen-bonds were subsequently identified with the FindHBond tool within Chimera (Pettersen et al., 2004).
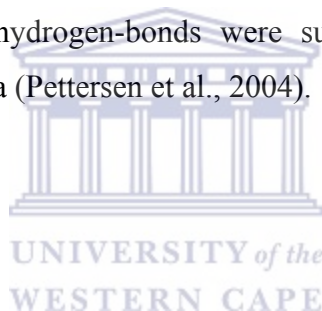
CHAPTER 3: RESULTS

3.1 Protein Purification

3.1.1 *Trial MNase*

Chicken nuclei were prepared as described in chapter 2, and a trial MNase digest carried out to determine the ideal time for producing chromatin of a suitable length. For hydroxyapatite (HA) chromatography, it was empirically determined that chromatin digested into predominantly mono- and dinucleosome lengths were best (data not shown). Over-digestion led to poor binding to the HA, presumably because the produced NCPs lack linker DNA – this resulted from using 400 U/ml of MNase as suggested by the established protocol (von Holt et al., 1989). Conversely, yields from long chromatin were low, because core histones are mechanically constrained by the DNA (40 U/ml MNase for 1 minute). Therefore, from the trial MNase time-course shown in figure 3.1 a digestion time of between 5 and 10 minutes was deemed suitable.
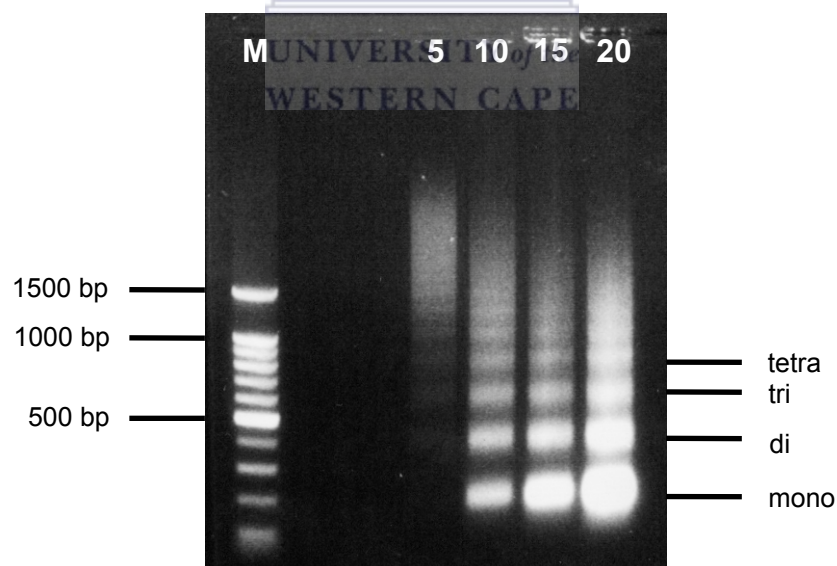

Figure 3.1 MNase time course digestion of chicken erythrocyte nuclei. Digestion times in minutes are indicated above the lanes and the nucleosome species produced to the side. The marker used was a 100 bp ladder (Promega).

3.1.2 *Bulk Digest*

Thus 14 ml of nuclei at 100 OD_{260nm}/ml were digested with 40 U/ml MNase for 7 minutes, and the soluble chromatin extracted. The length of this chromatin can be confirmed after hydroxyapatite chromatography without the need for laborious Proteinase K digestion and phenol extraction, as the proteins are separated from DNA by the procedure anyway. Figure 3.2C shows this DNA, and the average lengths of chromatin the bulk digest resulted in (which were slightly longer than expected from the trial MNase).

3.1.3 *Hydroxyapatite Chromatography*

HA chromatography works by immobilizing the DNA of chromatin, while the histone components are selectively dissociated by increasing salt concentrations. A total of 70 mg DNA/ml solubilized chromatin was loaded onto the column and fractions eluted with step gradients of salt (figure 3.2A). Linker histones were eluted first in the range 0.65 – 0.8 M NaCl, followed by the core histones at 3 M NaCl. Some H2A-H2B dimer depletion occurs at 0.8 M salt, but this is necessary to rid the sample of linker histone H5, which is seen to persist in later fractions (figure 3.2B). Finally, regenerating the column with 0.5 M potassium phosphate buffer liberated the bound DNA, whose length was visualized by agarose gel electrophoresis (figure 3.2C). The core histone fractions corresponding to peak 2 were pooled and concentrated to 2 mg/ml, and then dialyzed overnight against 2 M NaCl to reassemble complete octamers.
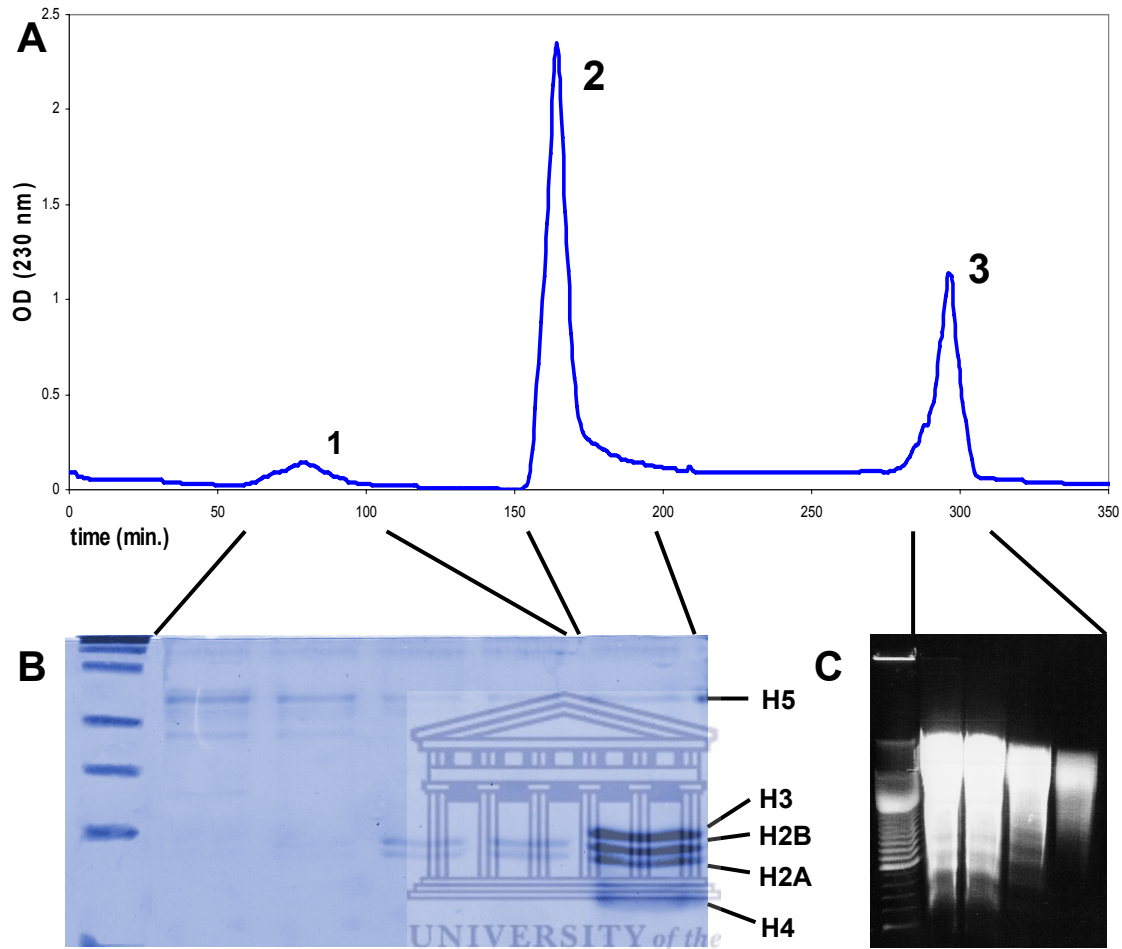
Figure 3.2 Isolation of histones via hydroxyapatite chromatography. **(A)** Chromatography trace was recorded at 230 nm. **(B)** SDS-PAGE gel of indicated fractions corresponding to peaks 1 and 2. Broad range marker (BioRad) was used. **(C)** 1% Agarose gel of fractions taken directly from peak 3 and stained with ethidium bromide.

3.1.4 *Sephadex G-100 chromatography*

The octamer is a 109 kDa complex, while H2A-H2B dimers are only 28 kDa, and thus the two species can be separated by size-exclusion chromatography. Additionally, this provides an extra purification step to remove any remaining linker histones. Sephadex G-100 has an exclusion limit of 100 kDa, so the octamer will exit in the void volume, while the dimer will enter the internal volume and be consequently retarded. Figure 3.3 shows the clear separation achieved between intact octamers and excess H2A-H2B dimers in peak 2 using this procedure.
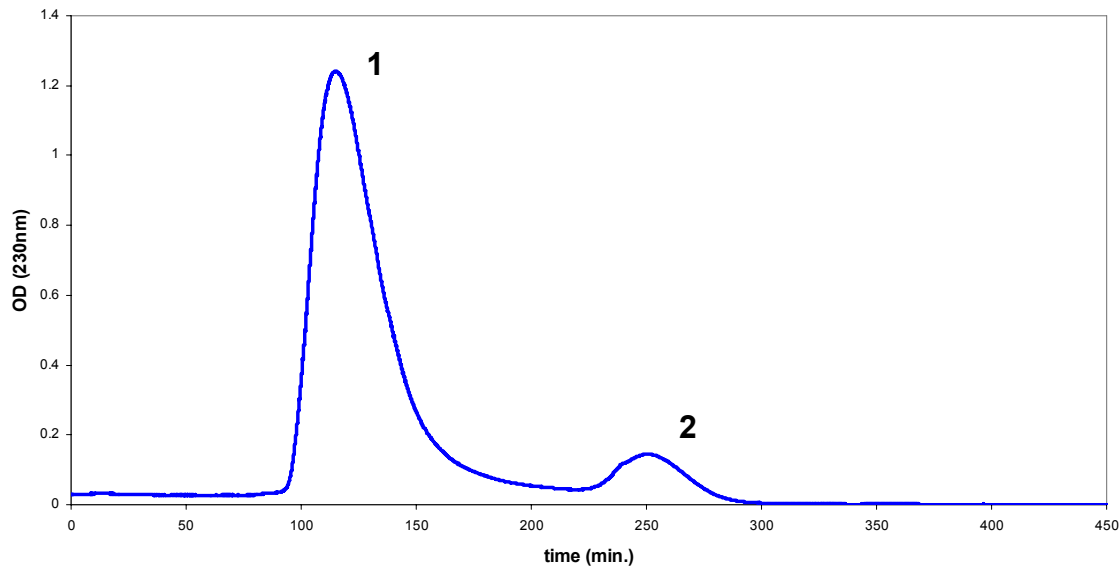
Figure 3.3 Separation of reconstituted histone octamers via Sephadex G-100 gel-filtration chromatography. Intact histone octamers are contained within peak 1, and are separated from free H2A-H2B dimers in peak 2 – as confirmed by SDS-PAGE analysis (data not shown).

3.1.5 *Octamer Crystallization*

Pure histone octamers were concentrated to 2 mg/ml, and dialyzed against 20 % ammonium sulphate. Any precipitates that formed were removed by centrifugation and this pre-crystallization procedure served to lower the protein concentration and remove any spontaneously formed protein aggregates. Final crystallization proceeded by slowly raising the dialysis buffer to 45 % ammonium sulphate. This procedure was performed slower than the 4 days reported (Greyling, 1987; Klug et al., 1980) as a precaution to produce extended helices. This was achieved by double dialysis and raising the dialysis buffer to the final concentration via buffer exchange over the course of two weeks. The appearance of a furry precipitate indicated that tubular crystals had finally been formed.
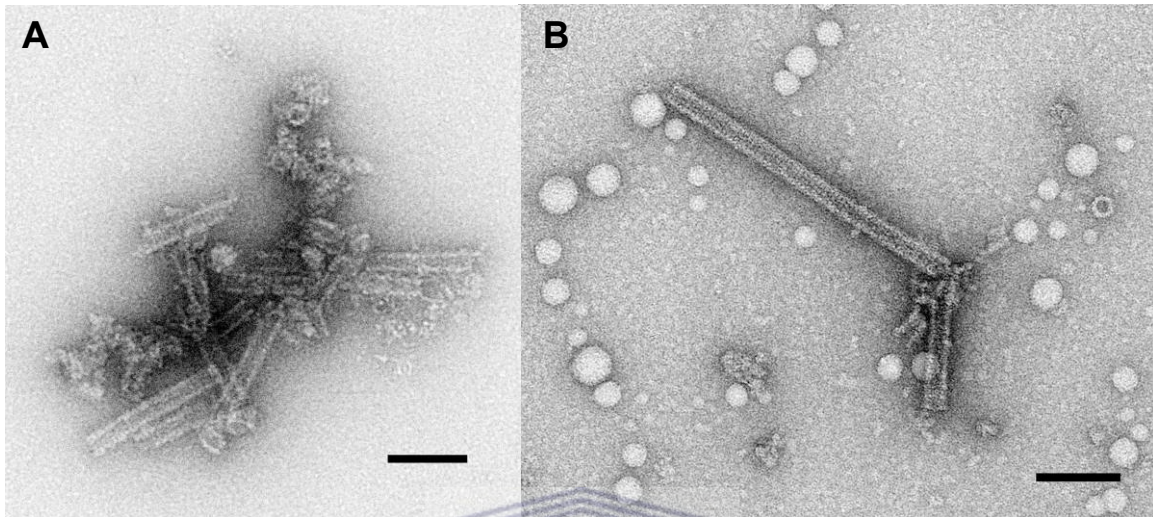
3.2 Electron Microscopy

3.2.1 *Negative Staining*

Figure 3.4 Negative staining of histone octamer precipitate. **(A)** Regular negative staining. **(B)** Modified negative staining by adding the stain to the precipitate before applying it to the grid. Magnification of 50k and scale bars represent 100 nm.

Figure 3.4A illustrates the results of regular negative staining whereby the sample was applied to the grid, prior to staining with 2% uranyl acetate (Harris and Scheffler, 2002). Under these conditions the 300 Å helices as described (Klug et al., 1980) were rarely observed, and the grid was instead dominated by ammonium sulphate crystals. When the octamer tubes did occur, however, they were aggregated together in patches of negative stain (figure 3.4A). To obtain helices more suitable for image analysis, a modification to the regular negative staining procedure was made as suggested by Klug et al. (1980) and detailed by Greyling (Greyling, 1987). The mixing of uranyl acetate and precipitate prior to grid application, led to a more suitable concentration, and less aggregation of helices (figure 3.4B). Uranyl acetate is a known fixative and allows helices to be stabilized (Zhao and Craig, 2003), while the ammonium sulphate is subsequently washed off. Thus, uranyl acetate serves to bridge the techniques of crystal growth and EM. However, excessive bubbling is seen to be a feature of this modified negative stain procedure, as is witnessed in figure 3.4B and also in the images of Klug et. al. (1980). This is suggested to occur from the boiling of trapped water under the beam and high vacuum of the TEM (Bozzola,

1998). This champagne artifact was therefore resolved by baking the grids at 50°C for at least 30 minutes, and the improved micrographs are seen in figure 3.5.

Originally it was conceived to attempt cryo-EM data collection on the helical tubes. However, this proved unsuccessful due to the presence of high salt concentrations, which impaired the vitrification procedure. Removing the salts, would necessitate first stabilizing the helices by glutaraldehyde cross-linking, which in turn, is impossible in the presence of ammonium sulphate. Thus the helices would need to be regrown in a buffer more suitable. Cryo-EM may have offered better resolution and fewer of the artifacts associated with negative staining, but the situation is still a good approximation for objects containing high symmetry, such as helices. This is because a single view of a helix offers many views of the individual subunits - the majority of which will be undistorted and allow restoration through averaging. Resultantly, artifacts generally apply more to the helix as a whole than the composing subunits.
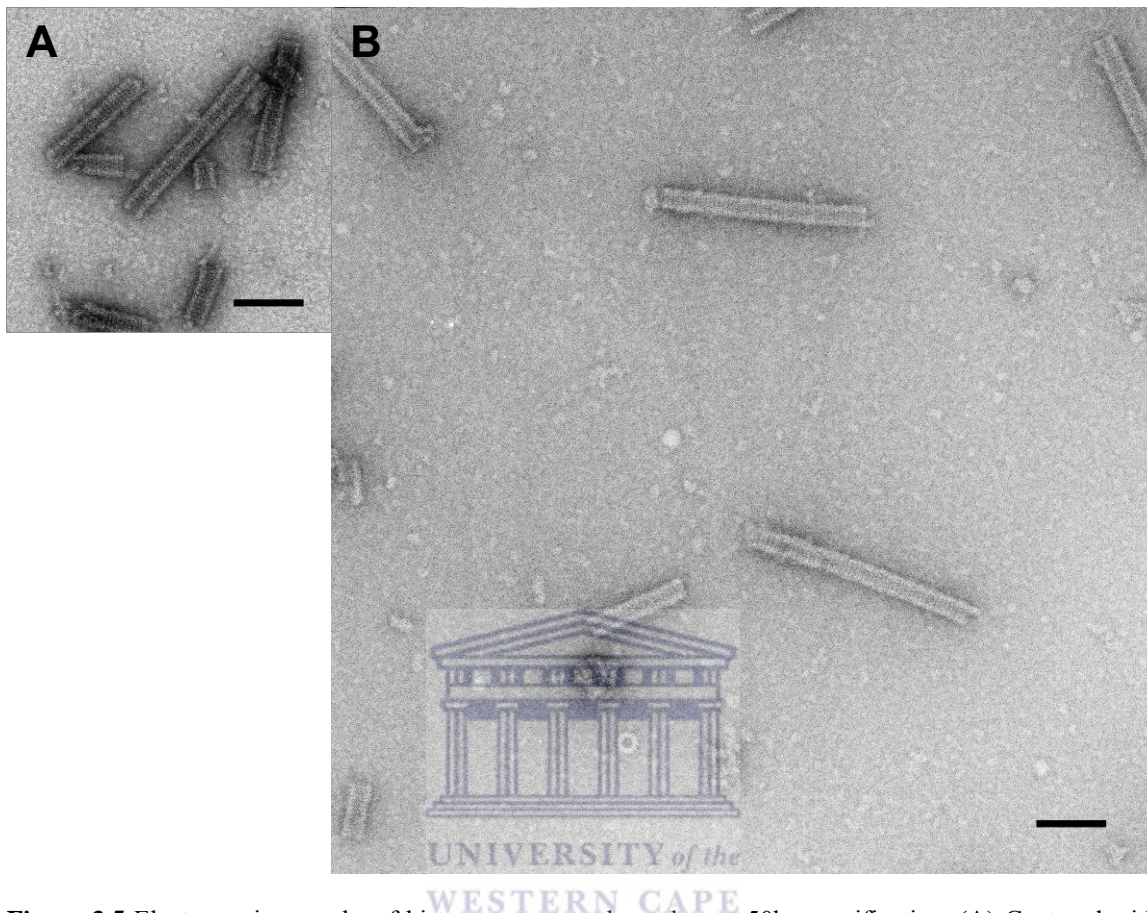
3.2.2 *Data Collection*

Figure 3.5 Electron micrographs of histone octamer tubes taken at 50k magnification. **(A)** Captured with the CCD camera. **(B)** Captured with film and scanned. Both micrographs are scaled to the same sampling to illustrate the difference in viewing area between the two formats. Scale bars represent 100 nm.

With the negative staining optimized, two datasets were subsequently collected using a minimum-dose procedure (Williams and Fisher, 1970). The first dataset of a 100 micrographs was collected using the CCD camera - of which an example micrograph is shown in figure 3.5A. An electron dose of 1000 e^-/nm^2 was needed to produce an optimal grey-level histogram and is still within acceptable limits for negative stain microscopy (Unwin and Henderson, 1975). The small field-of-view, however, makes the CCD camera unsuitable for collecting micrographs of helices, because a high concentration of helices is needed and often long helices are truncated (compared to the film format in figure 3.5B). Therefore, this dataset served primarily for initial image processing and 3D reconstruction, and instead data was recorded directly onto film negatives to form a

second dataset of 50 micrographs (figure 3.5B). Because of the increased sensitivity of film a lower dose rate of 100 e$^-$/nm^2 could be used, and meant that potentially more high resolution features were preserved from radiation damage within the structure. Although as a downside, the SNR was also visibly reduced when compared with the CCD data (figure 3.5B vs. A).

3.2.3 *Pre-Processing*

Helices were boxed into image stacks, and subsequently binned, filtered and normalized as described in section 2.2.4. Pre-centering in the x-dimension was performed via a reference-free approach (Penczek et al., 1992), and serves to lower the search range needed in the reference-based alignment step, and additionally makes the resultant alignment parameters more meaningful. Alignment in the x-dimension is aided by the negative staining meniscus which surrounds the helices, as it acts as an additional correlation penalty for misalignment. To avoid further image degradation, interpolation was avoided by only applying integer pixel x-shifts. The reference-free approach applied here is superior to similar model-based approaches, because no assumptions are made about the inner and outer helical diameters.

Visually the helices in the micrographs (figure 3.5) are indistinguishable from those reported by Klug et al. (Klug et al., 1980): They measure ~300 Å in diameter, and their lengths vary up to one micron; they are hollow, as indicated by the center of the helix being darker from the presence of stain (clearly seen in figure 3.5A); and obvious horizontal striations are a major visual feature of the helices and suggest they are composed of stacks of rings. Forming a 2D average is a quick way to boost the SNR of the data and reveal hidden features of the structure, and also gives an impression of the quality and potential resolution of the data prior to 3D reconstruction. Accordingly, calculation of a reference-free average from the entire CCD image stack (figure 3.6A) enhances these horizontal striations and can be measured to be 65 Å (after calibration of the CCD sampling). Figure 3.6A also reveals what appear to be individual subunits, which present an x-shaped front view (center of helix) and a donut-shaped side view (edges of helix). Subunits seem to be stacked vertically by a stagger (which would reflect

ΔΦ) which appears to repeat approximately every third ring. Of course, as the 2D average is formed without prior angular classification of the data, it reflects a rotationally averaged view.
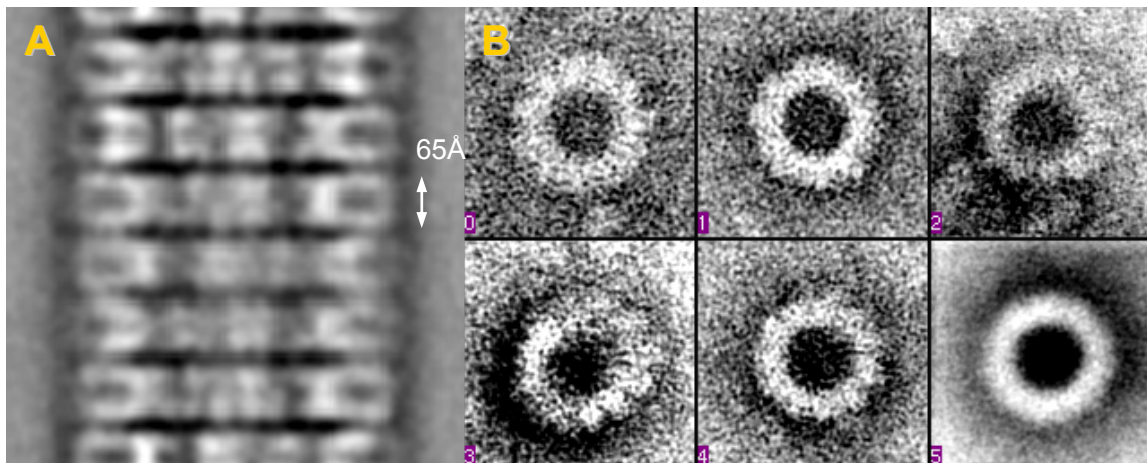

Figure 3.6 (A) Reference-free average calculated from 697 CCD images **(B)** Annular end-on views. Image 5 represents a reference-free average calculated from 23 such images.

Because of these periodicities displayed in the real-space 2D reference-free average, a power spectrum (amplitudes of a Fourier transform) would allow such features to be confirmed and more accurately measured. Such periodicities are also reinforced by using a long segment of helix for the power spectrum calculation. Figure 3.7D shows a power spectrum calculated for a single 1024×1024 pixel boxed helix, and it displays strong on-axis reflections at 65, 33, and 22 Å; and an off-axis reflection at 280 Å. On-meridional reflections arise from repeating structural motifs parallel to the fiber axis, which are unaffected by the rotational portion of the helical symmetry - such a basal repeat is the axial rise (Δz) per subunit of the helix and multiples thereof. Consequently, the largest horizontal striation measuring 65 Å corresponds to the rise of the octamer helix. This is further reinforced by the fact that an octamer measures 65 Å along its disc face (figure 1.5C) (Arents et al., 1991), and indicates that each ring is composed of octamers related by cyclic point group symmetry. Additionally, the 33 Å reflection being of second order to the 65 Å reflection and of equal intensity, further divides the density of each ring into two equal layers. This suggests the presence of a radial dyad axis, modifying the cyclic

point group to a dihedral point group. Some individual power spectra even show a further on-axis reflection at 17 Å, hinting at the potential resolution of the data.
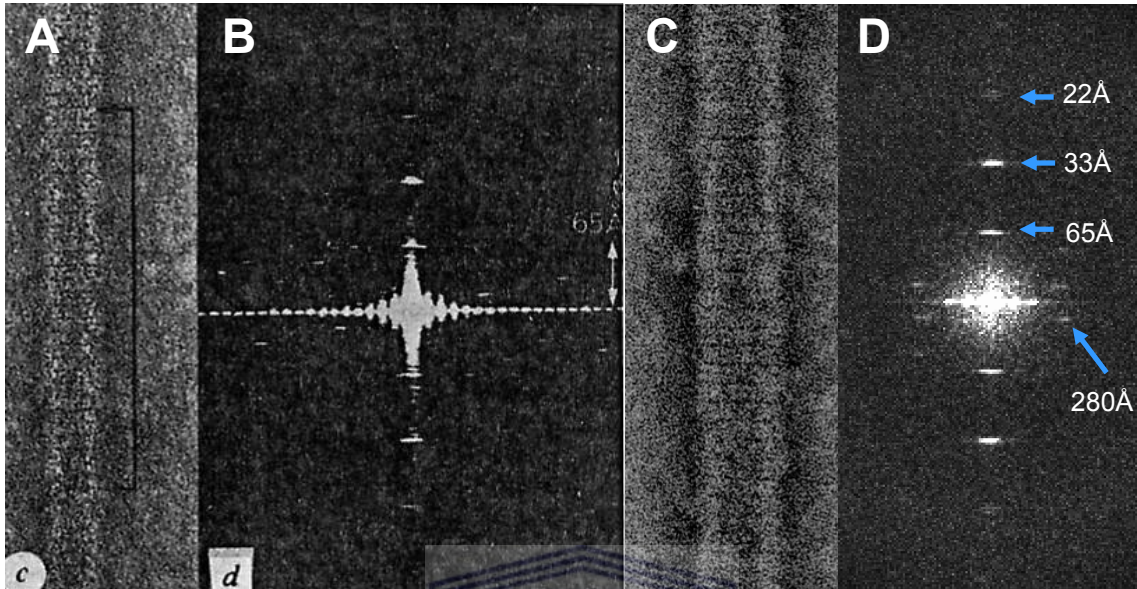

Figure 3.7 Power spectra. **(B)** Optical diffraction pattern of the helix shown in (A). Reproduced from Klug et al. (1980). **(D)** Computed power spectrum from helix shown in (C) (this study). Major meridional reflections are seen at 65, 33, and 22 Å. Because the 33 Å reflection is of second order to the 65 Å reflection and of similar intensity, the density of a single ring is divided into two equal layers - indicating the presence of a dyad axis. A single prominent off-axis reflection is seen at 280 Å, and corresponds to the pitch of the basic helix. Images are not to scale.

The first off-meridional reflection arises from the transform of helical lattice lines representing the most degenerate startedness of helix. These form oblique striations in the real-space image, reflecting the relative stagger of octamers between rings of the helix which repeat with a pitch of 280 Å/turn. Dividing this distance by the axial rise indicates a value of ~4.3 rings/turn for the pitch. The helical symmetry in this situation is not absolute though, because without determination of the cyclic symmetry an ambiguity exists as to what the axial rotation ($\Delta\Phi$) could be, and hence also the helical selection rule.

One way to try and confirm the number of subunits per ring is to look at the top views (figure 3.6B). These presumably represent end-on views of single rings, indicating stronger intra-ring contacts. Subunits seem to be joined together on the smooth inner circumference, while forming regular protuberances on the outer diameter in some

images (figure 3.6B images 2 & 3). Unfortunately, a reference-free average of 23 such top views does not enhance these protuberances and they cannot be explicitly counted. Similarly, individual top views enhanced by either 10- or 11-fold rotational averaging are ambiguous (data not shown). However, much like Klug et. al. (1980), the number can be estimated to be in the range of 10-12 subunits per ring. Additionally, it should be noted that single rings might not be truly reflective anyway, and might display greater point group heterogeneity, being outside the constraints of the helix.

Despite not being able to confirm the cyclic symmetry, these observations led to a favorable situation wherein at least the axial rise (Δz) and pitch could be predicted prior to a 3D reconstruction.

3.3 Preliminary 3D Reconstructions

3.3.1 *Start Model Creation*

The helices grown in this study closely resembled that described by Klug et al. (1980), and a start model was created for the IHRSR process using the octamer crystal structure (Arents et al., 1991), translated by their helical parameters. Based on their top view of a single ring, the octamer appears disc-shaped perforated by a large hole in the middle (figure 3.8C) - seemingly characteristic of the octamer when viewed down the superhelical axis (figure 1.5C). Hence this orientation was used and the octamer rotated about the center projected by its dyad axis, using a starting displacement needed to create a 10-subunit ring with diameter ~300 Å. This ring is then progressively copied by axial rotations and translations of -8.5° and 65 Å respectively, to generate a left-handed helix with 8 rings (needed to fill a 96^3 voxel box). Finally the atomic coordinates of the helix were converted to electron densities, down-sampled to 4.25 Å/pixel and filtered to 15 Å (figure 3.8B), to emulate the expected resolution of negative stain microscopy.
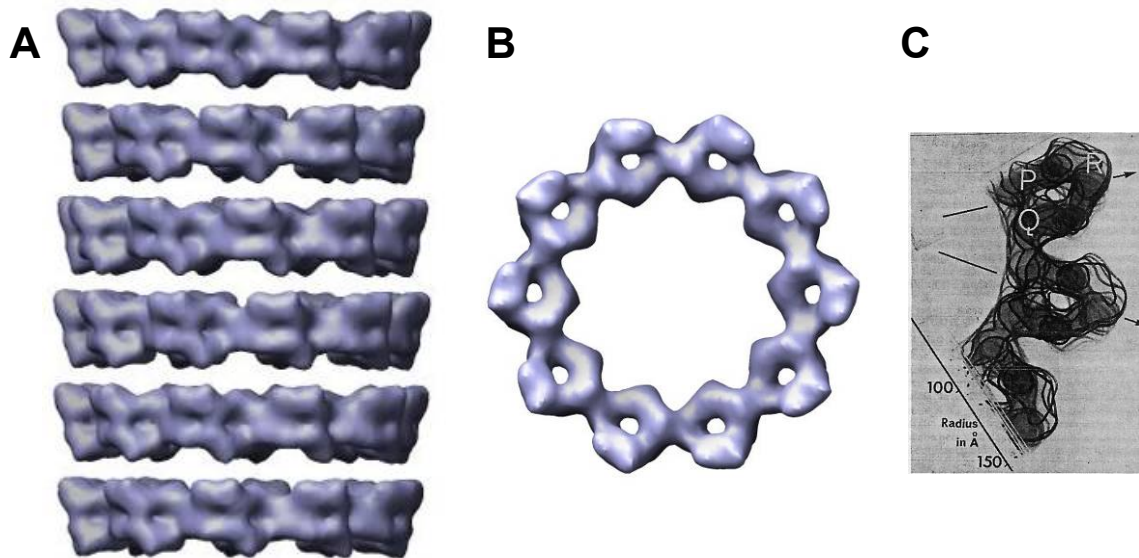
Figure 3.8 (A-B) IHRSR start model created by translating the octamer X-ray structure 2HIO (Arents et al., 1991), by the helical parameters of Klug et. al. (1980). The result is Gaussian-filtered to 15 Å - leaving a model which resembles their reconstruction **(C)**.

3.3.2 IHRSR of the CCD dataset with C_{10} symmetry

An attempt was made to reconstruct a preliminary 3D structure of the histone octamer tube from the CCD dataset, using the IHRSR procedure outlined in section 2.3. Briefly, because of the cyclic 10-fold symmetry, the start model, as described in the preceding section, only needed to be projected about the azimuth by 36° in order to cover the asymmetric wedge. Utilizing a 4° tilt increment led to the creation of just 9 references to which the entire CCD image stack of 697 images were aligned, during the multi-reference alignment steps. The alignment search range was set to 8 pixels or 65/2 Å to limit the search to only the nearest vertical alignment of rings vs. the reference and images were rejected if their mean translation exceeded a given threshold, or if their in-plane rotation deviated more than 15° from either 0° or 180° (after convergence only 2 images were rejected in this manner). No correlation threshold was enforced, since the in-plane rotation angle is a better discriminator of bad images for helices. Class-averages formed through angular assignment (figure 3.11A) were backprojected with imposed C_{10} symmetry, to form a 3D volume which was searched for- and symmetrized by- the underlying helical symmetry. Unlike the helical parameters which can be searched for each iteration round, the imposed C_{10} symmetry is a fixed assumption. The starting guesses used for the helical search were the same parameters used in the creation of the

start model: $\Delta z = 65$ Å, $\Delta\Phi = -8.5°$. The IHRSR procedure was then allowed to iterate over 50 rounds, although convergence was already achieved after roughly 7-10 rounds.

(**Note:** because the assumed sampling value for the CCD camera was incorrect, the true starting Δz value was actually 69.6 Å. All values pertaining to CCD reconstructions were subsequently adjusted to reflect the correct values at the specimen level, and thus aid direct comparison to the film reconstructions. See section 3.5.1).

The convergence and validity of the IHRSR method could be judged by the stability of the various iteration parameters, graphed in figure 3.9. The rise rapidly drops from the starting value of 69.6 Å to a stable value of 65.1 Å (figure 3.9A), as expected from the first meridional (figure 3.7). The axial rotation deviated little from the starting value and converged to an average value of -8.4° (figure 3.9A). The pitch, simply being a ratio of the axial rise and rotation, is still worth calculating for it can be directly confirmed by the power spectrum. Generally, it will also converge quicker than the other two parameters, because it represents the direct 2D visual feature of the long pitch helices (vs. axial rotation which is a 3D feature). It settles on a value of 280 Å, as measured for the first off-meridional (figure 3.7). In fact, none of these values would have needed to alter much, if the correct CCD sampling had been known from the start. Therefore all the values converged upon were completely expected, as was rationalized during the creation of the start model.

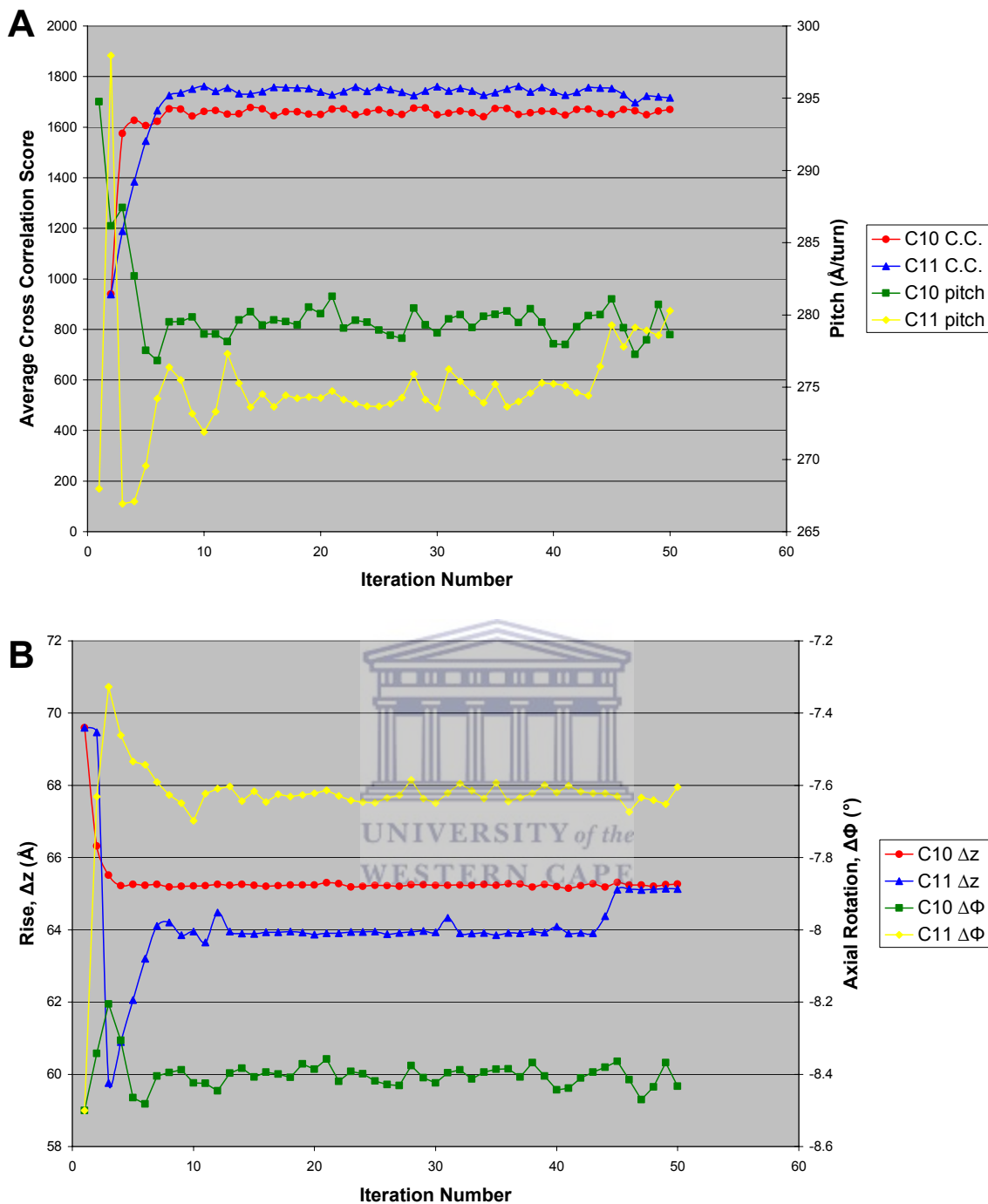**Figure 3.9** IHRSR parameters for both C_{10} and C_{11} CCD reconstructions, plotted over the 50 iteration rounds. **(A)** Average cross-correlation score and pitch. **(B)** Rise (Δz) and Axial Rotation ($\Delta\Phi$).

These results may be misleading, since the other parameters indicate problems with this reconstruction. The average cross-correlation score reaches a plateau after just one round,

and doesn't improve much beyond this (figure 3.9A), and means that the reconstructed volume is not properly iteratively refining. A histogram of the calculated x-shifts during the final round of alignment also shows serious problems (figure 3.10A), and the distribution is bimodal with a minimum centered on zero. This is completely unexpected as the boxed helices were already centered in the x-dimension via a reference-free approach. So in order to match the reference, helices are shifted away from the center, and should not happen if the reference is self-consistent with the images. This means the reprojected references are dissimilar to the reference-free average, and reference-bias is occurring through an induced fit. The y-shift histogram, on the other hand, was flat over its 8 pixel search range (not shown), as it is only dependant on the rise.

A test for self-consistency is the agreement between the class-averages formed and the projections of the model. Of course this is biased by the fact that images are aligned and classified based on their similarity to the reference projections. Still, the class-averages look less defined than the reprojections and contain very high noise levels (figure 3.11A & B). A fundamental of single particle techniques is that if similar views of a molecule are brought into common register, the average should have a vastly improved SNR. Also since the helices possess even-order symmetry the projections contain a mirror plane, which appears poorly-preserved in the class-averages.
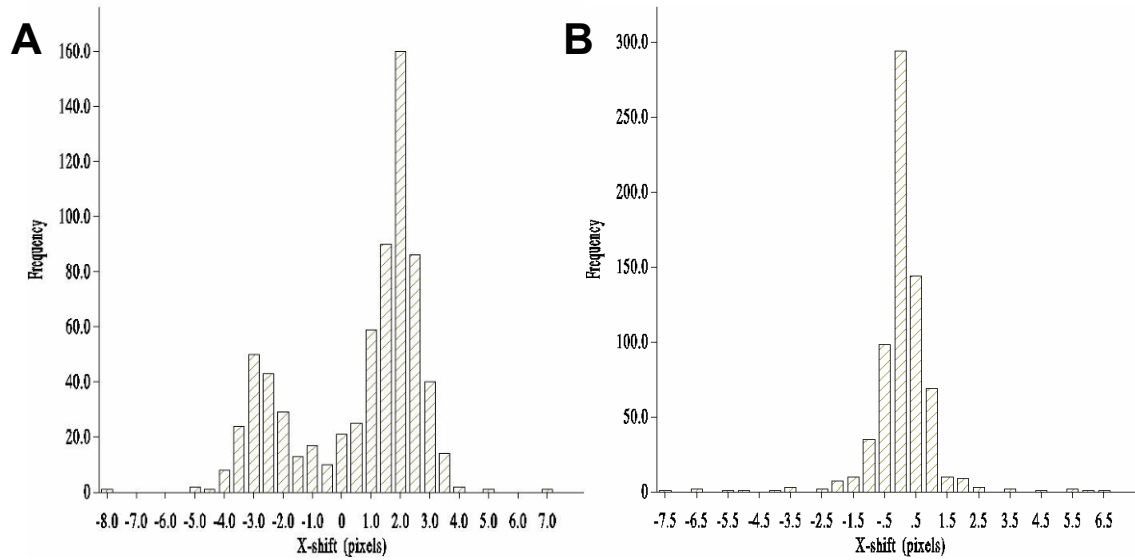
Figure 3.10 X-shift distributions. **(A)** C_{10} CCD reconstruction. **(B)** C_{11} CCD reconstruction.
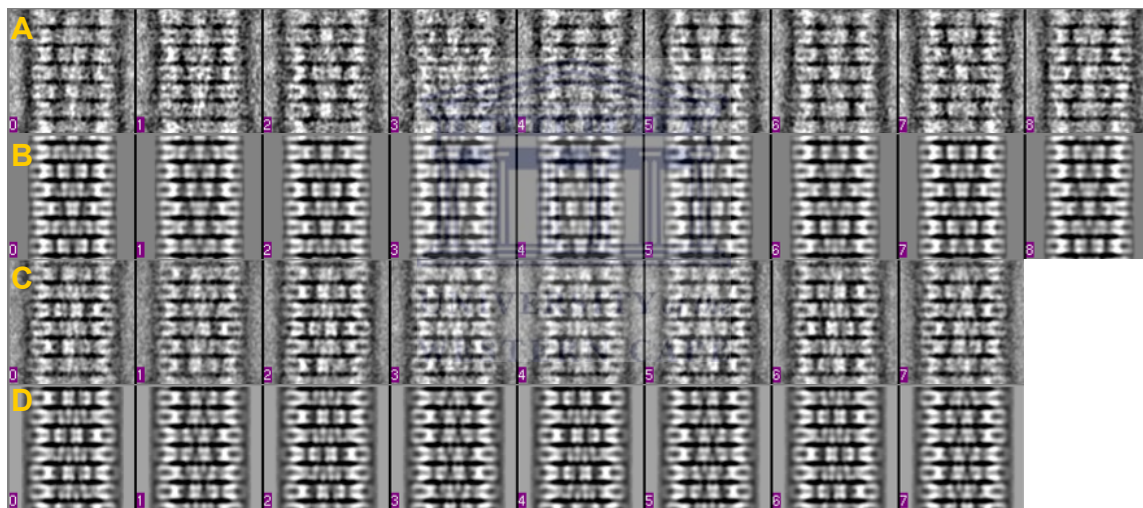

Figure 3.11 Class-averages and corresponding reprojections. **(A & B)** C_{10} CCD reconstruction. **(C & D)** C_{11} CCD reconstruction. Angular increments are 4° and 4.1° respectively.
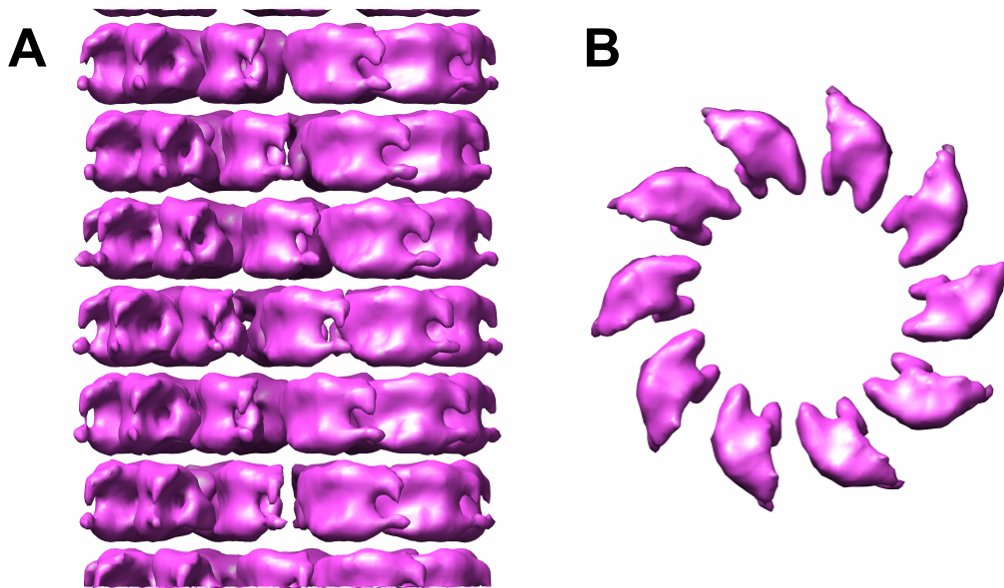
Figure 3.12 Reconstructed density from CCD dataset with imposed C_{10} symmetry after 50 iteration rounds. **(A)** Front view. **(B)** Top view of single ring.

Finally, although the reconstruction process can be invalidated by the alignment parameters alone, in this case the reconstructed density also violates important constraints. Figure 3.12 shows the final reconstructed volume, displayed at the correct density threshold calculated for the theoretical number of octamers included. Firstly, the individual subunits do not resemble histone octamers, with none of the characteristic views identifiable (figure 1.5). Secondly, subunits are seen to 'float' and display no lateral or vertical connectivity (at the correct threshold). Thirdly, histone octamers possess a dyad axis of symmetry which should symmetrize the front view (figure 3.11A); create a mirror axis in projection in the top view (figure 3.11B); and make the entire helix non-polar. None of these requirements are fulfilled. Thus, the entire reconstruction can be declared as artefactual - interestingly, considering the helical parameters had stably converged upon the predicted values. Consequently, a power spectrum calculated from a long projection of the reconstructed volume (similar to figure 3.29C red), also agrees back with the one calculated from a single raw image (figure 3.7D).

3.3.3 IHRSR of CCD dataset with imposed C_{11} symmetry

The simplest explanation for the previous reconstruction being artefactual is that the point group symmetry used was incorrect. The IHRSR procedure is robust in finding the

correct helical symmetry, but a fixed assumption about the point group symmetry could mean that the IHRSR process is prevented from ever converging to the true structure. Indeed, Klug and colleagues reported occasionally observing another class of helix containing 11 subunits per ring, but that these were less ordered and produced poorer maps (Klug et al., 1980). In a separate study, only C_{11} end-on views were ever seen, and a reconstruction based on a single image utilized C_{11} symmetry (Greyling, J. personal communication). Consequently, some questions were raised by that study as to the validity of the Klug et. al. reconstruction, and that the diffraction pattern was possibly wrongly indexed by assuming C_{10} symmetry. Thus, using the C_{11} point group instead for the following reconstruction was the simplest alternative.

The only changes implemented over the previous C_{10} reconstruction, was that the backprojection of class-averages was done with imposed 11-fold cyclic symmetry. The asymmetric wedge is hence reduced to 360°/11, and only 8 reference projections were needed spaced at an increment of 4.1°. The same start model was used despite being created with 10 subunits per ring. This is to avoid making assumptions as to how the helical parameters would change in response to the new point group.

Plotted on the same graph as the C_{10} reconstruction, the average cross-correlation score reaches a plateau more slowly and with a higher final value (figure 3.9A), indicating proper refinement (also no rejections occurred). The most striking change is that the axial rotation quickly converges on a value of -7.6° - almost a full degree from the starting value (figure 3.9B). The axial rise instead took 45 rounds to converge upon the final value, and therefore fortunate a full 50 iteration rounds were utilized, and this was probably related to the search range being too small to discover this global minimum earlier. This can be seen by the prominent climb in Δz from rounds 43 to 45 in figure 3.9B – prior to this adjustment, the value was quasi-stable and occasional spikes seen. The pitch being dependant on Δz also adjusts from 275 Å to 280 Å at this point (figure 3.9A). Both these two previous values are associated with a concomitant drop in cross-correlation score (figure 3.9A). Despite this, they at least agree better with both the previous reconstruction (although artefactual) and the expected values from the power

spectrum (figure 3.7D). Therefore, the only change brought about by the new choice of point group is that of the axial rotation, and an ambiguity thus exists because this value cannot be directly measured from the power spectrum.

The validity of this reconstruction over the previous one, and thus the true point group, can at least be resolved in real-space by other criteria. The class-averages and associated alignment statistics supply the strongest evidence here. Figure 3.10B displays the new x-shift histogram which now displays a proper Gaussian distribution centered about zero. The pre-centering has obviously worked well, with the majority of images located within a 1 pixel shift from the reference. As opposed to the C_{10} reconstruction, no obvious reference-bias is occurring and the reference-free center is similar to the reference-based center. The y-shifts of the two reconstructions are nonetheless equivalent (not shown), as this dimension is independent of pre-centering and rotational symmetry.

The x- and y-shift distributions additionally relate directly to the quality of class-averages formed. As a result of the better distribution of x-shifts, the C_{11} class-averages are a marked improvement over the C_{10} ones (figure 3.11 C vs. A). They display better defined subunits, an improved SNR, and a much lower background. Furthermore, the glide plane resulting from the projection of odd-order rotational symmetry is now better preserved than was the mirror plane in the even-order C_{10} averages. Interestingly, the C_{10} class-averages actually look astigmatic when compared to C_{11}, and with a concomitant loss of resolution in the x-dimension related to the misalignments.
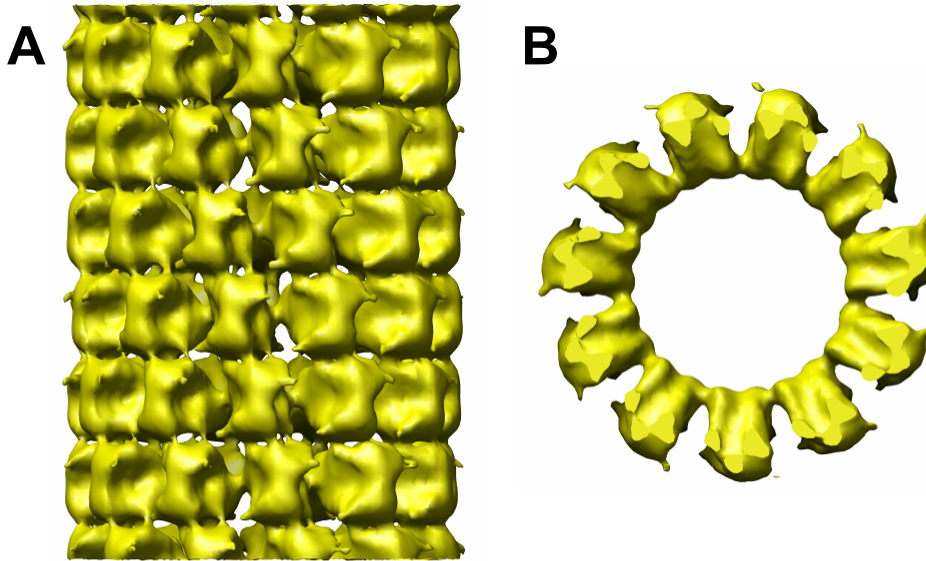
Figure 3.13 Reconstructed density using CCD data with imposed C_{11} symmetry after 50 iteration cycles. **(A)** Front view **(B)** Top view of single ring.

As a consequence of the anisotropy in the C_{10} class-averages, the map is seen to suffer from an apparent loss of resolution radial to the helical axis (figure 3.12). Solving this issue, the C_{11} map provides an improved density which is now readily interpretable (figure 3.13). The map is composed of clearly defined subunits of density which resemble histone octamers, identifiable by their three characteristic views (figure 1.5): seen directly toward the helical axis, subunits are propeller-shaped (figure 3.13A center of helix); toward the periphery of figure 3.13A, subunits are disc-shaped and possess a concave dent reflecting the position of the superhelical axis which is aligned tangentially to the helical circumference; finally, in top view (figure 3.13B), subunits display the characteristic wedge shape. The entire ring (figure 3.13B) also closely resembles the annular end-on views seen earlier in the micrographs (figure 3.6B), with the same shape and connectivity. The helix itself is stabilized laterally by this single central connection joining the octamers together into rings. Vertically, rings are stacked by what appear to be 3 regions of discrete connecting density (figure 3.13A). Because, in the micrographs helices are seen to break down into ring structures, the lateral interactions can be reasoned to be stronger and more significant. Ultimately, to confirm the orientation and connectivities of the octamers, the atomic structure of the histone octamer must be docked into the density.

The dyad axis was not enforced during the reconstruction, and therefore serves as a true independent test for the validity of the reconstruction (a test which the C_{10} structure failed). The dyad axis is identifiable in the front view and causes each blade of the propeller view to assume a similar identity (figure 3.13A). And in the top view (figure 3.13B), perpendicular to the dyad axis, each wedge is divided into two mirror-related halves (in projection). As a consequence of the dyad axis, the entire helix is also non-polar.

3.4 Confirmation of the C_{11} Point Group

Based on the evidence it appears that the helices observed in this study do in fact differ in their point group from the previous reconstruction (Klug et al., 1980). However, a possibility exists whereby the dataset could be composed of a heterogeneous mixture of these two types of helices as reported by Klug et. al. (1980). This is an especial risk since the dataset is composed of hundreds of individual helices, as opposed to Fourier-Bessel methods which typically rely on only a few well-characterized helices. The heterogeneity must therefore be disproven prior to acceptance of the C_{11} structure, since the presence of another helical form may degrade the quality of any reconstruction.

3.4.1 *Multi-Model IHRSR of the Film Dataset with Imposed C_{10}/ C_{11} Symmetry*
A reference-based algorithm was employed to in an attempt to sort and reconstruct the two possible point group helices simultaneously, and similar approach has been previously used to successfully sort variability arising from variations in helical pitch (Yu et al., 2001; Chen et al., 2004) or subunit composition (Galkin et al., 2002). Figure 3.14 outlines the procedure carried out on the film dataset (used because due to its larger size is more likely to encounter variability). 3500 images were aligned, assigned Euler angles, and now additionally classified by similarity to two projection sets derived from separate start models. The two start models differ by 10- and 11-subunits per ring, and consequently produce 9 and 8 references respectively when projected by a 4° and 4.1° increment respectively. The two sets of class-averages formed are then backprojected with imposed C_{10} or C_{11} symmetry into two separate models, which are then

independently searched for, and symmetrized by, their underlying helical symmetry. The new models are then used to form new references, and the entire process it repeated iteratively. In this way both models are free to discover and converge upon their true helical symmetry independently.
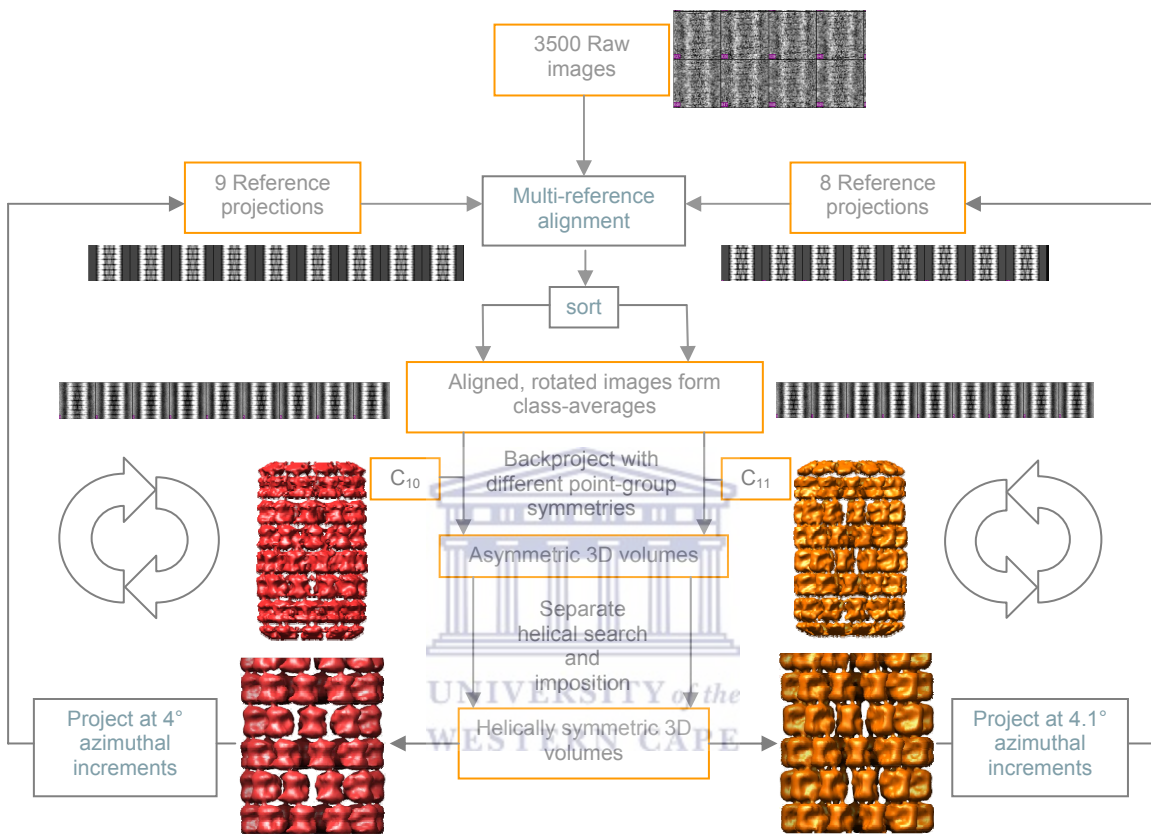

Figure 3.14 Multi-model IHRSR reconstruction scheme. It proceeds with two concurrent cycles of IHRSR which differ by back-projecting with either C_{10} or C_{11} symmetries, and separate helical symmetry searches.

From the starting helical value of $\Delta\Phi$, the two models quickly diverged to almost 1° apart (figure 3.15A). As with the separate CCD reconstructions, the C_{10} model stayed at -8.5°, while the C_{11} model dropped to a value of -7.6°. This has the effect that the pitch for the two helical forms was again in close agreement (figure 3.15B). Initial estimation of the sampling was much better for the film dataset, as the rise had remained at the starting value of 65 Å for both models (figure 3.15A). The average cross-correlation scores plateau rapidly for each model and then strangely oscillate in phase (figure 3.15B).
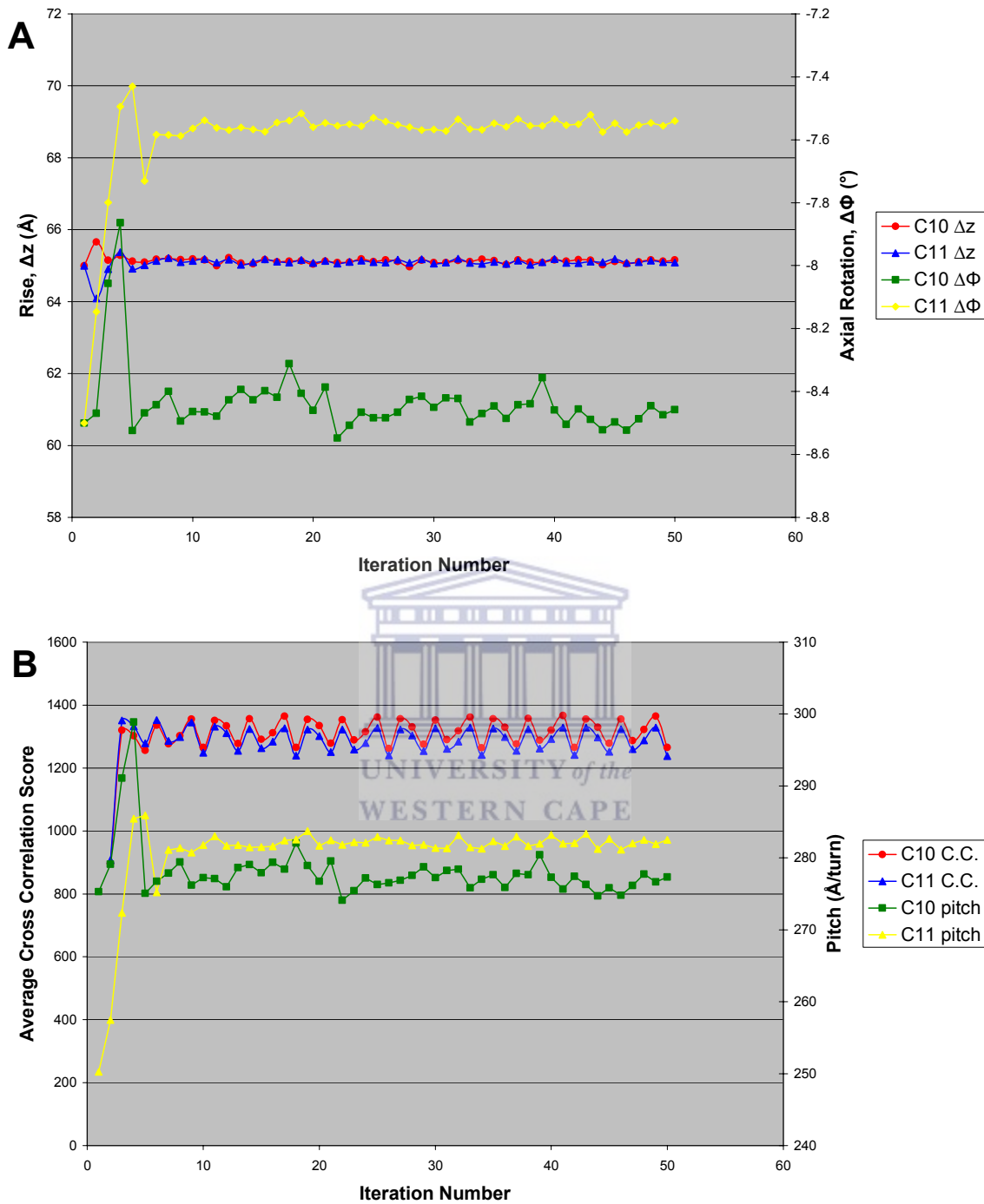
Figure 3.15 Resultant iteration parameters for the multi-model IHRSR, plotted over the 50 iteration cycles. **(A)** Average cross-correlation score and pitch. **(B)** Rise (Δz) and Axial Rotation ($\Delta \Phi$).

Figure 3.16 X-shift distributions for multi-model IHRSR film reconstruction. (A) C_{10} model (B) C_{11}

Figure 3.17 Class-averages and corresponding reprojections the multi-model IHRSR. **(A & B)** C_{10} CCD reconstruction. **(C & D)** C_{11} CCD reconstruction. Angular increment is 4° and 4.1° respectively.
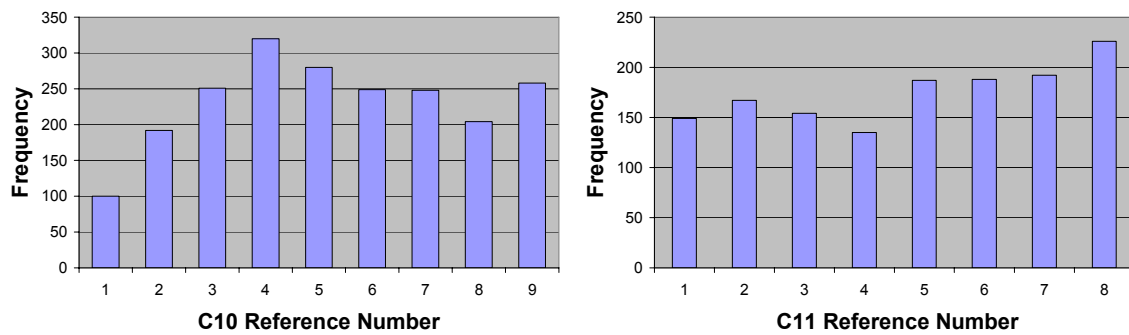
Figure 3.18 Reference distributions for the multi-model IHRSR.

Regarding the x-shifts and class-averages, similar conclusions can be made to the previous CCD reconstructions - that the C_{10} model is artefa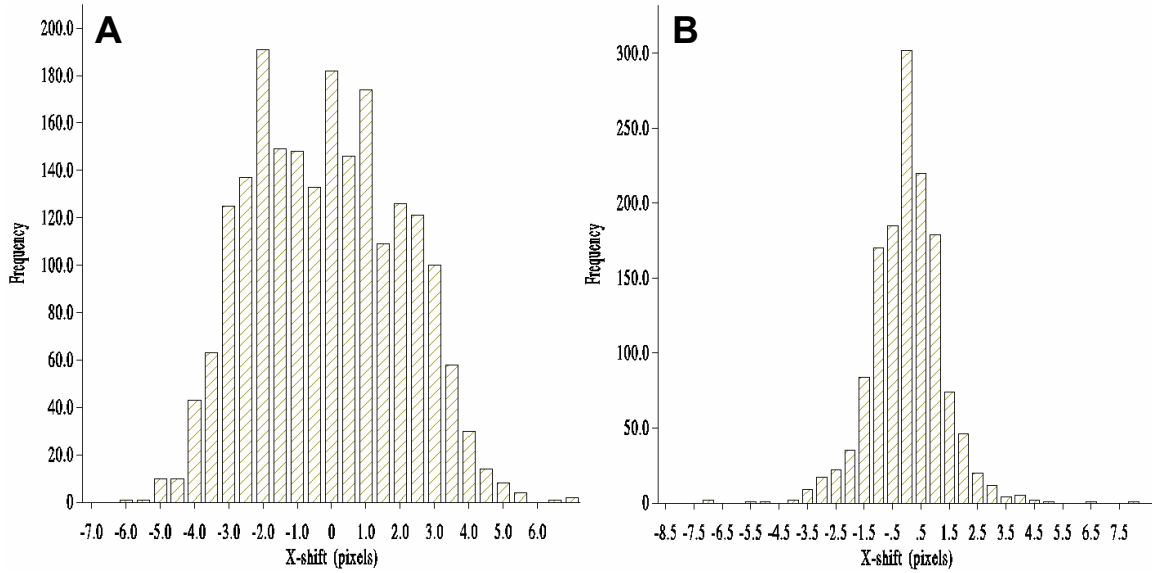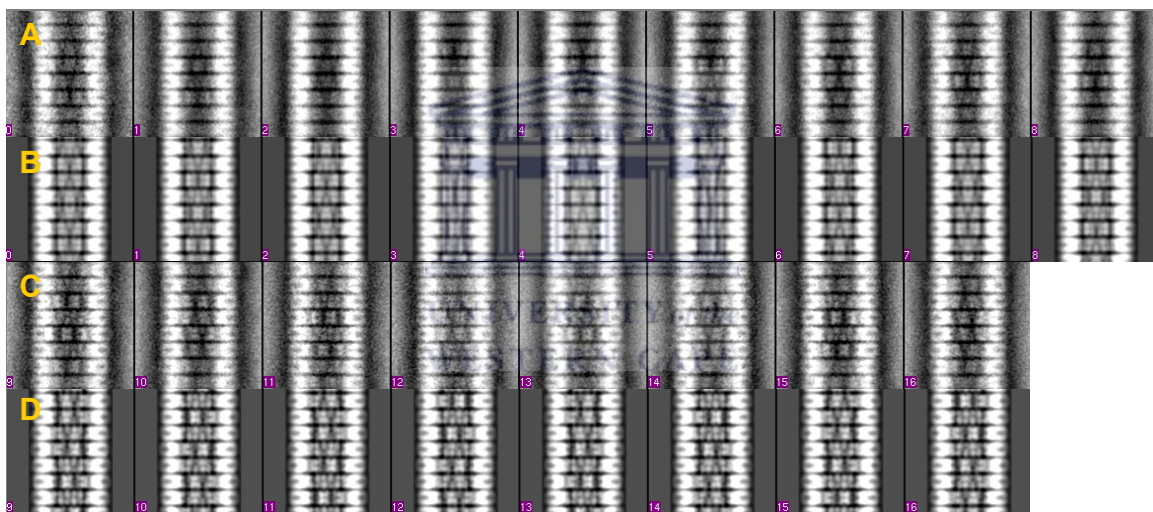ctual. The difference here is that images were free to choose between the two models based on correlation. X-shifts were more widely distributed (figure 3.16A) and consequent class-averages less defined for the C_{10} model (figure 3.17A). Additionally, the reference distribution also points to an artifact, since it is less flat for the C_{10} model (figure 3.18). An even distribution of views is expected, for the probability of each view should be equally likely.
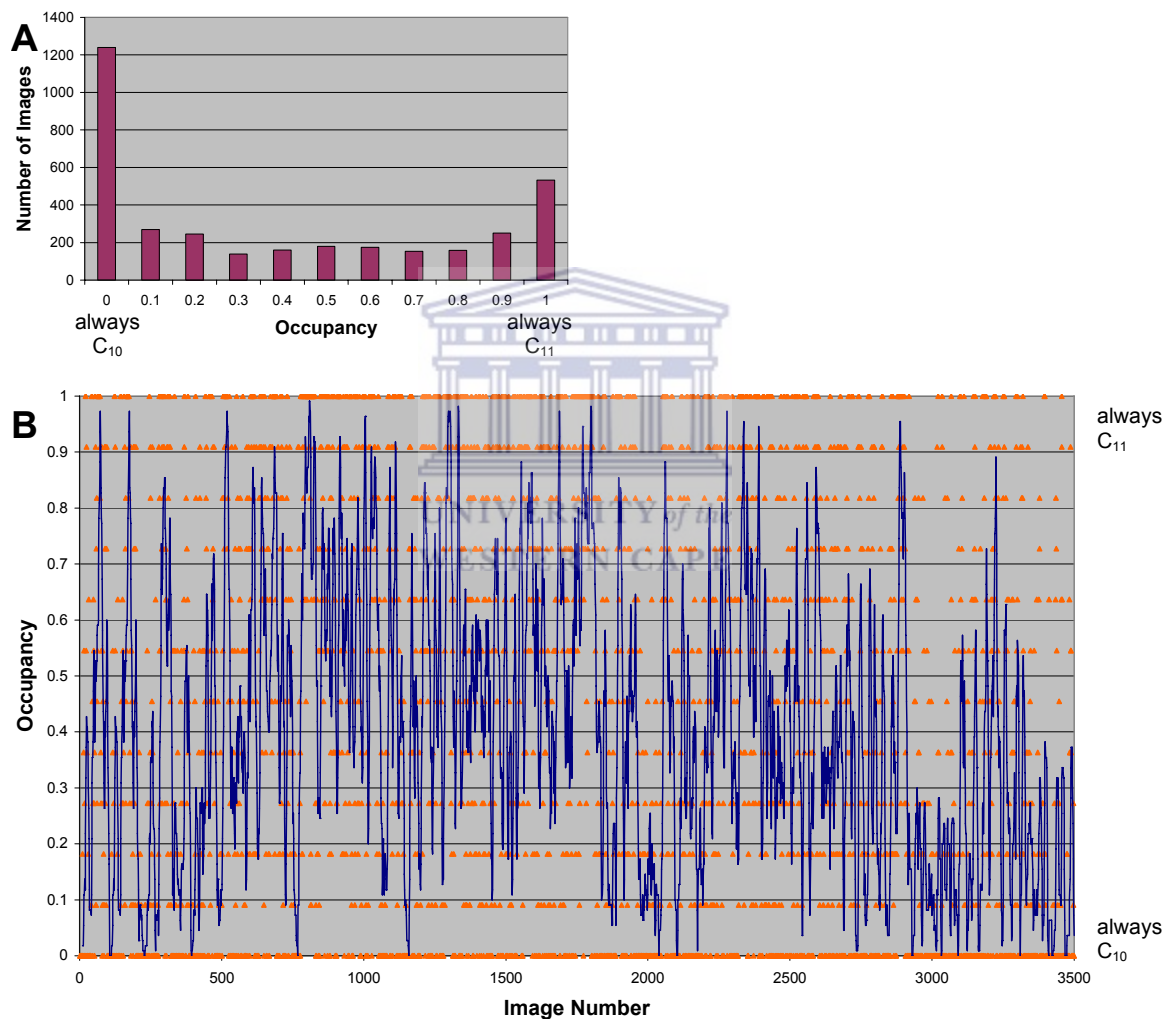
Figure 3.19 Multi-model IHRSR distribution statistics. **(A)** Average distribution of images between occupation of either C_{10} or C_{11} class-averages, calculated for the last 10 rounds of iteration. 0 reflects always belonging to a C_{10} class, while 1 always to a C_{11} class. **(B)** C_{10}/C_{11} class occupancy plotted per image (orange points). A trend line with a moving window of 10 images (blue) is added to identify data trends.

The map of this C_{11} reconstruction (figure 3.14 bottom right) is indiscernible from the previous CCD map and hence all deductions made thereof. However the C_{10} map has now changed, and is inconsistent with the one reconstructed from the CCD dataset (figure 3.14 bottom left). The subunits resemble distorted octamers by the presence of a rough dyad axis, but the large lateral separation between them indicates this is an artifact.

A test for the validity of the actual sorting process can be made from the distribution of images between the two reference groups. After convergence, each image is monitored each round for whether it was classified as belonging the C_{10} or C_{11} class, and assigned a value of 0 or 1 respectively. The average is calculated for the last 10 rounds and plotted as a histogram showing how the images are distributed between the two reference types. As after convergence images are expected to have a stable choice of reference, the distribution is expected to be bimodal. Figure 3.19A shows this is the case, but with an unexpected preference for the C_{10} reference. However, there exist a large number of unstable images termed class-jumpers, the sum of which is actually equal to the stable images. This is cause for concern.

If the population of images is truly a heterogeneous mix of these two types of helices, then they should correlate in blocks with the reference occupancy. This is because images are boxed sequentially along successive individual helices, accepting the minimum assumption that the point group is preserved within any given helix. Figure 3.19B plots a moving average with a window of 10 images of the occupancy, and reveals no significant clusters and that the images are randomly distributed between the two references models.

In conclusion, the C_{10} reconstruction is still artefactual even when the images are separated by the multi-model IHRSR procedure. The sorting itself was seen to be flawed, not because of problems with the approach, but rather because the dataset is probably not heterogeneous and composed only of the C_{11} form of helices.

3.4.2 *Eigen Images*

Multivariate statistical analysis was performed on the pre-centered film image stack using the IMAGIC package (van Heel et al., 1996). Briefly, images are plotted in multidimensional pixel space, and using hierarchal ascendant classification are grouped into clusters (van Heel and Frank, 1981). The vectors from the origin to the center of each cluster then represent the principal components of variation within the dataset, and can be expressed as eigenimages. Because the C_{10} and C_{11} helices are required to have different diameters, this would supply a major source of variation to the dataset. Figure 3.20 displays the eigenimages calculated, but no single eigenimage can be identified to obviously reflect a size difference. Image 3 is suspicious, but more likely reflects a variation in stain depth surrounding helices, as the striations contained within do not extend into the stain envelope. Another obvious source of variation between the C_{10} and C_{11} helices is the presence of a glide or mirror plane respectively (analogous to the phases being odd or even in Fourier-space – see section 3.4.3). However, none of the eigenimages representing subunits contain obvious mirror planes; only glide planes (images 8-12).
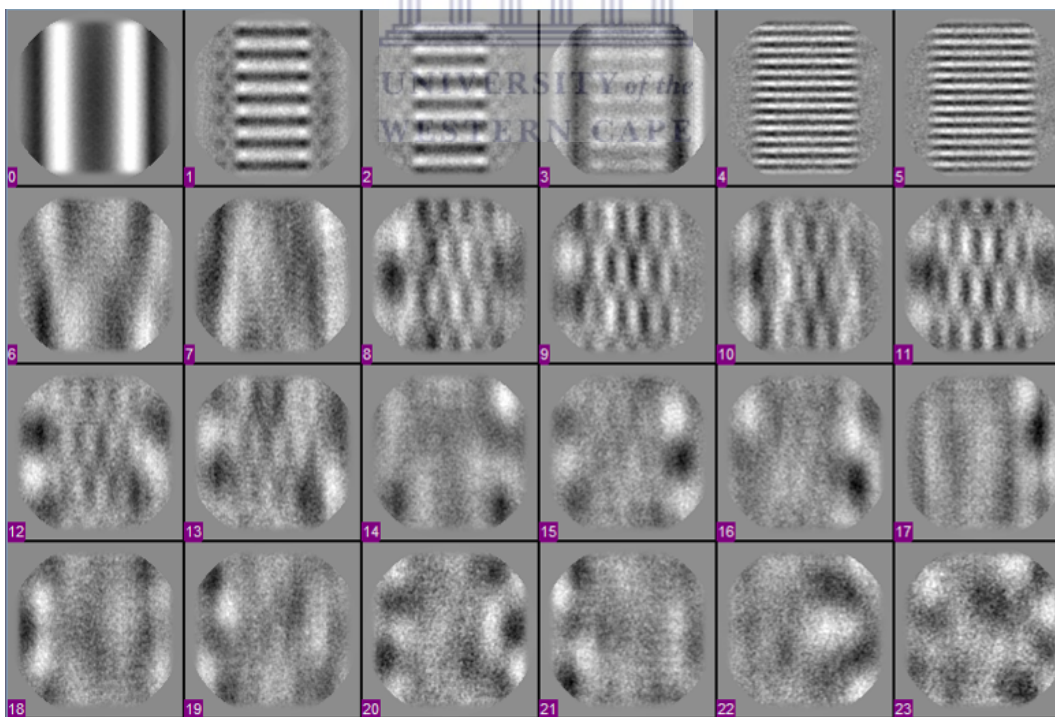

Figure 3.20 Eigenimages calculated from the film dataset. The principal components of variation which the Eigenimages represent are as follows: 0 - the global average; 1 & 2 - the rise Δz (65 Å); 4 & 5 - the second order of the rise (33 Å); 6 & 7 - misaligned images through in-plane rotations; 8-11 - the basic 11-start helix.

3.4.3 *Phases of the Fourier Transform*

The final most convincing analysis involves visualization of the phases of the Fourier transform along layer lines calculated for individual extended helices. In real space, a helix possessing rotational symmetry of even order, like the C_{10} helices, results in a mirror plane in projection. Odd rotational symmetry, however, like the C_{11} helices form a glide plane in projection. These symmetry planes are hard to identify in real-space through a noisy raw image (figure 3.6A), and the analysis is aided by migration to Fourier-space. Here the mirror-plane makes the phases for corresponding peaks on the left and right of the Fourier transform agree, and this effect is demonstrated using a projection from the artefactual C_{10} reconstruction (figure 3.21A). In opposition, since a glide-plane is simply a shifted mirror-plane, in Fourier-space it has the effect of flipping the phases of corresponding peaks by 180°. This can be seen in the $l = 4$ layer-line of a projection of the C_{11} reconstruction (figure 3.21B). The amplitudes along this layer-line may be similar, but the phases reveal the true differences.

For the actual analysis, 8 images of dimensions 1024×1024 pixels were boxed from the raw micrograph, ensuring enough long-range order to reinforce intensities in the Fourier-transform. The layer line used in this investigation was the $l = 4$ line, which contains the strong off-meridional 280Å intensities related to the most degenerate helix, with (presumably) n = 11 startedness (figure 3.7D). For all helices examined, the extracted $l = 4$ layer-line displayed phases which differed by ~180° (figure 3.21C). This is in contrast to the finding of Klug and colleagues (1980), wherein the phases were in close agreement for peaks across the meridional. In fact, this phase agreement was used as a measure of the quality of individual images. However, for this study it can be concluded that the order for this particular layer line is definitely odd. Thereby, the rings in this study definitely contain C_{11} symmetry, and represent a new helical class from those reconstructed by Klug et. al. (1980). Furthermore, although a small sample was used, no heterogeneity in the helix type was observed.
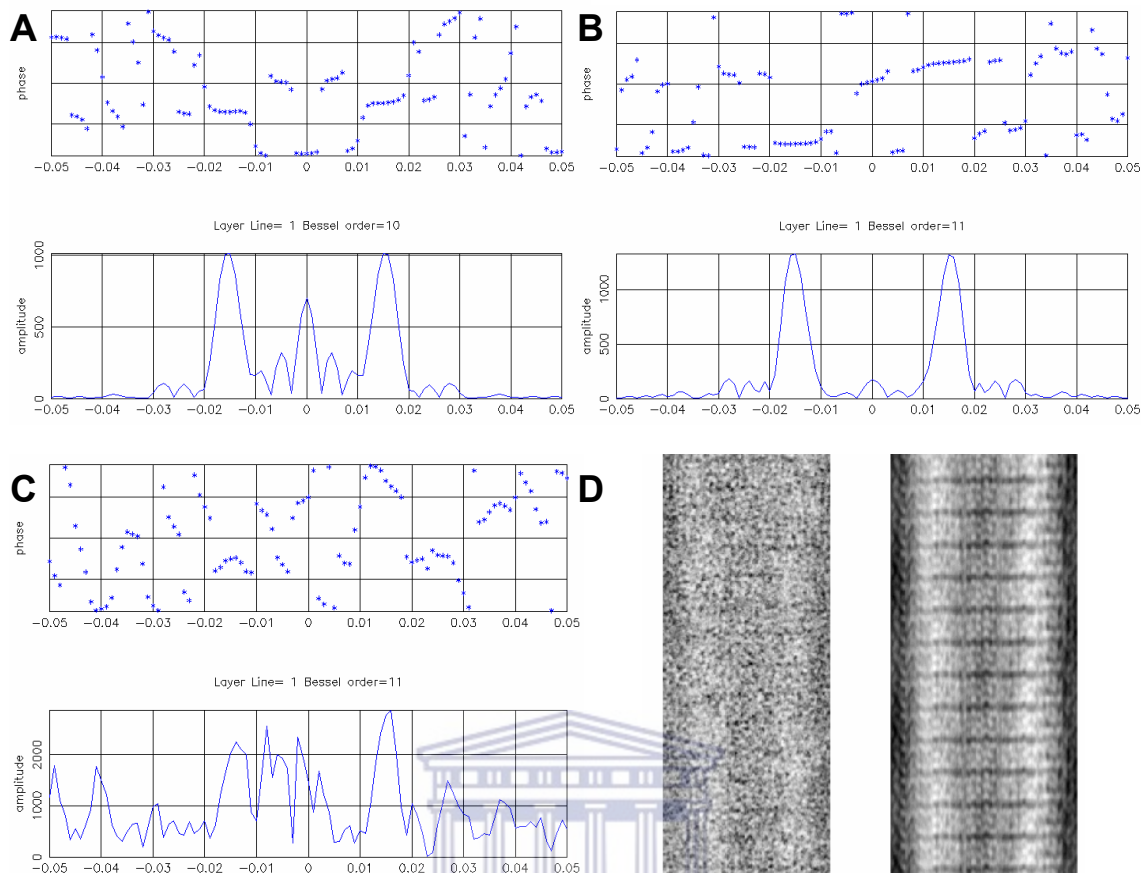
Figure 3.21 Amplitudes and phases of the Fourier transform, calculated from single images along the $l = 4$ layer line: **(A)** For the C_{10} model helix wherein phases are even for the maxima. **(B)** For the C_{11} model helix. Amplitudes are still identical the C_{10} model, yet the phases for the maxima now differ by 180°. **(C)** For the single raw image shown in (D), which is boxed from a raw micrograph. Here the phases clearly differ by close to 180°. **(D)** Quasi-optical filtration by back-transforming only diffraction occurring along selected layer lines, including $l = 0, 4, 17, 34, 51$.

The amplitudes of the 280Å intensities for the majority of helices are also unequal and this discrepancy can be regarded as weak evidence for the handedness of the helices (Moody, 1971). Partial staining means diffraction from the bottom of the helix is stronger and this specific pitch direction is enhanced over its mirror counterpart, which protrudes from the stain envelope. In this case, the amplitude for the reflection is generally larger on the left side of the Fourier-transform (unlike figure 3.21), corresponding to rightward-tilted long pitch helices being more predominately represented in projection. This in turn, indicates that the 11-start helix is left-handed, supporting the shadowing observation by Klug et al.

Further extraction of the all layer-lines which contain major diffraction intensities (at 280 Å, 65 Å, 32.5 Å, and 21.5 Å), and subsequent reverse Fourier transformation, resulted in a quasi-optical filtered real-space image (figure 3.21D). The major features of the helix could thus be reconstituted from just 4 layer lines, and interestingly reverse Fourier transformation of each major layer line individually, resulted in a real-space image resembling one of the eigenimages (figure 3.10). Thus the principal components of variation in the dataset are predominately the amplitudes of the diffraction intensities occurring on select layer lines. This makes sense, since they each represent a particular Bessel helical wave used to synthesize the helix, and MSA may present a novel way of identifying significant reflections during the indexing of helices. The false enhancement of weak reflections through flawed averaging techniques has already been seen to produce artefactual power spectra (Wang et al., 2006).

3.5 Final 3D Reconstruction with Imposed C_{11} Symmetry

Having eliminated the possibility that the film dataset were composed of a mixture of C_{10} and C_{11} helices, it was possible to proceed with a final reconstruction imposing just the C_{11} symmetry. This time a cylinder with radius 300Å was used as an *ab initio* start model, since the previous start models were incorrect anyway, and fewer assumptions are made this way. The first round still needs a starting guess for the helical symmetry however, and the parameters of Klug et. al. (1980) were used again. Similar to the previous C_{11} reconstructions, 8 references are formed by projection at a 4.1° increment of the asymmetric wedge, and backprojection was, of course, performed with imposed cyclic 11-fold symmetry.
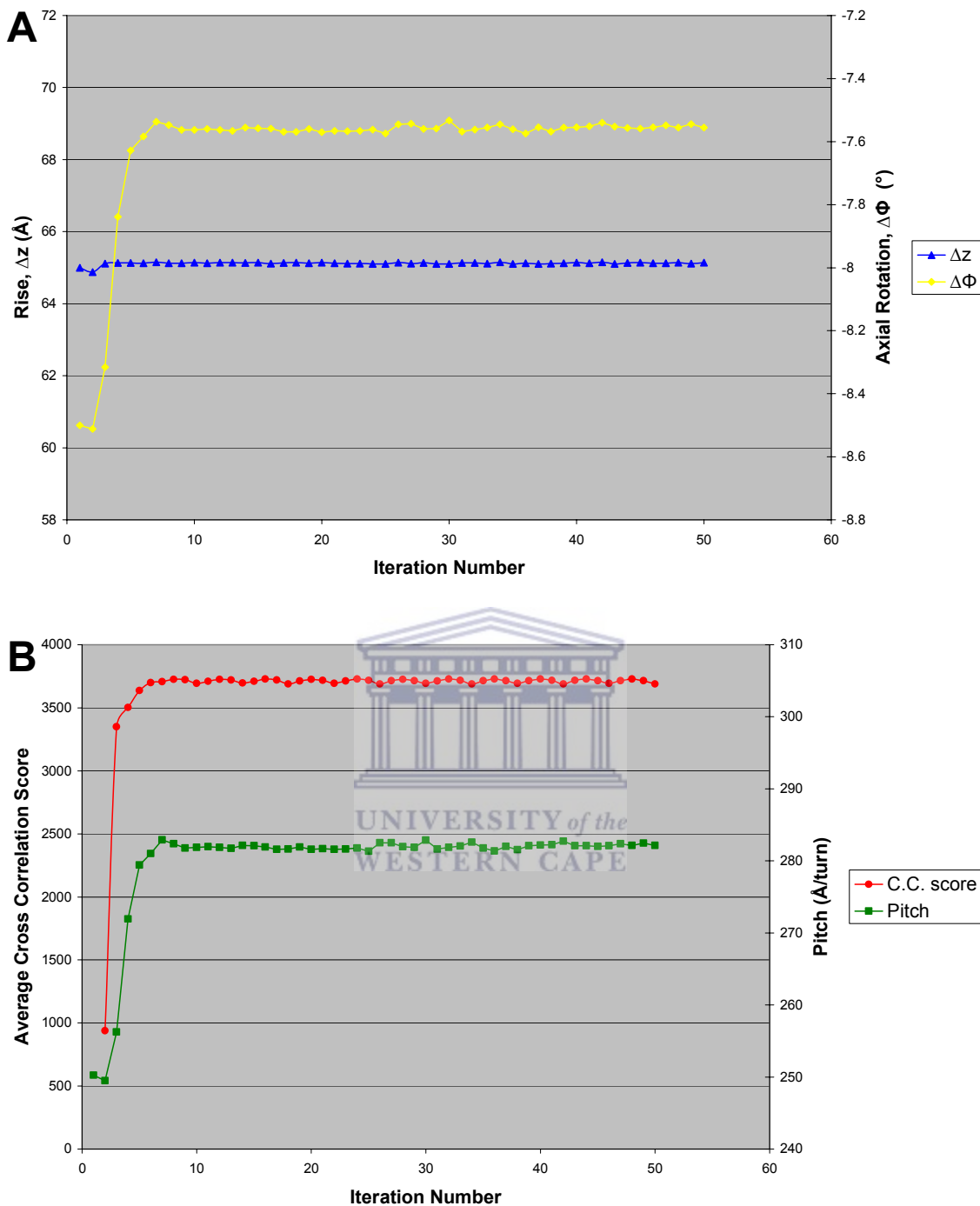
Figure 3.22 Iteration parameters for the final film IHRSR imposing C_{11} symmetry, recorded over the 50 iteration cycles. (**A**) Helical symmetry of axial rise (Δz) and axial rotation ($\Delta \Phi$) as found by least-squares searching. (**B**) Average cross-correlation score of all images, and pitch (Å/turn).

Judging by the various iteration parameters (figure 3.22), the reconstruction had converged and stabilized after just 7 rounds (quick, bearing in mind a featureless cylinder was used to start the process). Once stabilized, the average C.C. score was seen to oscillate slightly, and probably relates to a systematic rotation of the volume during symmetry imposition each round. The starting guess of the rise (Δz) was very good, and was stable at 65.1Å (figure 3.22A). In contrast, the axial rotation ($\Delta\Phi$) deviated by a full degree from its starting value and converged to an average value of -7.56° (figure 3.22A). Again, the point group and the axial rotation values combined to form a pitch of 281Å (figure 3.22B). Therefore the starting value of $\Delta\Phi$ was obviously wrong, because the pitch it denotes is 250Å, and is incompatible with the height of the $l = 4$ layer line intensities.

Well-centered x-shifts and an even reference distribution were a sign of a non-artefactual reconstruction (figure 3.23A & B). The class-averages used for backprojection match the reprojections well - similar to the previous C_{11} multi-model IHRSR, but now with less associated noise due to a greater number of images being used here (figure 3.23C vs. figure 3.17). One discrepancy versus the references though, is the central density of the helix which is weaker than the walls of the helix. This problem is not apparent in the CCD class-averages (figure 3.11), and probably relates to the better SNR of the raw images and/or stain depth in that case. It was hoped that the worse SNR of the film data, relating to the lower cumulative dose, would be made up for by the larger size of the dataset and better averaging during the formation of class-sums.
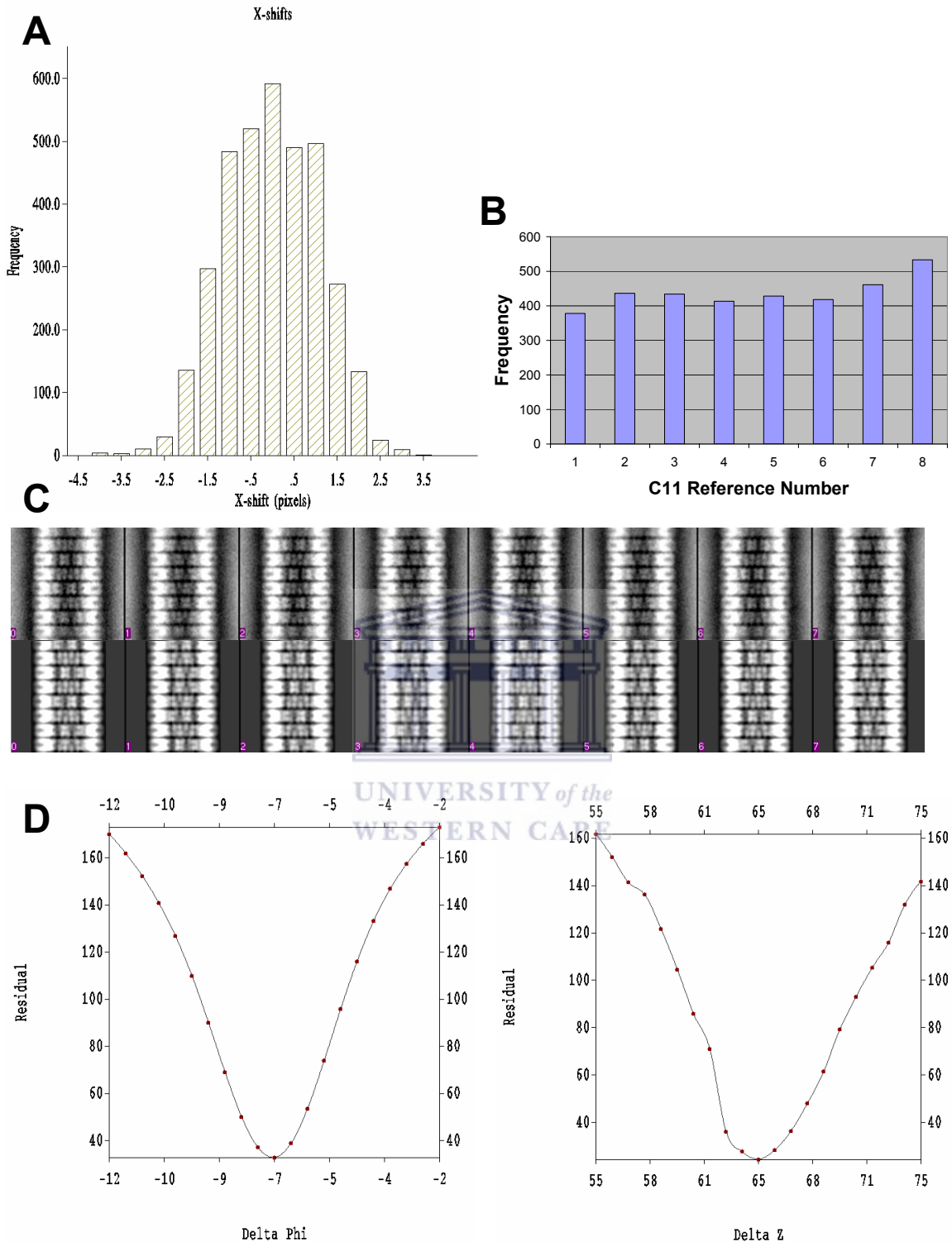
Figure 3.23 Various validity checks for the final C_{11} film IHRSR in the last iteration round. **(A)** X-shift distributions. **(B)** Reference distribution (images/class). **(C)** Class-averages (top) and corresponding reprojections (bottom). **(D)** Least-squares residual for the searching of underlying helical symmetry in the backprojected volume. Search radius is increased for illustrative purposes of demonstrating the large extent of the energy wells, indicating that the helical symmetry found is truly the global minimum.

The final reconstructed volume is incredibly well-defined, and in this sense has definitely benefited from the larger size of the dataset (figure 3.24). The symmetry conveyed by the dyad axis is enhanced: In the front view, octamer units display clear 2-fold symmetry especially in the 'propeller blades' (figure 3.24A); the same 2-fold being even more apparent on the interior view of the helix (figure 3.24B); and finally the 2-fold manifests itself as a mirror-axis in a top view of a single octamer ring (figure 2.24C). Thus, the octamer views are also clearly identifiable as before, to be confirmed later by the docking of the atomic structure into this volume. Vertical connectivity between rings is achieved by two distinct regions within each octamer, cross-sectioned in the top view (figure 3.24C). The horizontal intra-ring connections are, however, lost to higher threshold values. In conclusion, the histone octamer helix in this study can be described as an 11-start helix composed of stacks of octamer rings, related to each other by a left-handed axial rotation of -7.56° and a rise per ring of 65.1 Å. Combined this gives a pitch to the helix of 281 Å/turn or 4.36 rings/turn. The dyad axis of the histone octamer is identified to lie perpendicular to the helical lattice, as required by the constraints laid out for permissible helical line groups (Caspar and Makowski, 1981; Klug et al., 1958).
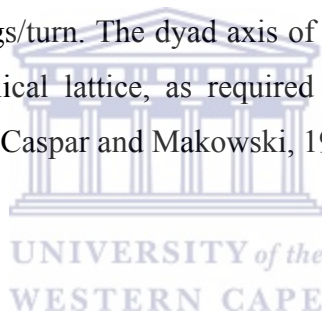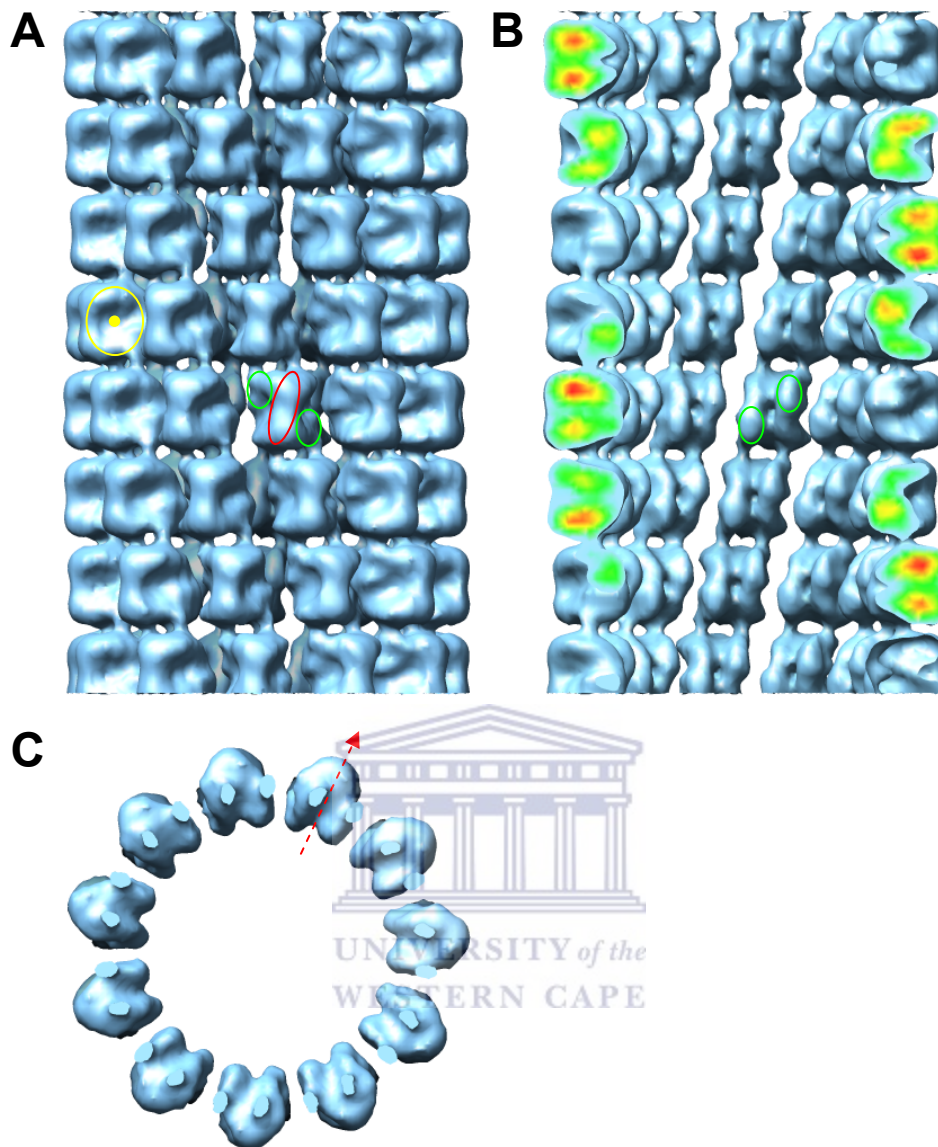
Figure 3.24 Final reconstructed density for the C_{11} IHRSR of film data. (**A**) Front view which displays characteristic propeller views (red ellipsoid) with flanking dimer 'hips' (green circles), as well as side disc views (yellow circle). (Compare with figure 1.5) (**B**) Cross-section through the center of the helix, revealing the inner lattice surface showing clear H2A-H2B dimer protuberances (green circles) across the dyad axis - characterized by a sharp indentation. Sectioned surface is colored according to internal density values. (**C**) Top-view of a single octamer ring displaying C_{11} point group symmetry. Two regions of density form the stacking connections between successive rings. The 2-fold dyad axis is indicated by the red arrow. Intra-ring connections are not visible at this threshold.

3.5.1 *Calibration of the CCD Sampling*

Having performed two independent reconstructions of the histone octamer tubes from both the CCD and film dataset, a large discrepancy was observed in the helical rise. This can be attributed to an inaccurate sampling value reported as 4.25 Å/pixel (after binning)

for the CCD camera. This value can be bettered though, because helices are an excellent calibration sample, and possess long range order and accurate rise values iteratively determined from hundreds of images. The film sampling is regarded as accurate at 4 Å/pixel since no adjustment was necessary when docking the atomic structure (See section 3.7), and both reconstructions are in agreement as witnessed by FSC (section 3.6.3). Therefore the discrepancy can be resolved by adjusting the CCD sampling to bring the rise value of the models into agreement. The average rise of the film reconstruction after convergence was 65.10 Å, while for the CCD it was 60.83 Å, and therefore the correct sampling for the CCD was calculated to be 4.55 Å/pixel. This reflects a 7 % error in the provisional value of 4.25 Å/pixel. All CCD reconstructions were retrospectively adjusted to reflect this change, but the starting guesses of the rise and pitch remain incorrect (unbeknown at the time and this emphasizes the importance of calibrating the sampling). Fortunately though, the initial power spectra were calculated from film micrographs (figure 3.7D).

3.6 Validation

The final film reconstruction procedure itself was validated by the convergence of the various iteration parameters: the average cross-correlation score reached a plateau; the helical parameters found were at the bottom of a large energy well; the input class-averages matched the reprojections; and the alignment statistics were favorable. A remaining test, however, is whether this is repeatable, and to what resolution as measured by FSC.

3.6.1 *Reproducibility*

For the second reconstruction, the same images and *ab initio* start model were used. Since the model was Fourier-filtered to 50 Å, this means they cannot agree, nor similarly bias the end reconstructions, to beyond this resolution. What separates the models then is the starting guess of the helical symmetry used in the first iteration round. Since the first reconstruction having been started at a ΔΦ of -8.5° and converged to -7.5°, it was decided to start the second reconstruction 1° on the other side of the expected end value at -6.5°.

This initiates the reconstruction on the opposite side of the energy landscape, and tests then the radius of convergence of the reconstruction. The axial rise value was not similarly altered, because of its unambiguous value calculated from the power spectrum.
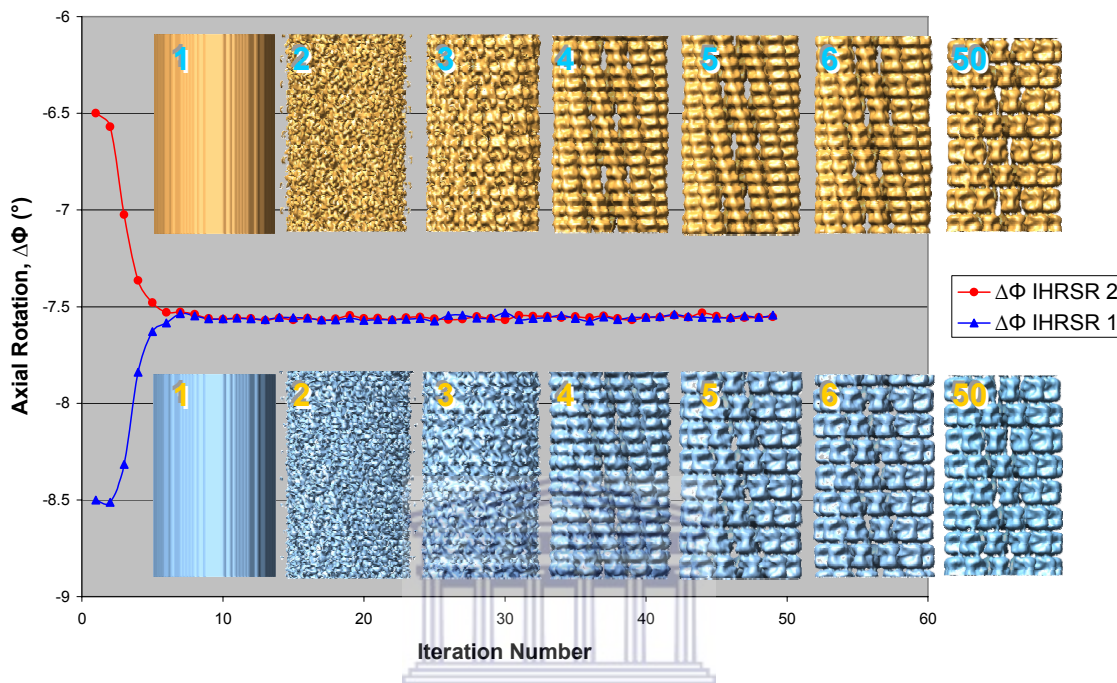
Figure 3.25 Convergence of two independent reconstructions started from *ab initio* cylinder start models and different values of axial rotation (ΔΦ), namely -6.5° and -8.5°.

Graphed in figure 3.25, the value of ΔΦ for the two independent reconstructions takes just 7 rounds to converge upon the same value, and the final values agree at -7.56° (an average calculated over the last 40 rounds). This rapid convergence is also confirmed by the increasing visual agreement of the two models over the initial rounds.

3.6.2 *Fourier Shell Correlation*

The resolution of these two independently generated volumes was measured by FSC to be 16 Å, using the familiar 0.5 criteria (figure 3.26) (van Heel and Harauz, 1986). This comparison was done between the asymmetric volumes, prior to the application of helical symmetry, in order to minimize the number of symmetric voxels in a given Fourier shell, otherwise leading to a spurious boost to resolution estimation. Still, a problem exists with this FSC as it never reaches zero correlation throughout the frequency range. The only

thing that can correlate as such high frequencies is the noise, and indicates that the noise component in the two reconstructions has become aligned (care was taken to remove common masking functions). This relates perhaps to the two reconstructions not being truly independent, by the fact that they are derived from the same image stack – testament to the idea of using half-stacks rather.
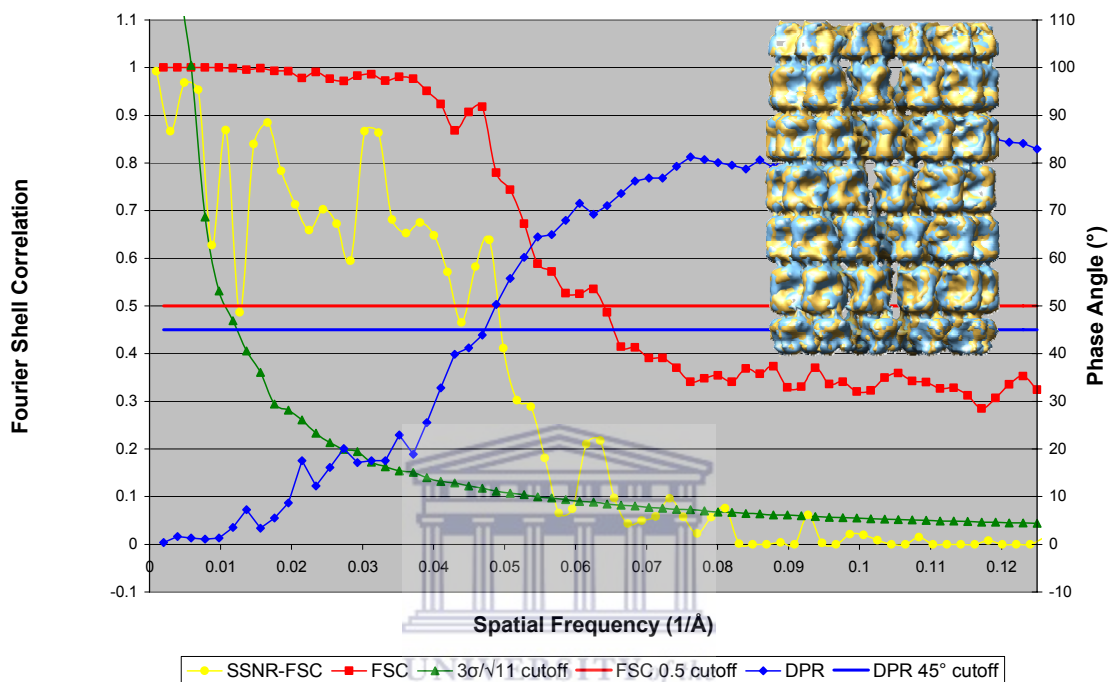
Figure 3.26 Resolution estimation of the final reconstruction by FSC, DPR and SSNR. Overlaid are the asymmetric volumes of the two independent film reconstructions. Estimates are 16 Å at 0.5 FSC, 20 Å by DPR and SSNR.

For this reason, two different resolution criteria, the spectral SNR (SSNR) and differential phase residual (DPR), were used in addition. The SSNR has an advantage over the FSC, in that it directly tests the consistency between the input class-averages and the reconstructed map, therefore overcoming the need for independent reconstructions (Unser et al., 2005). It is more convenient to express the SSNR as a FSC curve using the relation:

$$FSC = SSNR / 1+SSNR.$$

At a value of 0.5, which then corresponds to a SNR = 1, the resolution estimate is a more realistic 20 Å (figure 3.26), and more trustworthy since it now indicates zero correlation

at high frequencies. Equivalently the DPR 45° cutoff also indicates the same resolution, and nicely follows the derivation that the DPR 45° is equivalent to a SNR = 1 (Radermacher, 1988). A 3σ noise threshold curve is also included as an alternative to the fixed-value cutoff used for the SSNR, and is multiplied by the square-root of the imposed symmetry, i.e. $\sqrt{11}$. Using this more optimistic measure, the resolution is estimated at ~18 Å (Orlova et al., 1997). Thus, taking into consideration all these different measures, the final resolution can be stated to be in the range of 16-20Å, depending along which axis is being considered in polar coordinates.
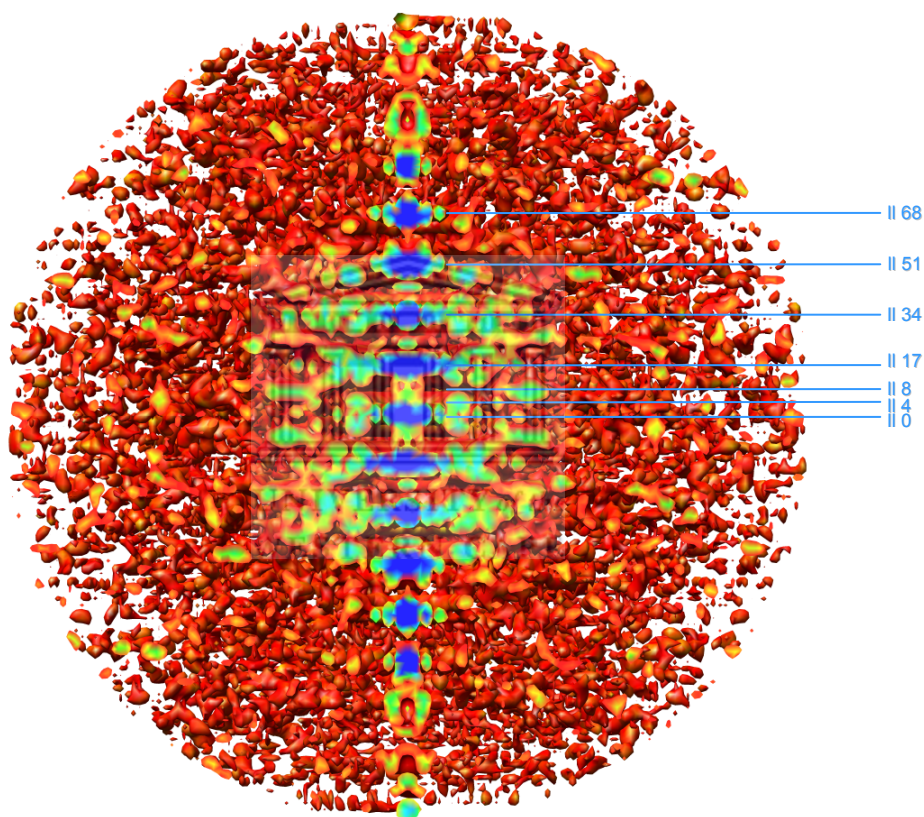
Figure 3.27 Cross-section through the volumetric SSNR contoured at a SNR = 1. Signal-to-noise is seen to be better along the z-axis, with peaks occurring at discrete layer plane spacings labeled with blue lines. Layer spacings are labeled according to the selection rule of Klug et. al. (1980). Sectioned surface is colored according to SNR: 4 = red, 8 = yellow, 16 = green, 32 = turquoise, 64 = blue.

Interestingly, the SSNR can also be expressed as a 3D volume, and displays the angular distribution of SNR at a given threshold in Fourier-space (figure 3.27). The resultant volume at SNR = 1 shows anisotropy, obviously relating the backprojection geometry of a single-axis tilt series. This anisotropy is related to better sampling of structure factors

on the z-axis, but does not mean that Fourier-space has not been properly covered elsewhere. Some questions must then arise as to the validity of FSCs when applied to helical objects, since it involves the calculation of correlation within spherical shells. Also, distinct on-axis striations and ellipsoids arise from a better SNR occurring on the distinct layer planes. Better correlation is therefore expected when a given Fourier shell intersects any given layer plane in the 3D Fourier transform, leading to FSC oscillations.

3.6.3 *FSC: film versus CCD*

Instead of repeating the above independent reconstructions using half-stacks, we can achieve a similar result by FSC between the film and CCD reconstructions. These two volumes are truly independent since they are generated from separate starting models, and furthermore exhibit unique noise statistics relating to the different capture methods (which cannot become correlated as before). A downside is that the FSC determines the resolution of the worse model - in this case the CCD model. Comparison first necessitated the down-sampling of the CCD reconstruction from 4.55 to 4 Å/pixel via interpolation, and this asymmetric volume (without imposed helical symmetry) was then aligned to the film volume via cross-correlation to produce good visual agreement (figure 3.28 see inset). The FSC between these volumes drops below 0.5 at 24 Å and crosses the $3\sigma\sqrt{11}$ threshold soon after at 21 Å, and accordingly fits with the estimation of the film reconstruction being better than this.
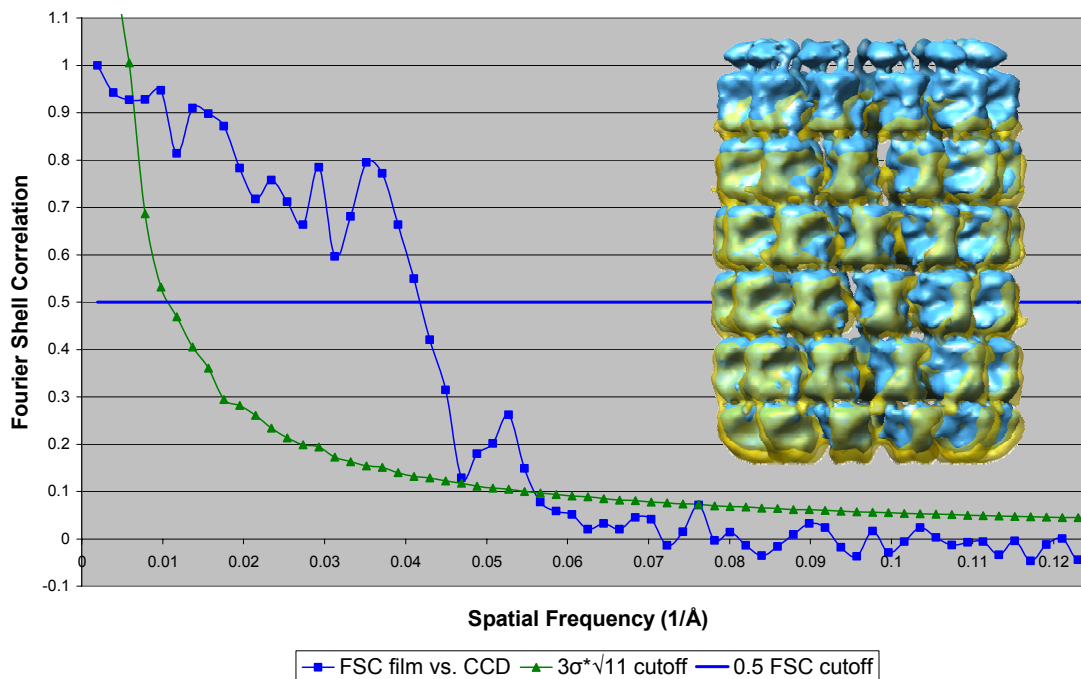
Figure 3.28 Resolution estimation by FSC between the film and CCD reconstructions. Overlaid are the two aligned volumes: blue – film map, yellow – CCD map. Resolution is estimated at 24 Å and 21 Å by 0.5 and 3σ FSC cutoffs respectively.

3.6.4 *Power Spectra Agreement*

A final test of self-consistency is the agreement between the observed power spectrum and the computed power spectrum from a projection of the final reconstruction. This is an unbiased measure of the structure, since the positions of the diffraction maxima are linked to the helical symmetry, and therefore should be identical. This is especially true of IHRSR, where the positions were never used in the determination of the helical symmetry - like in the indexing of the Fourier transform of Fourier-Bessel methods. Since the amplitudes of the complex transform are shift-invariant, it permits the averaging of power spectra from individual image segments in order to boost the SNR beyond that of a single filament (figure 3.7D), however the phase information of the complex transforms cannot be similarly averaged and are ignored. Thus the averaged amplitudes alone form the 'gold standard' for IHRSR (Wang et al., 2006), wherein one works with image stacks composed of small boxed segments of helix.
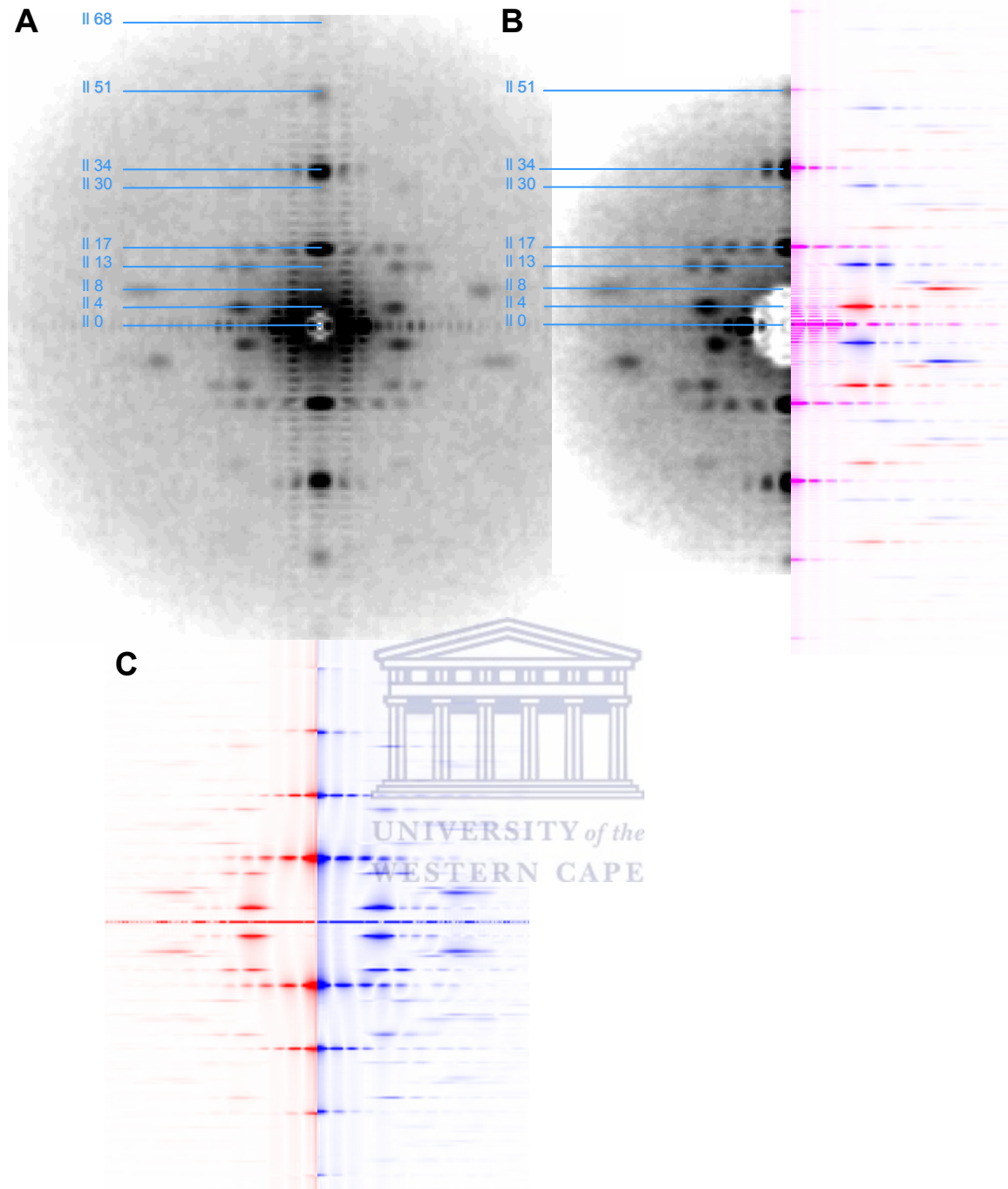
Figure 3.29 Observed power spectrum generated from 3500 segments of the film data **(A)** and 697 segments of the CCD data **(B).** Superimposed power spectrum computed from a projection of the final C_{11} film reconstruction is overlaid on the right. For clarity diffraction from near (red) and far (blue) lattice surfaces are separated by color. Approximate layer spacings are indicated by the index *l*, according to the helical selection rule of Klug et. al. (1980) **(C)** Side-by-side power spectra calculated from projections the C_{10} (red) and C_{11} (blue) models from the multi model IHRSR procedure (section 3.3.4.1). The various spectra were brought to scale by adjusting their respective layer line spacings.

The averaged power spectra of the observed CCD and film data are indistinguishable from each other, and well as from the power spectrum computed from a projection of the final C_{11} reconstruction (figure 3.29A&B) in the position and spacing of layer line maxima. It is surprising then, that although these power spectra agree, that they equally agree with the power spectrum of the artefactual reconstruction with imposed C_{10} symmetry (figure 3.29C blue vs. red). In this sense the power spectrum is not a good discriminator, and there exists a fundamental ambiguity wherein the true point group cannot be resolved. As such, the 'gold standard' of the observed power spectrum fails to identify the C_{10} reconstruction as artefactual.

The observed power spectrum also indicates that the film images diffract to a higher resolution, with a faint meridional reflection at $l = 68$, corresponding to a spacing of 16 Å (figure 3.29A). In contrast, the last reflection in the CCD data only occurs on $l = 51$ with a spacing 21 Å (figure 3.29B). Since a helix is defined by diffraction only on discrete layer lines, the last observable reflection is perhaps a better estimator of resolution than FSC (at least along the helical axis).

The principal maxima on layer lines $l = 4$ and $l = 8$ in figure 3.29A show a clear deviation from mirror symmetry about the meridian (or across the equator in figure 3.29B). When compared with the model power spectrum, these two maxima in the top-right quadrant have shifted inwards (this comparison is made bearing in mind that contributions from the edge of the helix broaden the principal maxima outwards for the model power spectrum). Knowing the absolute hand of the basic helix (left-handed), and therefore the sign of the order these reflections represent ($n = -11$ and -22 respectively), they can be deduced to represent diffraction arising from the top surface of the helix. Thus the decrease in distance from the meridian, or R_m value, represents a broadening of the helical lattice lines they represent as designated by their respective (n, l) values. This broadening is seen as a manifestation of an increase in the effective radius of the helix (r_e) (Trus and Steven, 1984), since the lattice is confined to a circumference of $2\pi r_e$ once indexed. The physical cause of this can attributed to the collapse of the upper surface of the helix during staining and dehydration, and is plausible for a large hollow helix as

such. This explanation is favored since an increase in the real-space radius was later confirmed by the docking results. The bottom surface, on the other hand, is largely undistorted, a fact substantiated by the broader peak-width of the right-handed reflections, especially noticeable with the maxima at $n = 22$, $l = 8$ (figure 3.29B, top left quadrant). Again this effect arises from contributions of closely spaced lattice lines near the edge of the helix and thus indicates a more cylindrical surface for the bottom surface - probably because it is better preserved by the stain. Conversely, the narrowness of the near side peaks indicates a flattened profile for the top surface.

3.7 Docking

Docking of an X-ray structure into an EM map serves the purpose of allowing detailed analysis of the helical intermolecular contacts formed and their possible biological roles. Figure 3.30 illustrates the sequential docking scheme used wherein after each round of docking, new constraints were discovered and applied in the successive round to improve the overall accuracy. Ultimately, this led to an atomic model of the histone octamer helix, which could be used for hypothesis development.

3.7.1 *Rigid Body Docking*

Initial docking was performed using a single copy of the histone octamer crystal structure (2HIO) and the EM map truncated to 3 rings - thus containing density for 33 histone octamers. This approach thus disregards any possible constraints implied by the helical and line group symmetries, and would usually lead to problems through partial occupation of multiple available subunit densities. In this case however, it is not expected to be a problem since individual subunits are well segregated, and fitting contrast is additionally enhanced through the use of a Laplacian edge filter. No constraints meant that rigid body docking required a full 6-dimensional cross-correlation search with a 15° Euler angle increment within Situs (Wriggers et al., 1999). After docking the IHRSR helical symmetry could then be retrospectively applied to the best solution in order to generate the completed helix.
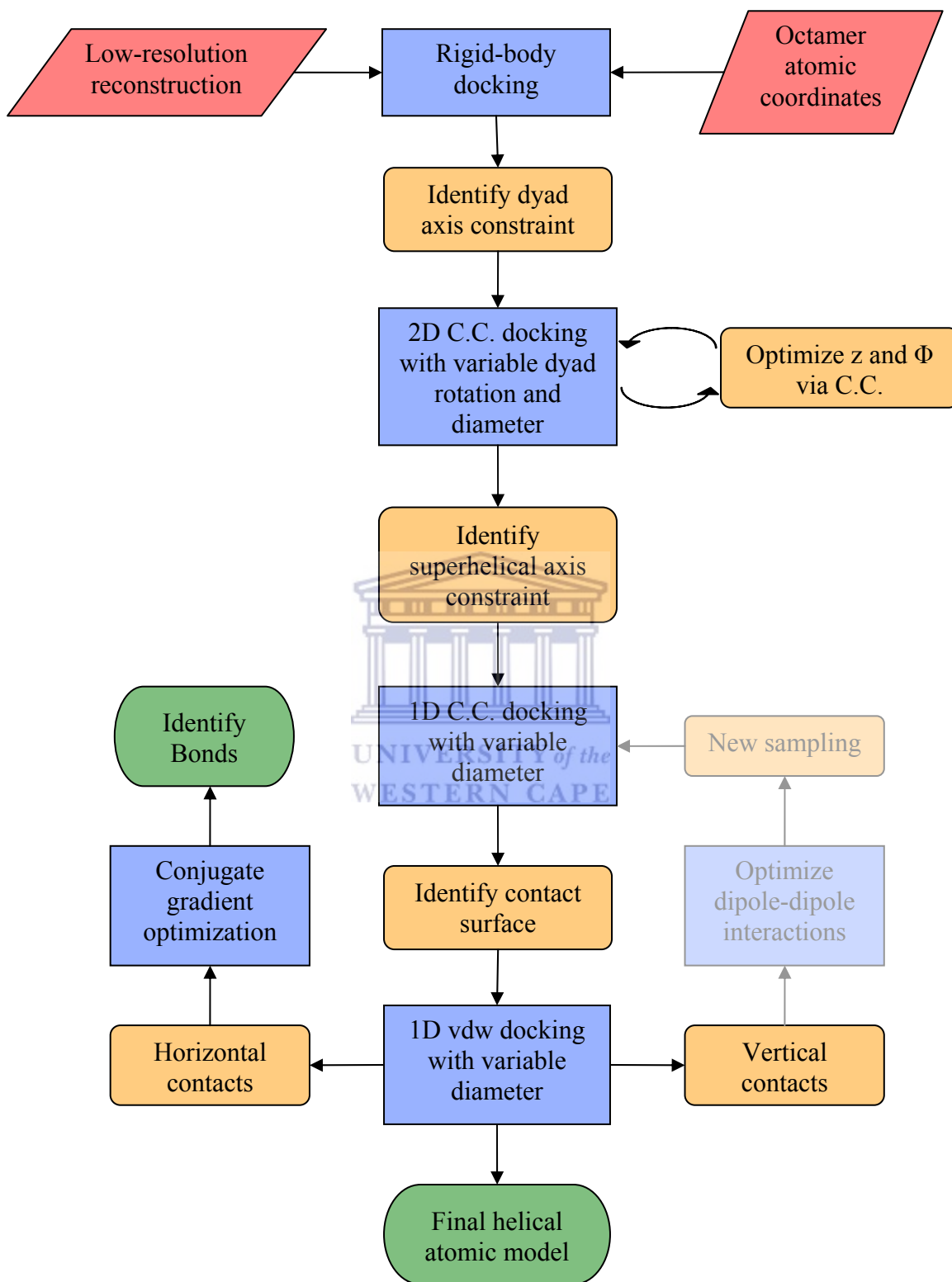
Figure 3.30 Complex docking scheme used to produce an accurate atomic model of the histone octamer helix.
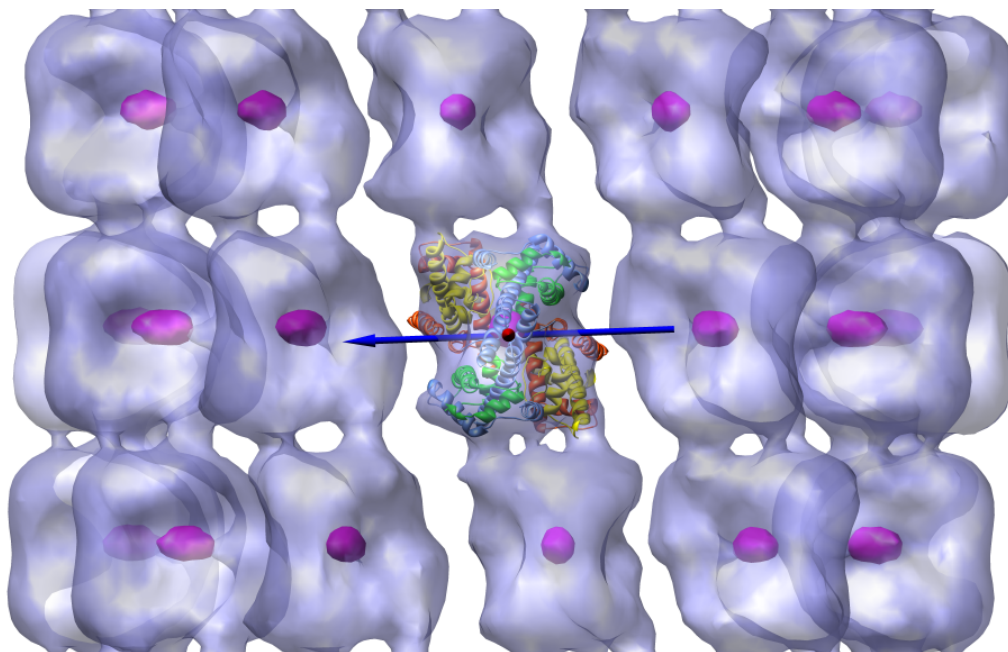
Figure 3.31 Rigid-body docking result using Situs (Wriggers et al., 1999). Translational correlation peaks are shown as purple ellipsoids. The superhelical and dyad axes are shown as blue and red arrows respectively (dyad arrow is seen head-on).

Overall, this resulted in an excellent fit of the crystal structure with no major unaccounted density (figure 3.31), although the atomic structure is seen to be shifted too far radially inward, relative to the map. This can be attributed in part to the Laplacian filter, which would favor breaking the single smaller surface area of the map over partially violating both. The top 33 solutions found by Situs all relate to the redundant positions of octamers linked by helical symmetry, as seen by peaks in the translational cross-correlation function (figure 3.31). Their rotational orientations differ slightly, and with the best solution the dyad axis (red arrow) is located just 1° from the x-y plane and on the radius of the helix, while the superhelical axis (blue arrow) is at 4° to the x-y plane and perpendicular to the radius. Importantly, the binary choice of dyad axis orientation can now be assigned with the H3-H4 tetramer located on the helical periphery. The crystallographic line group to which this helix belongs to can now be unambiguously assigned as $D_{11}S_{-4.36}$ - following the nomenclature of Caspar (Caspar and Makowski, 1981). D_{11} represents dihedral symmetry with an 11-fold rotational point group operation, while S_{σ} equals the screw symmetry, where the σ represents the relative rotation between successive rings calculated by the equation $2\pi / \Phi n$.

3.7.2 *Constrained 2D Docking*

The newly discovered orientation of the dyad axis meant that the docking problem could be reduced to a simple 2-dimensional search. This follows on from the radial projection of a helix resembling a 2D crystallographic plane group and therefore allowing only certain symmetry operations (Klug et al., 1958). Any 2-fold operator can only occur in the plane of the 2D lattice; thus when rolled back into a helix, the dihedral symmetry operator is constrained to lie on a radius perpendicular to the z-axis. Furthermore, the relative positions of subunits are also constrained in the docking, since they are known from the IHRSR procedure. Consequently, the only variables that remain are the in-plane rotation and translation (corresponding to a variation in helical diameter) about and along the dyad axis of the histone octamer.
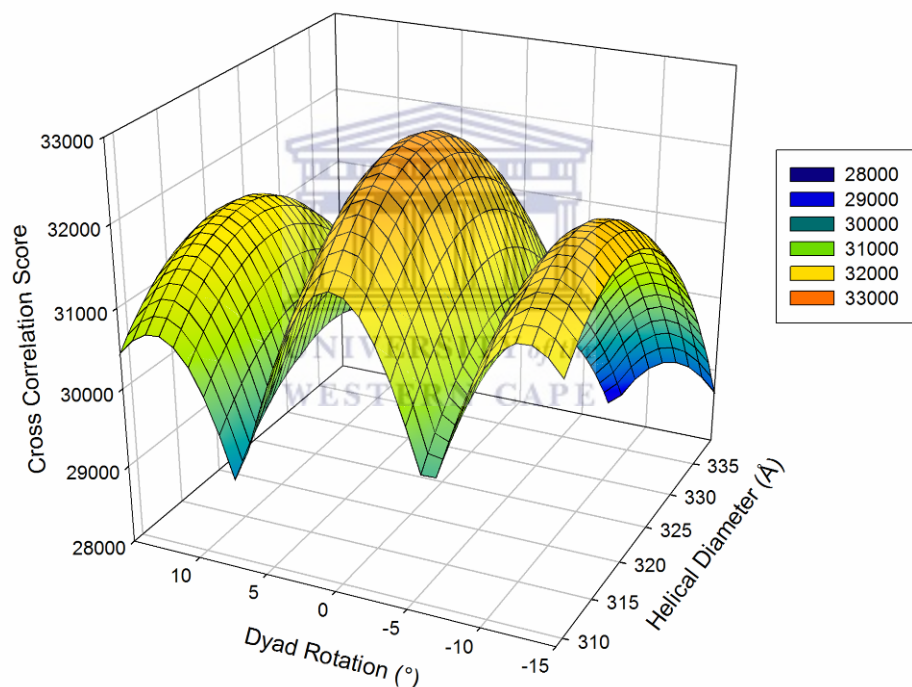
Figure 3.32 Two-Dimensional cross-correlation surface, produced by constrained docking. Notably the surface displays a peak at a helical diameter of 320 Å, and a dyad rotation of 0° - and is additionally mirror-symmetric about this point.

In the actual procedure used here, the atomic model was systematically rotated and translated about the dyad, generated into a helix of 6 rings (therefore accounting for all density), filtered to 16 Å, and cross-correlated with the EM map. In this manner a 2D cross-correlation surface was formed as a function of dyad rotation and helical diameter (figure 3.32). This procedure was iterated once after alignment in Φ and z was re-

optimized between the map and the atomic model helix, since this fit is not independent of the 2D dyad axis fit (resulting in a 1 Å decrease in helical diameter). The 2D correlation surface displays a clear maximum at a dyad rotation of 0°, at the position that the superhelical axis is perfectly aligned with the x-y plane. The surface is additionally symmetric about this position, meaning that the alignment of the superhelical axis forms a reflection point in correlation for rotations about the dyad axis. Mechanistically, because all the superhelical axes of octamers in a ring are now in the same plane and intersect each other, it means that the individual core histones within any given ring form a pseudo-continuous left-handed spiral (figure 3.33B). For these reasons, the superhelical axis can be regarded as an additional constraint, and the docking problem reduced to just one dimension. Of course this is simply a line profile of the previous 2D surface, with a peak at a helical diameter of 320 Å.
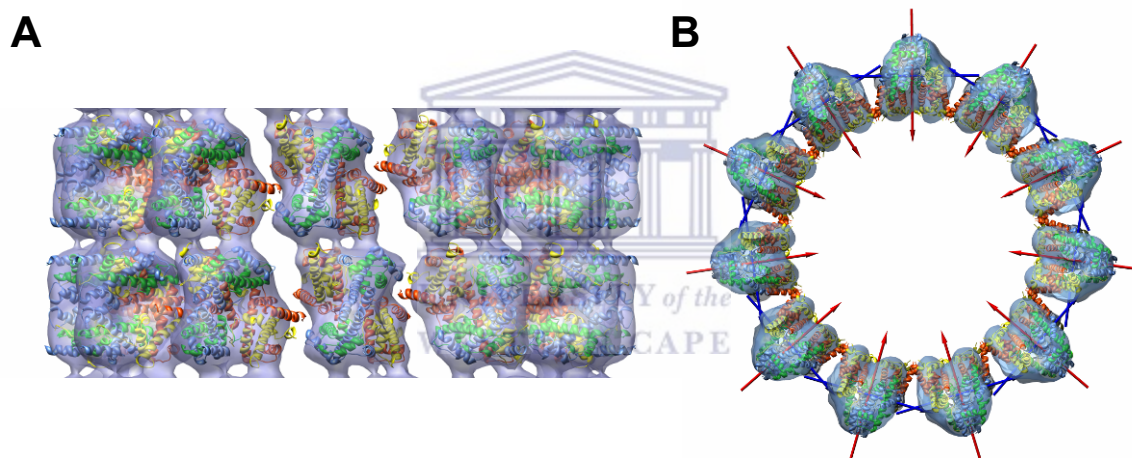
Figure 3.33 (A) Optimal 2D-docked result utilizing both helical and dyad axis constraints. This solution represents the peak in the 2D correlation surface shown in figure 3.32. The result is truncated to 2 rings for display. **(B)** Single docked ring seen from above, dyad axes (red) point toward the helical axis, while adjacent superhelical axes (blue) intersect each other to form a pseudo-continuous LH-protein spiral.

Overall this docking is similar to the Situs result, except the positions of the octamers are now, because of the constrained axes, more accurately placed relative to the map density at a slightly wider radius (figure 3.33A). Interactions within the helix can be broken down into two major components: *inter-ring*: vertical interactions between rings; and *intra-ring*: horizontal interactions within rings. Vertically, the generated helix produces a good staggered fit with closely paired surfaces. However, the intra-ring contacts are separated by large gaps, although their surfaces are still clearly complementary. Incorrect sampling

of the film could account for some of this discrepancy, but since the sampling would alter both horizontal and vertical interaction distances, it cannot account for a discrepancy in just a single direction. Instead, this can be attributed to artefactual flattening of the helix in negative stain as already diagnosed from the power spectra (figure 3.29). Because each class-average representing a view of the helix in projection is broadened, through backprojection the 3D volume itself is consequently widened in diameter. Ultimately using this broadened EM map through a cross-correlation based docking procedure, results in placing the atomic structure at too great a radius. For this reason the resolution of the map about Φ is also expected to be poorer, and also offers an explanation for the loss of density representing horizontal connectively in the EM map. Since flattening is confined only to x-z plane, it does not change any z-spacings, and hence resolution in this dimension is better and any vertical interaction distances can be regarded as accurate.

3.7.3 *Comparison of the reconstructed density with the atomic model helix*

For FSC comparison the best 2D cross-correlation fit was used, to compensate in part for the helix broadening. Visually the two helices agree well with each other, although the 16 Å value used to filter the model helix is perhaps slightly optimistic and this resolution is probably only achieved in the z-dimension. Well-conserved regions include: the overall symmetry created by the dyad axis; the propeller shape formed by the central tetramer, and the flanking 'hips' formed on either side by the two dimer pairs; and the two distinct regions of vertical connectivity between octamer rings. Inconsistencies in the EM map include: the loss of horizontal connecting density intra-ring; excessive protuberances on the inner diameter leading to a knobbly appearance; and slight differences in the outer surface of the octamer (seen in red in figure 3.34).
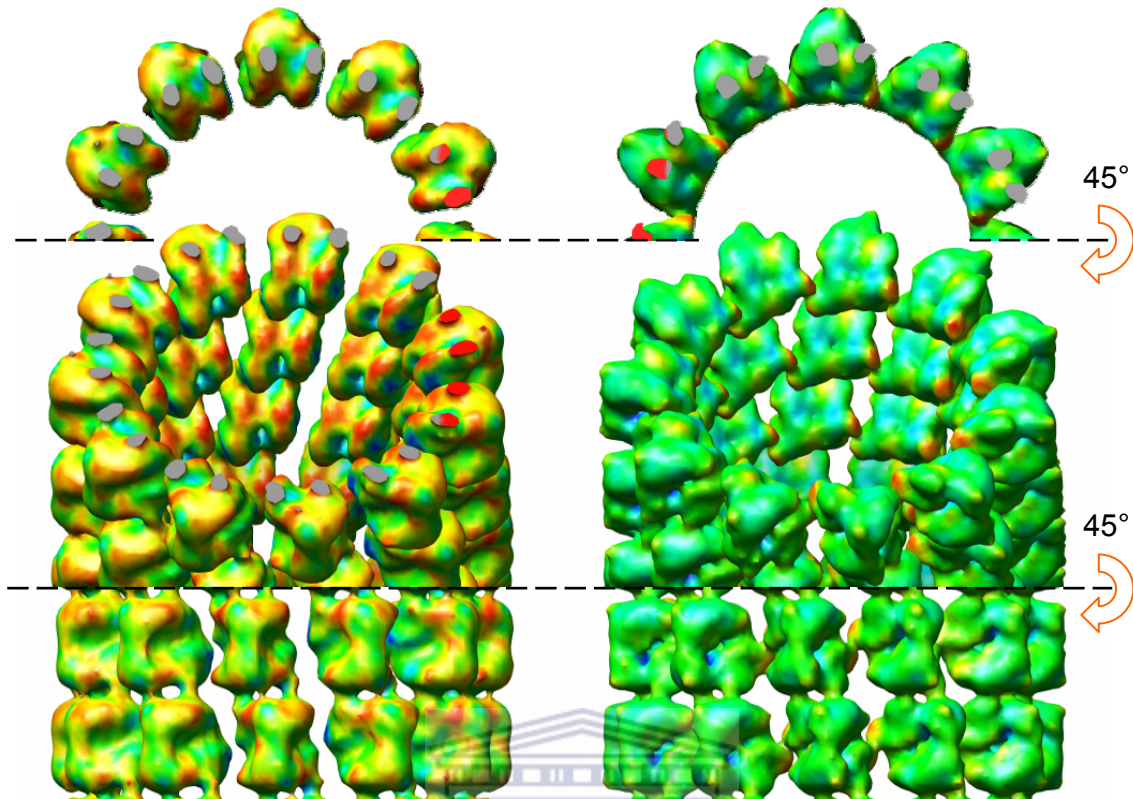
Figure 3.34 Visual comparison of the reconstructed histone octamer tube (left) vs. one generated from the pdb (2HIO) using the same helical parameters and docking results (right). Maps are colored by difference, blue being the most conserved and red the least so. The density generated from the atomic model helix was Gaussian filtered to 16 Å.

In order to quantify this agreement, these two volumes were overlaid by alignment in Φ and z and compared via cross-resolution FSC. Using an FSC threshold of 0.71, because the X-ray helix contains no noise, cross-resolution is estimated to be 27 Å (cutoff point is chosen where the FSC last crosses the 0.71 threshold). This poor resolution estimation although can be regarded as a limitation of comparing a negative stain cast structure, to that of a true coulomb potential.
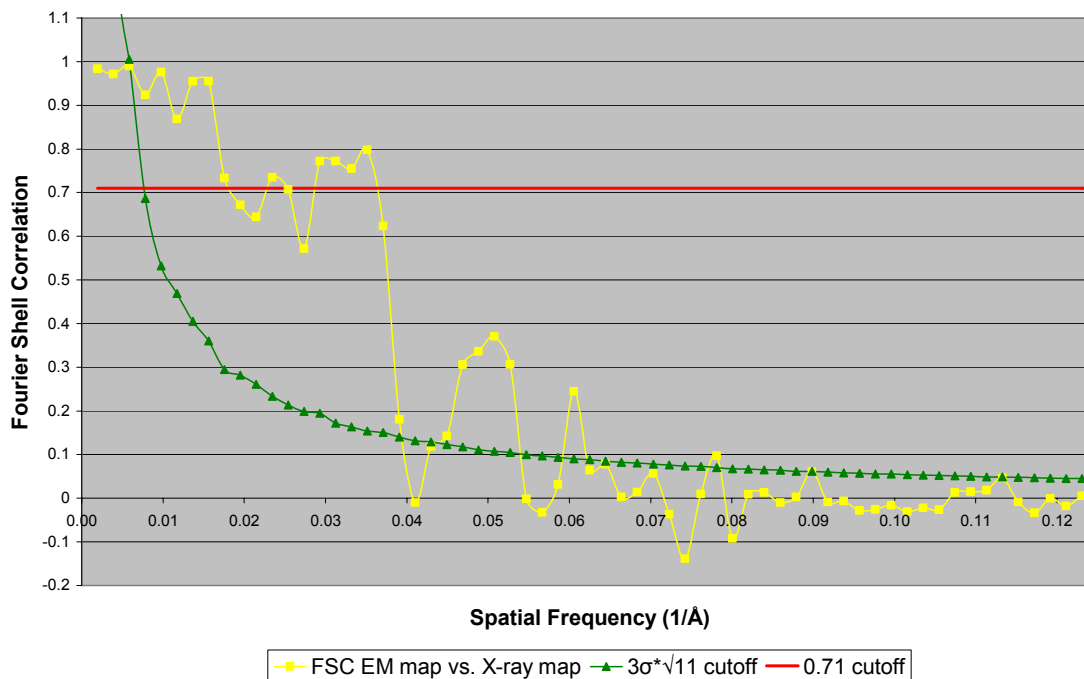
Figure 3.35 Cross-resolution estimation by Fourier shell correlation between the EM reconstruction and a model helix created from docking the octamer crystal structure (2HIO) and subsequent filtration to 16 Å. Resolution estimation by a 0.71 FSC criterion is 27 Å.

3.7.4 *Van der Waals Docking*

A clear potential mating surface could be seen in the intra-ring interactions involving H2B-αC helices across a 2-fold axis (red helix in figure 3.36A). The flattening artifact had been seen to boost the helical diameter artificially, and precluded complete optimization of the docking results in this dimension. The reconstructed EM map had served its purpose in the identification of the numerous constraints and vertical positioning. Correlation-based searching was discarded in favor of docking using the direct van der Waals (vdw) surfaces, which is simplified considering the search was reduced to just one dimension. The helical diameter was thus decreased empirically until the vdw surfaces first contacted, resulting in a helix with a refined diameter of 300 Å. Therefore the flattening had artificially increased the real-space diameter by 20 Å or approximately 7 % over its true value.
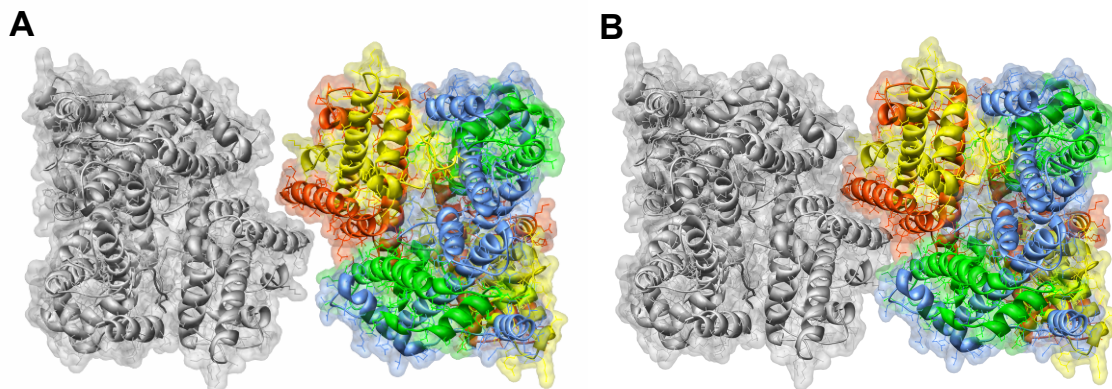
Figure 3.36. Complementary surfaces between neighboring octamers *intra-ring*, viewed down the 2-fold interaction axis. **(A)** Correlation-based docking result **(B)** Van der Waals docked surface. A 4-helix bundle is formed out symmetry-related pairs of H2B-α3 and αC helices across the 2-fold axis of interaction.

3.8 Octamer-Octamer Interactions

Interacting octamers are labeled separately by primed and unprimed symbols as used by Luger et. al. (1997), and because of the 2-fold axis across the interfaces, all interactions described occur twice by symmetry.
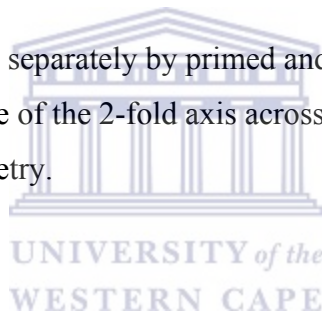
3.8.1 *Horizontal Interactions*

The horizontal interactions are solely responsible for the association of octamers into rings, and set the angle for the enclosed point group symmetry. The overall fit generated by the vdw docking of these adjacent octamers extends over a large interaction surface with dimensions of 29 Å × 19 Å, and constitutes a total buried surface area of 544 Å^2 inaccessible to solvent. Complementarily between the two surfaces can be seen by serial z-sections through the interface in figure 3.37(B-G). The interface is largely hydrophobic owing to a balance of 58 % non-polar to 42 % polar amino acids, which are individually tabulated below (table 3.1). Most side chains are donated by histone H2B, and are mainly threonine (32%) and lysine (31%) residues, followed to a lesser extent by serine (15%) and arginine (13%).

Segment	Residue	% Surface Area
H2A-αN	Ser19	1.26
H2A-L0	Gly22	0.68
H2A-α2	Tyr50	0.91
H2B-L2	Arg86	0.21
H2B-L2	Thr90	5.96
H2B-α3	Ser91	3.45
H2B-α3	Arg92	8.65
H2B-α3	Gln95	2.91
H2B-α3	Arg99	4.67
H2B-αC	Val111	0.65
H2B-αC	Thr115	9.83
H2B-αC	Lys116	7.03
H2B-αC	Val118	2.39
H2B-αC	Thr119	11.52
H2B-αC	Lys120	12.63
H2B-αC	Thr122	5.13
H2B-αC	Ser123	9.68
H2B-C-tail	Ser124	0.86
H4-L2	Lys77	11.57

Table 3.1 Table of interface residues, between adjacent octamers in the vdw docked octamer structure. Due to the 2-fold interaction axis, the same residues are involved on the second octamer. Results are generated from the protein-protein interaction server. (http://www.biochem.ucl.ac.uk/bsm/PP/server).

Another means of visualizing the interface, is via a dividing surface which bisects the two vdw surfaces, and is colored by the separation distance. This surface displays a distinctive red 'chicken foot shape' (figure 3.37A), which arises through the closest contact formed between the symmetry-related pairs of the distinctive V-shaped helices of H2B-α3 and αC - across the 2-fold axis.

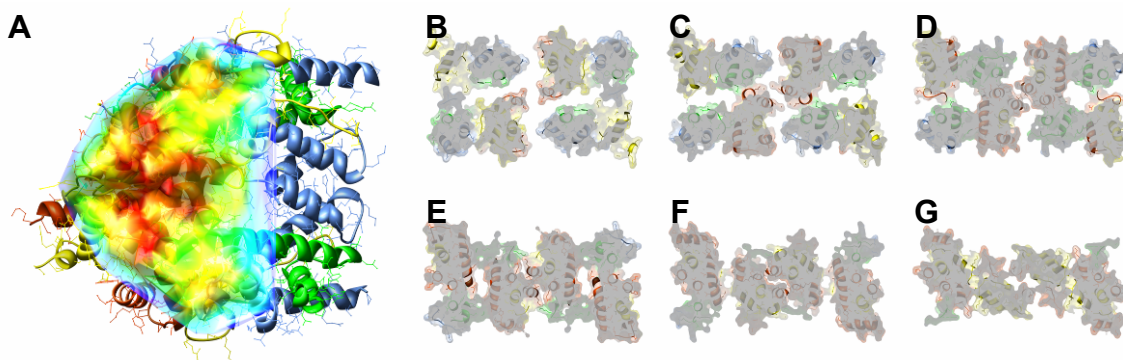**Figure 3.37** Horizontal interaction surfaces. **(A)** Interface dividing surface. Interface distances are labeled by color: red, yellow, green, and blue ≤ 4, 8, 12, and 16 Å respectively. Note the 'red chicken foot' shape formed through 2-fold related pairs of H2B-α3 and αC helices. **(C-H)** 5 Å thick serial sections of the interface shown in (A) sectioned by increasing depth down the 2-fold interaction axis

The major contributor to this quaternary interaction consequently comes from the histone H2B α3 and αC helices. The terminal H2B-αC helices intercalate with each other directly over the 2-fold interaction axis and interpenetrate the grooves created between H2A-H2B dimer and H3-H4 tetramer pairs. Because the two adjoining H2B-αC helices pack closely, the tertiary structural motif appears to be a ridges-in-grooves mode of interaction. Furthermore, this interaction also places the N-terminus of H2B-α3 close to the C-terminus of the opposite H2B'-αC helix, and hence this entire tertiary structural motif can be described as a 4-helix bundle (figure 3.36B). Other minor interactions include parts of H2A-αN; and a single lysine (K77), which manages to extend from H4 to the interface.

When the electrostatic potential is calculated for the surface of the histone octamer, it also follows a rough 2-fold trend about the H2B-αC helix. The center of this alpha-helix, along the 2-fold, is largely neutral, but above and below it opposite charge patches dominate (figure 3.38B). An acidic patch is formed above H2B-αC, in a hollow centered over H2A-α2 and flanked by the helices H2A-αN and αC. A basic patch, on the other hand, is formed below the H2B-αC, between this helix and H3-α1 in a groove located over H2B-α3. The residues making up the character of these two charged patches are tabulated below (table 3.2). Because of the 2-fold interaction symmetry, this brings these two oppositely charged regions into close opposition when octamers faces stack together (figure 3.38A). Although most of these residues are probably not in a position to form direct salt-links (especially D90, E91, and E92), they still provide a contribution to the combined electrostatic potential of the charge patch, and since these electric fields extend into space they can be expected to ultimately supply binding energy. The 2-fold nature of these charge potentials additionally ensures that only this particular orientation of octamer stacking is favorable, and possibly guides the assembly of octamers down a potential energy surface.
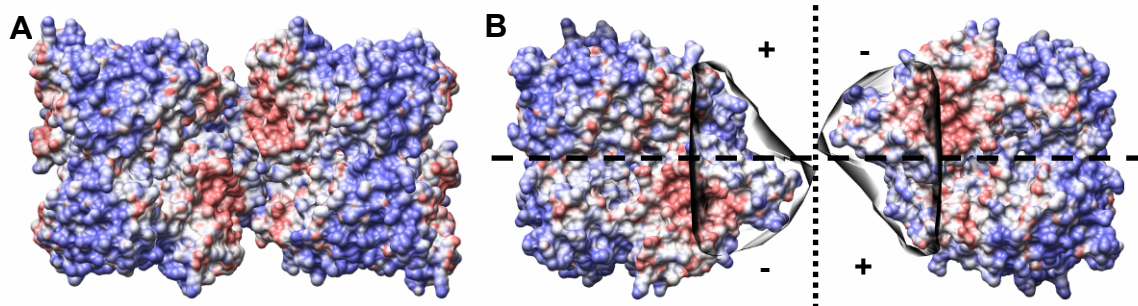
Figure 3.38 Surface-mapped electrostatic potentials. **(A)** Adjacent octamers seen down the 2-fold axis **(B)** Interface folded out by 90°. The interface dividing surface shown as a clear window, to aid identification of contacting regions. Oppositely charged patched are visible within this window, above and below the 2-fold axis (dashed horizontal line). The electrostatic potential ranges from +7.0 (blue) to -7.0 (red) kcal mol$^-$ e$^-$.

Acidic Charge Patch		Basic Charge Patch	
Segment	Residue	Segment	Residue
H2A-α2	Glu56	H2B-α3	Arg86
H2A-α2	Glu61	H2B-α3	Arg92
H2A-α2	Glu64	H2B-α3	Arg99
H2A-αC	Asp90*	H2B-αC	Lys108
H2A-αC	Glu91*	H3-L2	Lys79
H2A-αC	Glu92*	H4-α2	His75
H2B-αC	Glu105	H4-L2	Lys77
H2B-αC	Glu113	H4-L2	Lys79

Table 3.2 Table of positively- and negatively-charged residues, contributing to the electrostatics of the charged patches on the octamer face. *These residues are implicated in the H4 N-terminal tail-acidic patch interaction previously described (Luger et al., 1997).

3.8.2 *Energy Minimization*

A molecular mechanics simulation was performed in order to optimize the side chains for this new quaternary interaction between octamers, since these regions were previously engaged in very different contacts within the crystal lattice (Arents et al., 1991). In this way, the specific bonds which stabilize this particular positioning of octamers could be identified. Only the 4-helix bundle formed out of symmetric pairs of H2B-α3 and αC were included in the simulation, since this is the only proper structural motif formed. This modeling step is, of course, hypothetical and was performed as a final optimization step.

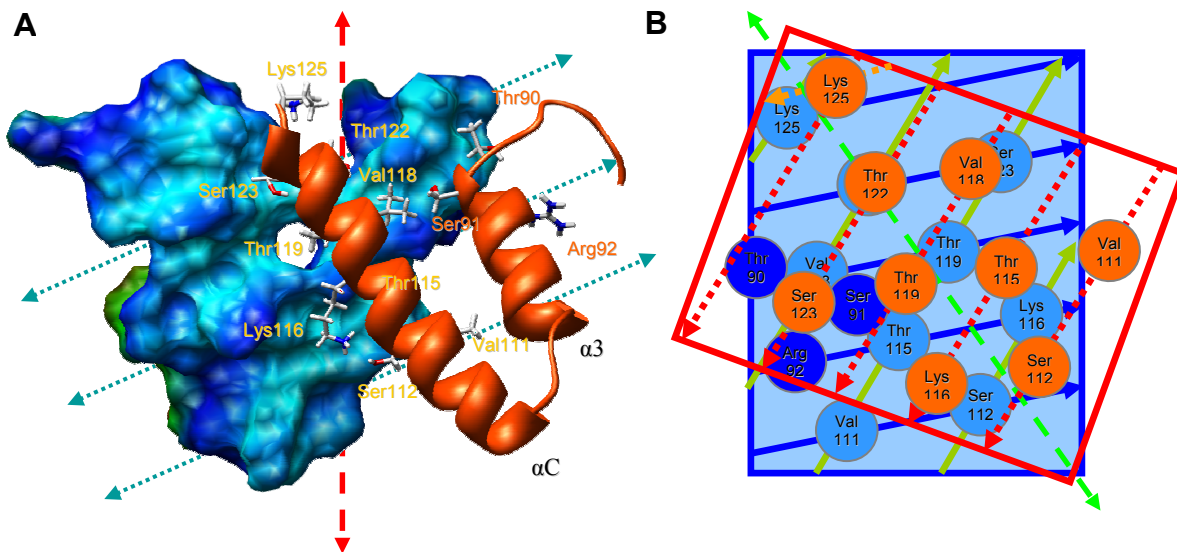**Figure 3.39 (A)** Optimized 4-helix bundle interaction. Molecular surface colored according to height, thus the i + 1n ridges appear dark blue and grooves light blue and additionally indicated by blue dashed lines. **(B)** Schematic of the 4-helix bundle shown in (A). αC helices are angled at 85° to each other, enabling the alignment of i + 1n and i + 4n helical grooves of both top (red) and bottom helices (blue), because a 2-fold axis passes through the center (green dashed arrow). This results in extensive interdigitated packing of residues between the chains of H2B and H2B'. Residues colored by a darker shade originate from the H2B'-α3 helix and additionally sandwich serine 123.

The results show that the majority of residues engage in extensive hydrophobic interactions via the intersection of ridges and grooves derived from both α3 and αC helices across the interface. Because of their characteristic V-shape, these two helices are approximately aligned and so the i + 1n ridges and grooves (the basic helix) can extend across both helices, forming a corrugated triangular surface (figure 3.39A). When the two αC helices from neighboring octamers then pack together at an angle of 85°, residues from the i + 4n ridges mutually interpenetrate into hydrophobic pockets created by these i + 1n grooves of the opposite helices and visa versa. This arrangement is a non-canonical helix-helix packing mode, and is a compromise in satisfying the alignment of i + 1n with i + 4n grooves from both helical surfaces simultaneously, as required by the 2-fold interaction axis. The most extensive zigzag intercalation occurs over this 2-fold axis, and is best illustrated by the schematic in figure 3.39B. Threonine and valine residues play significant roles here, and are intimately surrounded by 3 or 4 other hydrophobic residues. The most interesting interactions are those of threonine 119 and 122, whose methyl groups insert into pockets formed between both opposite α3 and αC helices; and thus cement together the entire 4-helix bundle.

A helical wheel plot has previously characterized the amphipathic nature of the H2B-αC helix, and it was expected to simply face the solvent with the hydrophilic face (Chantalat et al., 2003). Although part of the hydrophilic sector is engaged in the 4-helix bundle here, the dual nature of threonine, serine and lysine residues are well exploited in this environment, and effectively only the aliphatic portions are utilized. Thus, these residues have the ability to satisfy both constraints, and can rearrange according to their local environment. Still, exposure to solvent of the remaining more charged portion of the hydrophilic face is still maintained in this interaction.

Relatively few hydrogen-bonds are formed in this interaction, and then mainly at the periphery of the 4-helix bundle, and only to main-chain carbonyl groups. The terminal amides of arginine 86 (H2B-α3) are engaged in hydrogen-bonds with the main-chain carbonyl groups of both lysine 125 and serine 123 respectively (H2B'-αC); while the terminal hydroxyl group of serine 91 (H2B-α3) forms a hydrogen-bond to the main-chain carbonyl group of threonine 119 (H2B'-αC). Threonine 119, in turn, hydrogen-bonds through its hydroxyl group to the main chain carbonyl of threonine 115 (H2B-αC), and because this occurs close to the 2-fold axis, bonds form a parallelogram shape over the axis. Although not included in the simulation, lysine 77 from H4 could be manually identified to hydrogen bond to the main chain carbonyl group of serine 19 from H2A.

Donating segment	Residue	Receiving segment	Residue
H2B-α3	Arg86 NH1	H2B'-αC	Ser 123 O
H2B-α3	Arg86 NH2	H2B'-αC	Ser 123 O
H2B-α3	Arg86 NH2	H2B'-αC	Lys125 O
H2B-α3	Ser91 $O_\gamma H$	H2B'-α3	Thr119 O
H2B-αC	Thr115 O	H2B'-αC	Thr119 $O_\gamma H$
H4-L2	Lys77 N_ζ	H2A'-αN	Ser19 O

Table 3.3 Table of inter-octamer hydrogen-bonds achieved through geometric optimization.

3.8.3 *Vertical Interactions*

In contrast to the horizontal contacts, the vertical contacts form over two small surface areas (figure 3.40A), and are well-preserved features between the EM map and the docked octamer structure. Few residues are involved, and in fact, the only identifiable bond is a charge-dipole interaction between the terminal amide of lysine 74 of H2A-L3 (at the end of α2), and the main chain carbonyl group of leucine 60 of H3'-L1 (near the α1 helix).

Instead, the interaction is governed rather by clear dipole-dipole interactions between the extended α2-helix of H2B, and that of H3-α1 (figure 3.40B). The ends of these helices are aligned such that the carboxy-terminus of H2A-α2 is opposite the amino-terminus of H3-α1, and their oppositely-charged helix-dipoles are brought into close proximity. However, it is unclear exactly how the dipoles are passed along - whether by the main chain residues at ends of helices (N-H$^+$ and C=O$^-$) or by a terminal amino acid. Additionally, basic residues (K74, K75, and R77) at the C-terminus of H2B-α2 may quench the negative dipole effect somewhat; while at the N-terminus of H3-α1 similar residues (R63, K64) may help to strengthen its dipole potential. Seen in the electrostatic-mapped surface potential (figure 3.40B), some quenching has already weakened the negative dipole of H2B-α2.

Unlike the horizontal contacts, these octamers are offset by $\Delta\Phi$ over the 2-fold interaction axis. In the absence of DNA, this staggering is a way of successfully satisfying the electrostatics of the octamer surface usually engaged in DNA-binding (figure 3.40B). Ultimately, this satisfaction of electrostatics through dipole-dipole interactions is what sets the relative rotation between rings, and thus the pitch for the entire 11-start histone octamer helix.
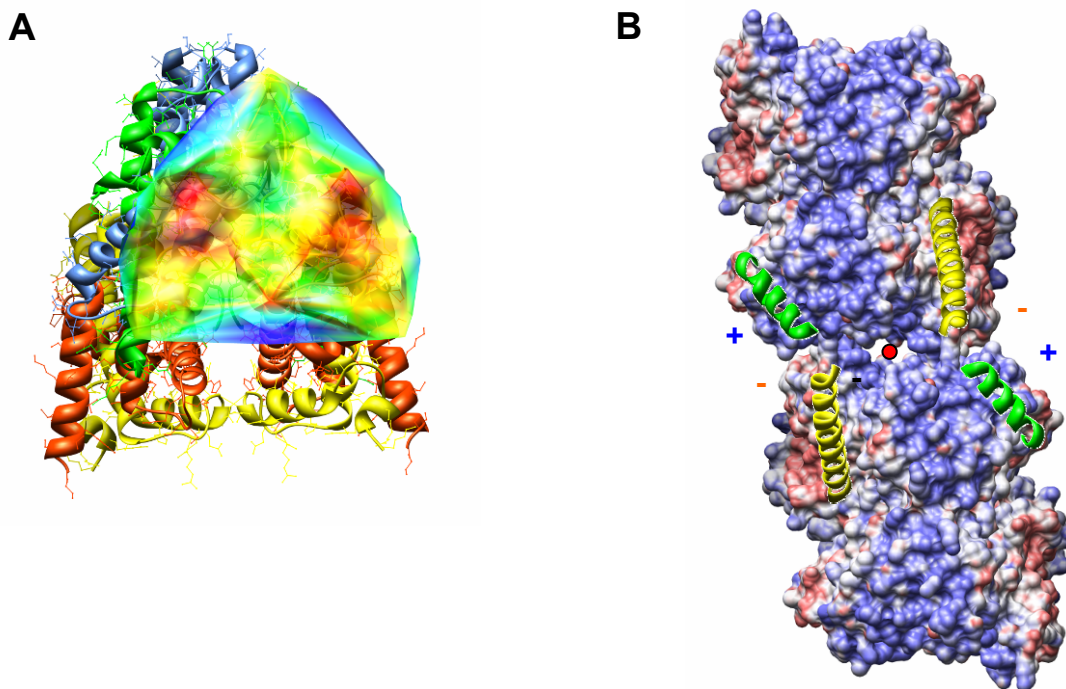
Figure 3.40 Vertical contacts. **(A)** Interface-dividing surface. Closest contact is formed at the two regions representing the dipole-dipole interaction (in red). Interface distances are labeled by color: red, yellow, green, and blue ≤ 4, 8, 12, and 16 Å respectively. **(B)** Surface-mapped electrostatic potential with overlaid ribbon models of the helices of H2A-α2 (yellow) and H3-α1 (green). (H3 is colored in green for contrast – against the usual coloring scheme.) 2-fold interaction axis is indicated by a red dot.

Although it may have be possible to optimize the z-spacing (and hence the sampling in Å/pixel) on the dipole-dipole distance alone, it was unclear how the actual dipole is passed on and to what distance. Feeding back of the adjusted sampling may have potentially improved the 2D correlation docking (figure 3.30); but since the horizontal contacts were finally optimized disregarding the EM map, there would be little use in doing this. Besides, the vertical contacts were *de novo*, in good agreement with features in the EM map.

CHAPTER 4: DISCUSSION

Histone octamers were successfully purified and crystallized into helical tubes, EM data were collected and the structure solved by iterative helical real-space reconstruction (IHRSR) to a resolution between 16-20 Å (allowing for anisotropy). The resulting density was characterized as an 11-start helix and the slower, more buffered crystal growth employed here is possibly responsible for the production of a homogenous set of this form of helix.

4.1 Resolving the Ambiguity

4.1.1 *Relative Ring Stagger*

A question arises as to the validity of the reconstruction of Klug et. al (1980), since the reconstructions attempted with imposed C_{10} symmetry in this study were shown to be artefactual, and yet these reconstructions still converge to the correct helical parameters of Klug et. al (1980). This is an especial concern since the two helix point groups are indistinguishable at the level of the power spectra, on which indexing is based in Fourier-Bessel methods unless cognizance is taken of the phases of the complex transforms of individual fibers. Fortunately, in this case, the IHRSR iteration parameters were powerful enough in identifying these artifacts, without consulting the actual reconstructed volume. Instead a C_{11} point group was shown to be definitely correct and the atomic structure of the histone octamer could be docked with high accuracy into this volume.

The main discrepancy in helical symmetry between these two structures is that the axial rotation value ($\Delta\Phi$) is different. If the two reconstructed helices are indeed different or the C_{10} symmetry is wrong in the case of Klug et. al. (1980), then different structural contacts should govern the helical construction or should produce steric clashes respectively - particularly between adjacent rings. The historical helix could be recreated in atomic detail using their helical parameters and a reduced diameter based on an assumed $^{10}/_{11}$ decrease in circumference ($2\pi r$). Overlaying the result in figure 4.1 reveals

the vertical interactions between octamers are virtually superimposable for a single stack and that the relative stagger between rings is practically the same.
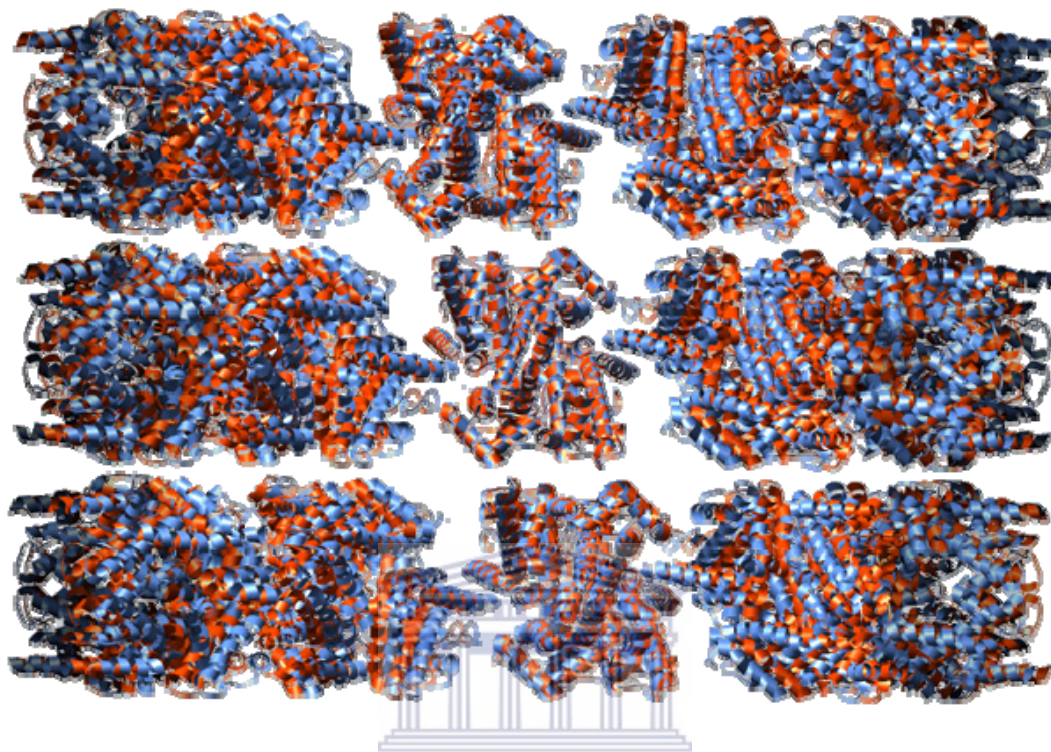
Figure 4.1 Atomic models for C_{10} (blue) and C_{11} (red) helices. The relative stagger between rings is the same for a single aligned stack of octamers and means that the same structural interactions govern the construction of these two helices. PDB coordinates are from 2HIO (Arents et al., 1991).

This may seem obvious since the relative rotation between rings is predicted to be the same according to the equation:

$$\Delta\Phi = 360° / (\text{point group} \times \text{relative rotation})$$

Where the relative rotation is the number of rings per turn (helical twist), and was found from the power spectra as the ratio of pitch (280 Å) over rise (65 Å). Based on this equation, the axial rotation ($\Delta\Phi$) therefore *must* then be different for the two helices to be similar and possess the same structural interactions between rings. This means that the reconstruction and the point group symmetry reported by Klug et. al. (1980) is essentially correct, and it is a common occurrence for labs to report differing point group symmetries for helices composed of the same biological molecule. The strongest undeniable evidence

for a C_{10} point group in their case had been the even phase agreement across the Fourier transform, seen for a few well-characterized helices used in their Fourier-Bessel reconstruction.

Now knowing the exact orientation of the histone octamer within both helices meant that the orientation originally used in creation of the start model was wrong by a factor of 90° (Section 3.3.1). This can be attributed to an anomalous feature in the reconstruction of Klug et. al. (1980) of a hole in the center of the octamer (figure 3.8), which was initially wrongly assumed in our study to represent the superhelical axis. This anomaly probably arose due to the radial uncertainty of Bessel terms, which is a known disadvantage of Fourier-Bessel reconstruction methods.

4.1.2 *Helical Nets and Cylindrical Projections*

Cylindrical projections were made for both the final C_{11} reconstruction and recreated atomic model of Klug et. al. helix respectively, by co-axial projection of their electron density perpendicular about the z-axis (figure 4.2). The result is a representation of the helices folded out into two dimensions and axial rotation is thus more conveniently represented as a translation. Finally, the cylindrical projections are overlaid with their respective helical net plots, which denote the lattice positions formed from the helical symmetry values of Δz and $\Delta\Phi$.

When the two cylindrical projections are scaled by octamer number, rather than by rotation angle, they are seen to represent the same 2D lattice (figure 4.2) (initially the scaling only differed because helices possess slightly different radii). This result demonstrates that these two helices are constructed out of the same 2D lattice, joined at different points when rolled into a helical tube to form their respective point groups. They are joined flush as opposed to at an angle, which is more common for helical assemblies, making the number of octamers per ring independent of the helical screw symmetry. The two helix types therefore share the same screw symmetry and can be described by the helical line groups of $D_{10}S_{-4.36}$ and $D_{11}S_{-4.36}$ respectively (Caspar and Makowski, 1981).

The differences between D_{10} and D_{11} intra-ring interactions are therefore expected to be minimal and logically the inter-ring contacts are also maintained.
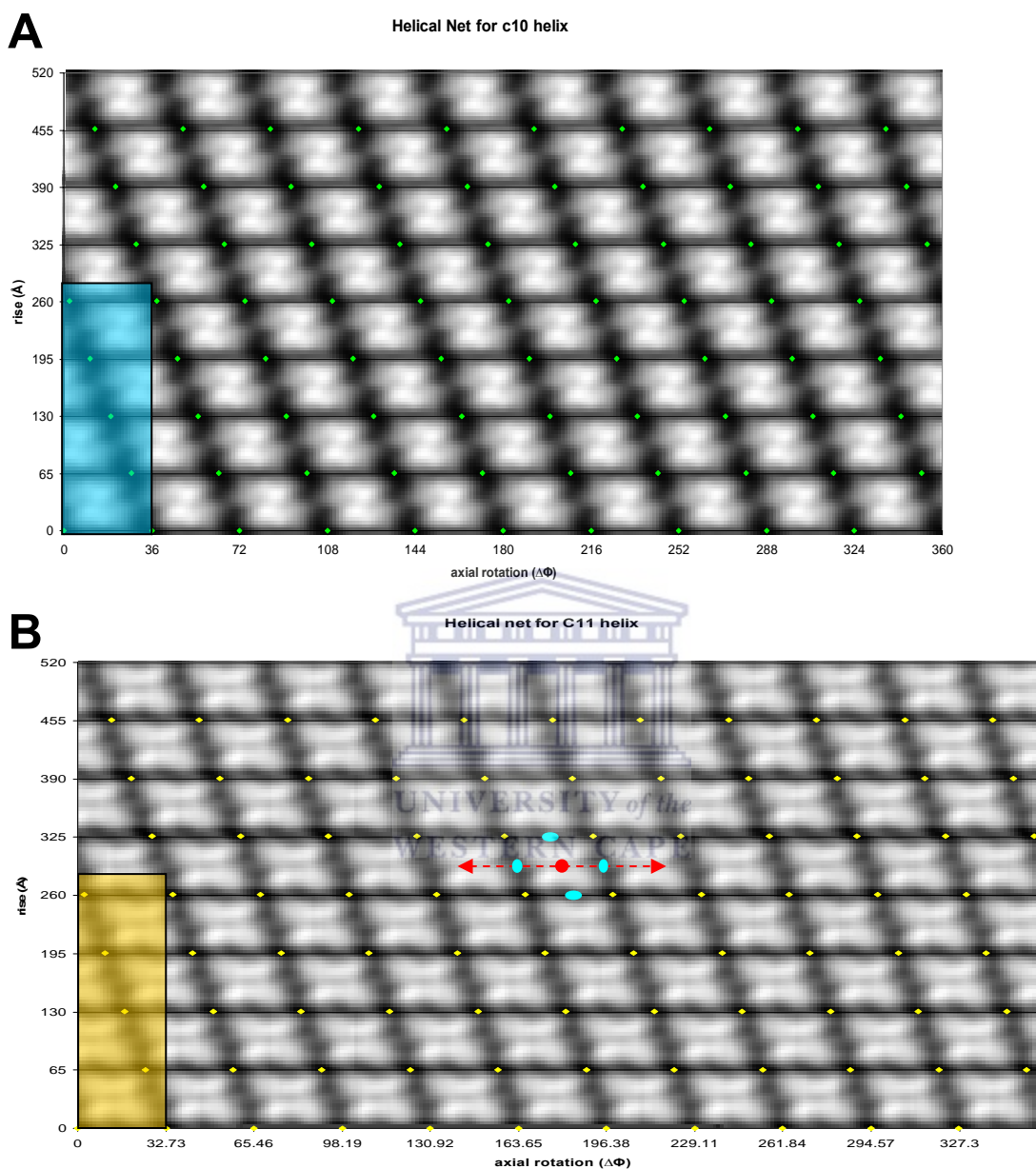
Figure 4.2 Helical nets overlaid onto the cylindrical projections for **(A)** Atomic model created from the helical parameters of Klug et. al. (1980) and with C_{10} point group symmetry **(B)** Actual reconstructed helix from this study with C_{11} symmetry. Differences in appearance are due to projecting negative stain cast (B) vs. a true coulomb potential (A). The 280 Å pitch of the helix is identical between helices and is indicated by the colored block. Lattice positions are labeled green in (A) and yellow in (B). Additional degenerate 2-fold axes are labeled in blue (B). The dyad axis projects out the plane and is labeled as a red dot, and the superhelical axis is shown as a red horizontal line (B).

In this representation the helical line group resembles the crystallographic plane group *p*2, and similarly shares its symmetry elements. The basic unit cell is oblique and there

are five 2-fold rotational axes, one at each corner of the unit cell and one within the repeating motif of the octamer itself. It is clear now why the dyad axis of the octamer is constrained to lie in-plane in order to be effectively described by the helical line group, and is responsible for the formation of the dihedral symmetry of the point group as opposed to more degenerate cyclic symmetry. Additional degenerate symmetry elements are also generated between the lattice positions, and describe the 2-fold axes which occur between octamer interactions (labeled in blue on figure 4.2B). Furthermore, the superhelical axis projects horizontally, and passes through these 2-fold elements lying between horizontally-opposed octamers, and thus preserves the screw axis describing the pseudo-continuous left-handed spiraling of histone domains within octamer rings.

4.1.3 *Computed Diffraction Explains the Ambiguity*

When a computed power spectrum of the C_{11} IHRSR is overlaid with an *n,l* plot constructed using the selection rule of Klug et al. (1980), the approximate positions of the principle maxima from the top surface of the helical lattice are predicted (figure 4.3). Since the power spectra for C_{10} and C_{11} were already shown to be the same in section 3.4.4, it means that this *n,l* plot equally explains the diffraction from both forms of helix, and arises because the specific point group is encoded separately from the selection rule. This relates back to the two helices belonging to the same 2D lattice and the *n,l* plot is simply an approximation to diffraction from a completely flattened lattice (Moody, 1967). This explains the ambiguity in the power spectra completely, and represents a unique situation wherein the IHRSR 'gold standard' of the observed power spectrum fails to recognize an artefactual reconstruction (Wang et al., 2006).
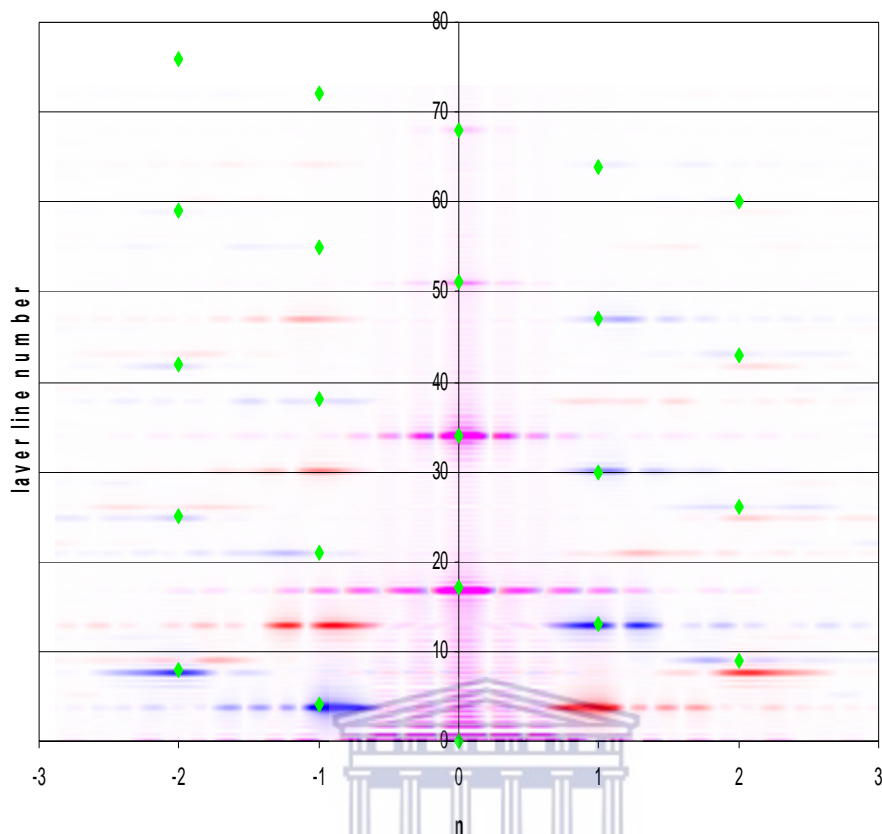
Figure 4.3 Computed diffraction from C_{11} IHRSR with superimposed n,l plot constructed from the selection rule $l = 4n' + 17m$ of Klug et. al. (1980). The real order n is equal to 10/11 times n', since the same selection rule approximately describes both helices and is independent of the point group symmetry. The n,l plot predicts well the positions of the principle maxima in the diffraction arising from the upper surface of the helix (blue).

4.1.4 *Refined Indexing*

The study by Klug et al. (1980) had reported a wide variance in the relative rotation value measured for individual helices (4.0 – 4.4). A value of 4.25 ($^{17}/_4$) was settled upon since it was the simplest rational approximation for a selection rule. The IHRSR method has the advantage that the helical symmetry is better determined, since it is calculated via a least-squares fit to a real-space backprojected volume derived from thousands of images, leading to an accurate value for the relative rotation of 4.36 rings/turn. Since it was demonstrated these two helices are derived from the same 2D lattice, the relative rotation should ideally be identical. The discrepancy can be attributed to inaccurate measurements from the power spectra and the rounding off this value too heavily in order to create a

rational fraction, affecting the final selection rule. This in turn affects the accuracy of assigning layer line spacings, and concomitant helical symmetry allocated to the various Bessel functions. In real-space this results in smearing, caused by summating helical waves at imprecise helical symmetries. Instead the improved ratio of 4.36 can be better described by constructing a new selection rule:

$$l = 25n + 109m$$

This refined selection rule was constructed purely to highlight a primary limitation of Fourier-Bessel methods, in that indexing is always an approximation based on user-made decisions such as the arbitrary assignment of a helical repeat value. An exact helical repeat does not exist for helices though, since there is nothing to maintain such long range order. The real-space approach of IHRSR is superior, because helical parameters are more accurately determined and are not dependant on the rationalization of layer spacings to simple fractions. This refined selection rule only represents a small change over the old one and consequently does not affect any previous deductions, although it does predict the diffraction in figure 4.3 better.

4.1.6 C_{10} *Reconstructions*

Having explained the construction principles behind the C_{10} and C_{11} form of helices, we can take a retrospective look at the C_{10} reconstructions, and particularly at the failings of the multi-model IHRSR sorting algorithm.

Because C_{10} and C_{11} helices are comprised of the same 2D lattice, they also share similar real-space information in projection – particularly the 280 Å pitch. Then through backprojection with the incorrect 10-fold symmetry applied, it results in a volume that when searched for underlying helical symmetry, converges upon the parameters of Klug et al. (1980) - this being the only helical symmetry consistent with this point group which preserves the real-space information of the pitch. Thus convergence can be achieved through a quasi self-consistency between images and reprojections. Additionally, in this case, it explains how the correct 2D lattice can be preserved even though the repeating

motif contained within is artefactual, and therefore still produces the correct power spectrum.

In case of a mixture of point groups, a multi-model IHRSR procedure was employed to separate the helix types and reconstruct them simultaneously. There was already strong evidence against mixture, including: no mirror plane represented in the eigenimages; no phase agreement between symmetric peaks within the Fourier-transforms of eight extended helices; and finally, like the previous independent C_{10} reconstructions, the multi-model C_{10} volume was clearly artefactual. Additionally, if a mixture did indeed exist, the C_{10} form would definitely be expected in the minority. Therefore, it was of especial concern why the majority of images were artefactually assigned to C_{10} classes, and the sorting was seen to fail since C_{10}/C_{11} image classes were not correlated in blocks - this failing in itself evidence against a mixture.

The explanation offered is that discrimination by cross-correlation between C_{10} and C_{11} forms, in this case, was not sensitive enough to sort the images. This is because they both represent the same 2D lattice, compounded by the fact that flattening and partial staining further decrease the discrimination between these helix types. Failure of the C.C. sorting most likely relates to normalization issues, wherein all images are initially treated to represent the same integral projected density. However, when images of C_{11} helices are backprojected with C_{10} symmetry, they result in a more concentrated density and brighter helical walls in reprojection (compare figure 3.17 B & D), in turn attracting more images to that class in successive iteration rounds. The cross-correlation is therefore mainly based on alignments to the helical periphery and the bimodal x-shift peaks at +/- 3 pixels validate this suspicion (figure 3.16A).

A suggestion for a similar situation in the future is to perform a prior characterization of individual helices, based on the phases of their Fourier transforms, before segmentation into short boxes.

4.2 Biological Insights

The major results from this study differ from those of Klug et. al (1980), since the X-ray structure for the histone octamer has since been solved and it is now possible to use a hybrid approach to discover novel inter-octamer helical contacts in atomic detail.

4.2.1 *H3-H2A Dipoles*

The inter-ring contacts discovered are staggered and comprise the 2-fold related helix-dipole interactions between the opposite helices of H2B-α2, and H3-α1. These interactions, however, would be sterically occluded by DNA within the nucleosome and bear no direct physiological relevance to the chromatin fiber (figure 4.4). This illustrates that *in vivo* only the extended tail domains can pass between adjacent gyres of DNA on this surface of the nucleosome in order to mediate such interactions between successive helical turns of the chromatin fiber. Here these interactions are simply electrostatically driven as way of satisfying the powerful helix dipoles within the context of crystal tubes in the absence of DNA. It might be speculated however, that a similar mechanism might be used to quench these dipoles by histone chaperones and assembly factors during nucleosome assembly/disassembly.
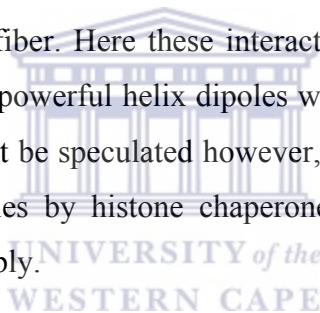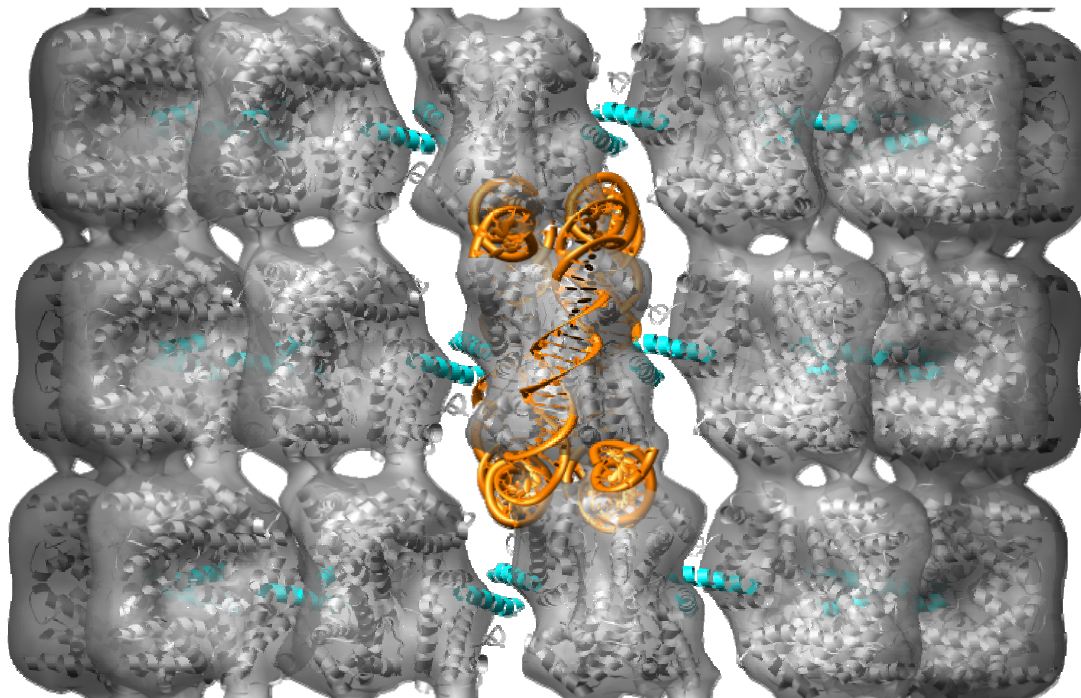
Figure 4.4 Steric occlusion of vertical interactions by DNA. Unoccluded H2B-αC helices of the horizontal contacts are indicated in blue. PDB coordinates for octamers are from 2HIO (Arents et al., 1991) and for DNA are from 1AOI (Luger et al., 1997).

4.2.2 H2B-α3/αC Helix Interactions

Following the final C_{11} reconstruction, a hybrid docking approach involving a 2D correlation search and vdw-docking procedure was used to produce a pseudo-atomic model of the histone octamer helix. A final energy minimization was carried out when it was realized that the vdw-docking had placed adjacent octamers close enough for a ridges-in-grooves interaction between their respective H2B-αC helices. This led to discovery of the complete structural motif of a 4-helix bundle, formed between 2-fold symmetric pairs of H2B-α3 and αC helices.

4.2.2.1 *Relevance of the Horizontal Contacts*

The horizontal H2B-α3/αC interaction discovered in this study is deemed more relevant than the various 3D crystal contacts discussed in the literature review (section 1.2.8) for numerous reasons. Firstly, since the contacts are only constrained by the helical line group, octamers are not forced into a distinct 3D packing arrangement, thereby repressing what seems to be an inherent helix forming propensity - as these octamer tubes were

- 112 -

originally formed in attempt to produce 3D crystals (Klug et al., 1980). This might mirror an *in vivo* mechanism used in the formation of the 30nm chromatin fiber. Secondly, the dyads of adjacent octamers are now orientated on the same side, meaning this arrangement is more compatible with 30 nm chromatin fiber models. Thirdly, the interaction itself is a reasonable one, involving the well-defined structural motif of a 4-helix bundle, wherein stability is afforded by the intercalation of a high number of residues via a ridges-in-grooves model. Presumably the angle set by this ridges-in-grooves model is also incompatible with any 3D space group, and would hence prohibit it from ever being visualized by an X-ray structure. Finally, although the octamers were crystallized alone, the horizontal contacts formed, occur between the would-be octamer faces of the nucleosome and are not obscured by DNA (figure 4.4).

It should be noted that the interaction from this study does, however, share some common features with the crystal structures of the yeast-NCP (White et al., 2001) and the tetranucleosome (Schalch et al., 2005). The stacking modes are similar, in that adjacent nucleosomes are flush and superhelical axes are close to juxtaposition, although not perfectly so, as required to preserve the 2-fold symmetric H2B-α3/αC 4-helix bundle seen in this study. As a consequence, although all structures make use of the H2B-αC accessory helix to mediate their contacts, they do so via differing mechanisms. The low resolution (9 Å) of the tetranucleosome, however, prevented visualization of the actual contacts made.

4.2.2.2 *Inverted Solenoid Model*

Generally, there has been a drive to discover novel nucleosome-nucleosome interactions, since they would provide strong constraints for the modeling of the 30 nm chromatin fiber. However, as discussed (section 1.2.8), the arrangement of nucleosomes within a 3D crystal is probably not reflective of compact chromatin. Still, it is predicted from force measurements of chromatin (Cui and Bustamante, 2000), that a compact nucleofilament should contain these elusive short-range interfacial contacts which dock adjacent nucleosomes. There is minimal evidence in this regard so far, and the H4 tail being flexible, only serves as a weak constraint. The H2B-αC helix is not involved in DNA-

binding or octamer assembly, and currently possesses an unassigned function in chromatin, and is therefore an excellent candidate to carry out these nucleosome-nucleosome interactions.

The internucleosomal 4-helix bundle shown here supports, *prima facie*, both 2-start and solenoid modes of flush nucleosome stacking. Within the 4-helix bundle, the interaction angle is presumably fixed at 85° by the ridges-in-grooves contained within, and has perhaps evolved as a mechanism of restraining internucleosomal angles. It appears rational then to preserve this interaction, and use it as an added modeling constraint when building a 30 nm fiber model. This angle of the interaction is, however, seemingly incompatible with current fiber models, since it places the central dyad base pair of the nucleosome facing outwards and relates to the packing of octamers into rings which are opposite to those expected from wedge-shaped molecules. This would set incorrect entry/exit angles for nucleosomal DNA, and is only resolved though placement of linker DNA in a new conformation on the outside of fiber. Therefore, direct transplantation of this interaction into the context of the 30 nm chromatin fiber, produces a new inverted solenoid model (see figure 4.5 for details).
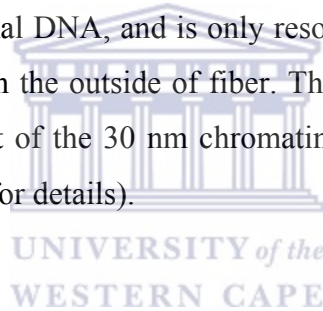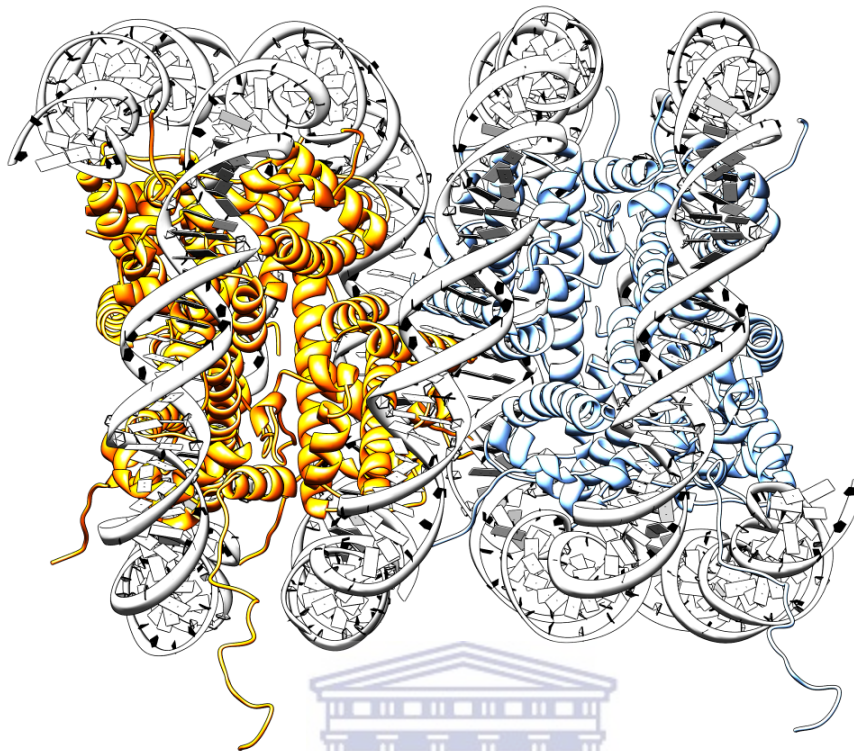
Figure 4.5 Proposed inverted solenoid model. A two nucleosome segment is viewed from the outside of the would-be fiber. Nucleosomes are linked via the H2B α3/αC 4-helix bundle discovered in this study and the central base pairs of the pseudo-dyad face the exterior. Approximately 30 bp of DNA links adjacent nucleosomes on the outside of the fiber, and continues the superhelical trajectory induced by the supercoiling of DNA around the histone octamer. Atomic coordinates used are from 1AOI (Luger et al., 1997).

The new model possesses numerous beneficial features which represent an improvement over the classical solenoid model (Finch and Klug, 1976). This previous solenoid model places the central dyad base pair on inside of the fiber where the entry/exit strands of adjacent nucleosomes are in close proximity. As a result, only very short stretches of linker DNA can be accommodated between neighboring nucleosomes, otherwise an unreasonably sharp bending would be required and would also violate the continuous supercoiling of DNA suggested by this model. Nonetheless, there has never been direct structural evidence to constrain the linker DNA conformation within a solenoid model. Instead, it makes more sense to place the pseudo-dyad of the nucleosome facing outward from the fiber, where there is more space for the accommodation of linker DNA. In this

way, a larger stretch of linker DNA connects nucleosomes, and exceeding this length, looping is uninhibited at the exterior of the fiber.

It was previously suggested, based on the importance of 4-helix bundles in histone octamer construction, that the remaining α3 helix of H2A might extend these rules into high order structure through the formation of a final internucleosomal 4-helix bundle (Luger et al., 1997). Here a similar idea applies, but rather through the terminal α3/αC helices of H2B, and is more rational since this extension is more peripherally located, and also does not require a rotation about the superhelical axis in order to bring these helices into contact (preserving the dyad orientations). The alignment of superhelical axes in the formation of nucleosome stacks, results in the continuous spiraling of histone dimers (analogous to a coiled-coil structure, where the z-axis is not straight, but helical). Within the inverted solenoid model then, the continuous protein and DNA superhelices mirror each other, and suggest the driving force for high order structure assembly is actually dictated by the underlying protein-protein interactions. Thus, the driving force behind octamer tube assembly is the same as in the postulated 30 nm fiber. Additionally, unlike the in 2-start model where DNA twist sets the angle between nucleosomes and the linker length sets the helical diameter, here the protein fixes both internucleosomal angle and helical diameter.

Since the angle set is the same as within the octamer rings at 10/11 nucleosomes per turn, a similar value is predicted for the number of nucleosomes per turn of the chromatin fiber, and agrees well with recent EM measurements of nucleosome density per unit length (Robinson et al., 2006). A greater angle, however, would produce irreconcilable steric clashes between nucleosomal DNA within the interior of this inverted solenoid model, and thus make it incompatible with a more generally-accepted value of 6 nucleosomes per turn (Gerchman and Ramakrishnan, 1987). However, this would be true of any solenoid attempting to include such an interfacial angle. In fact, this is a primary limitation of the solenoid model and since there has been no attempt to model the linker DNA it has thus far escaped controversy. Therefore, no model of the 30 nm fiber should

be considered seriously if it excludes the all important linker DNA trajectory, like that of Robinson et. al. (2006).

Instead, the H2B-$\alpha3/\alpha C$ 4-helix internucleosomal 4-helix bundle should rather be considered within the context of a 2-start zigzag model, especially in light of the recent evidence for this model (discussed in section 1.3.6). But, it requires that this interaction hinges in the opposite direction - the feasibility of which remains to be modeled. Some flexibility is already evident though the difference in the C_{10} vs. C_{11} interaction angles (3.28°), although a much larger angle would be required. The various classes of ridges-in-grooves interactions can occur at different angles and a crossing between H2B-αC helices at a more conventional 50° could support this notion. A large surface area is potentially gained by hinging into a more wedge-shaped packing mode and the electrostatics also supports this scheme by providing complementary charge patches further into the interface (figure 3.38B).

Removal of the supercoiled linker DNA, through the use of straight linkers, relieves these steric clashes encountered in the classical solenoid model, and the angle can be henceforth increased to accommodate a value closer to 6 nucleosomes per turn. The supercoiling of histone subunits is still preserved as the diving force in this model, but is however not mirrored by the DNA. The expected linear dependence of fiber diameter on linker length often quoted as evidence against the various 2-start chromatin models (Robinson and Rhodes, 2006). Here specific protein-protein interactions could resolve this conflict, by instead setting the angle between adjacent nucleosomes and thereby limiting the fiber diameter. Variability of linker DNA length could then be accommodated within the fiber interior by changes in the entry/exit angles (Bednar et al., 1995).

4.2.2.3 An Envisaged Folding Pathway

If the interaction from this study does indeed occur in native chromatin then it predicts certain pathways for the assembly of the chromatin fiber. Fibers are seen to exist in equilibrium between three distinct states characterized by sedimentation rates as the 29S

extended; 40S moderately-folded; and 55S maximally-folded forms (Hansen, 2002). What is proposed here is that the final bistability that occurs between the loose and compact chromatin conformations, arises through an interplay between histone tails and octamer faces in what is suggestively-termed here as the Tail & Face model.

The N-terminal tails act as the primary determinant of the moderately-folded conformation and operate mainly through the charge neutralization of DNA, allowing the close approach of nucleosomes; and specific long-range protein-protein interactions - like the H4 charge-patch interaction. Nucleosome locations are hereby loosely constrained as a tethered helical chain. This places adjacent nucleosomes within the radius of convergence, whereby the final docking of nucleosome faces is thermodynamically steered by the 2-fold complementary electrostatic lobes (figure 3.38B) and additionally ensures that only this specific orientation is favorable. Therefore the same interaction with dyads reversed is not expected to be relevant i.e. in the intercalation of nucleosomes between neighboring 30 nm fibers. Importantly, the role of the H4-charge patch or any of the tails is not mutually exclusive, nor expected to be occluded by the interfacial interaction within a fully-compacted fiber.
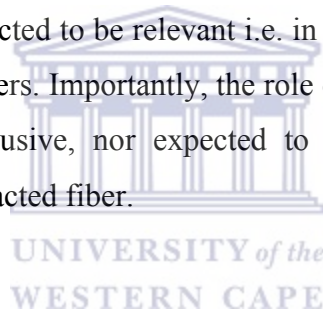
4.3 Future Work

A possible future goal could include improving the preservation and resolution of this study through use of cryo-negative staining and data collection on a FEG-TEM. This would be done with the express aim of better visualizing the positions of the histone tail interactions, since structural information pertaining to the tail domains is currently limited. The salt concentrations which stabilize the histone octamer tubes would prohibit regular cryoEM (as was attempted).

The modeling of the 30 nm chromatin fiber using this newly discovered H2B-α3/αC interface needs to be extended into all variations of solenoid and 2-start models, while possibly trying to include the H4-charge-patch interaction. This new constraint would make modeling the chromatin fiber easier than previous attempts, and can be considered

in atomic detail. A static constraint is also better than a flexible tail constraint and probably more directly applicable to a fully-compacted chromatin structure. A search is definitely needed for additional helical constraints particularly from newer defined 'Widom 601' arrays, to provide accurate helical parameters to further constrain model building.

A possible new function has been assigned to the H2B-αC 'accessory' helix which needs biochemical confirmation. An obvious choice would be to introduce disulfide cross-links within a compact defined nucleosome array, in an analogous technique as done to verify the H4-charge patch interaction (Dorigo et al., 2004). Cross-links stabilizing the array would verify the proximity of such-predicted residues and additionally confirm the startedness of helix that the H2B-α3/C interaction preserves.

Sequence conservation is expected from important residues within the H2B-α3/αC 4-helix bundle, including threonines 115,119,122 and valine 111 which intimately stabilize the bundle. There also are important predictions that the histone code presents and targeted modifications to this region may help disrupt the chromatin fiber in preparation for transcription. Currently, most known post-translational modifications to this H2B-αC region possess unassigned functions (Arg86, Arg92, Arg99, Lys116, and Lys120).

REFERENCES

Adkins,N.L., Watts,M., and Georgel,P.T. (2004). To the 30-nm chromatin fiber and beyond. Biochim. Biophys. Acta *1677*, 12-23.

Amos,L.A. and Klug,A. (1975). Three-dimensional image reconstructions of the contractile tail of T4 bacteriophage. J. Mol. Biol. *99*, 51-64.

Angelov,D., Vitolo,J.M., Mutskov,V., Dimitrov,S., and Hayes,J.J. (2001). Preferential interaction of the core histone tail domains with linker DNA. Proc. Natl. Acad. Sci. U. S. A *98*, 6599-6604.

Arents,G., Burlingame,R.W., Wang,B.C., Love,W.E., and Moudrianakis,E.N. (1991). The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. Proc. Natl. Acad. Sci. U. S. A *88*, 10148-10152.

Ausubel,F., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J., and Struhl,K. (2002). Short Protocols in Molecular Biology. (New York: John Wiley & Sons).

Baker,N.A., Sept,D., Joseph,S., Holst,M.J., and McCammon,J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. U. S. A *98*, 10037-10041.

Bang,E., Lee,C.H., Yoon,J.B., Lee,D.W., and Lee,W. (2001). Solution structures of the N-terminal domain of histone H4. J. Pept. Res. *58*, 389-398.

Bednar,J., Horowitz,R.A., Dubochet,J., and Woodcock,C.L. (1995). Chromatin conformation and salt-induced compaction: three-dimensional structural information from cryoelectron microscopy. J. Cell Biol. *131*, 1365-1376.

Bharath,M.M., Chandra,N.R., and Rao,M.R. (2003). Molecular modeling of the chromatosome particle. Nucleic Acids Res. *31*, 4264-4274.

Bohm,L. and Crane-Robinson,C. (1984). Proteases as structural probes for chromatin: the domain structure of histones. Biosci. Rep. *4*, 365-386.

Bozzola,J.J. (1998). The Champagne Artifact in Negative Staining TEM. Microscopy Today *98*, 24-25.

Bustamante,C., Zuccheri,G., Leuba,S.H., Yang,G., and Samori,B. (1997). Visualization and analysis of chromatin by scanning force microscopy. Methods *12*, 73-83.

Butler,P.J. and Thomas,J.O. (1980). Changes in chromatin folding in solution. J. Mol. Biol. *140*, 505-529.

Butler,P.J. and Thomas,J.O. (1998). Dinucleosomes show compaction by ionic strength, consistent with bending of linker DNA. J. Mol. Biol. *281*, 401-407.

Case,D.A., Cheatham,T.E., III, Darden,T., Gohlke,H., Luo,R., Merz,K.M., Jr., Onufriev,A., Simmerling,C., Wang,B., and Woods,R.J. (2005). The Amber biomolecular simulation programs. J. Comput. Chem. *26*, 1668-1688.

Caspar,D.L. and Makowski,L. (1981). The symmetries of filamentous phage particles. J. Mol. Biol. *145*, 611-617.

Chantalat,L., Nicholson,J.M., Lambert,S.J., Reid,A.J., Donovan,M.J., Reynolds,C.D., Wood,C.M., and Baldwin,J.P. (2003). Structure of the histone-core octamer in KCl/phosphate crystals at 2.15 A resolution. Acta Crystallogr. D. Biol. Crystallogr. *59*, 1395-1407.

Chen,Y.J., Zhang,P., Egelman,E.H., and Hinshaw,J.E. (2004). The stalk region of dynamin drives the constriction of dynamin tubes. Nat. Struct. Mol. Biol. *11*, 574-575.

Cirillo,L.A., McPherson,C.E., Bossard,P., Stevens,K., Cherian,S., Shim,E.Y., Clark,K.L., Burley,S.K., and Zaret,K.S. (1998). Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. EMBO J. *17*, 244-254.

Clark,D.J. and Kimura,T. (1990). Electrostatic mechanism of chromatin folding. J. Mol. Biol. *211*, 883-896.

Cosgrove,M.S., Boeke,J.D., and Wolberger,C. (2004). Regulated nucleosome mobility and the histone code. Nat. Struct. Mol. Biol. *11*, 1037-1043.

Crowther,R.A., De Rosier,D.J., and Klug,A. (1970). The reconstruction of a three-dimensional structure from projections and its application to elctron microscopy. Proceedings of the Royal Society of London A *317*, 319-340.

Crowther,R.A. and Klug,A. (1975). Structural analysis of macromolecular assemblies by image reconstruction from electron micrographs. Annu. Rev. Biochem. *44*, 161-182.

Cui,Y. and Bustamante,C. (2000). Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure. PNAS *97*, 127-132.

Davey,C.A. and Richmond,T.J. (2002). DNA-dependent divalent cation binding in the nucleosome core particle. Proc. Natl. Acad. Sci. U. S. A *99*, 11169-11174.

Davey,C.A., Sargent,D.F., Luger,K., Maeder,A.W., and Richmond,T.J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. J. Mol. Biol. *319*, 1097-1113.

Dorigo,B., Schalch,T., Bystricky,K., and Richmond,T.J. (2003). Chromatin fiber folding: requirement for the histone H4 N-terminal tail. J. Mol. Biol. *327*, 85-96.

Dorigo,B., Schalch,T., Kulangara,A., Duda,S., Schroeder,R.R., and Richmond,T.J. (2004). Nucleosome arrays reveal the two-start organization of the chromatin fiber. Science *306*, 1571-1573.

Dutnall,R.N. and Ramakrishnan,V. (1997). Twists and turns of the nucleosome: tails without ends. Structure. *5*, 1255-1259.

Egelman,E.H. (2000). A robust algorithm for the reconstruction of helical filaments using single-particle methods. Ultramicroscopy *85*, 225-234.

Finch,J.T. and Klug,A. (1976). Solenoidal model for superstructure in chromatin. Proc. Natl. Acad. Sci. U. S. A *73*, 1897-1901.

Flaus,A., Luger,K., Tan,S., and Richmond,T.J. (1996). Mapping nucleosome position at single base-pair resolution by using site-directed hydroxyl radicals. Proc. Natl. Acad. Sci. U. S. A *93*, 1370-1375.

Frank,J., Radermacher,M., Penczek,P., Zhu,J., Li,Y., Ladjadj,M., and Leith,A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J. Struct. Biol. *116*, 190-199.

Froimowitz,M. (1993). HyperChem: a software package for computational chemistry and molecular modeling. Biotechniques *14*, 1010-1013.

Furrer,P., Bednar,J., Dubochet,J., Hamiche,A., and Prunell,A. (1995). DNA at the entry-exit of the nucleosome observed by cryoelectron microscopy. J. Struct. Biol. *114*, 177-183.

Galkin,V.E., Orlova,A., VanLoock,M.S., Rybakova,I.N., Ervasti,J.M., and Egelman,E.H. (2002). The utrophin actin-binding domain binds F-actin in two different modes: implications for the spectrin superfamily of proteins. J. Cell Biol. *157*, 243-251.

Gerchman,S.E. and Ramakrishnan,V. (1987). Chromatin higher-order structure studied by neutron scattering and scanning transmission electron microscopy. Proc. Natl. Acad. Sci. U. S. A *84*, 7802-7806.

Greyling, H. J. The reconstitution of the histone octamer. 1987. University of Cape Town. MSc Thesis/Dissertation.

Hamiche,A., Schultz,P., Ramakrishnan,V., Oudet,P., and Prunell,A. (1996). Linker histone-dependent DNA structure in linear mononucleosomes. J. Mol. Biol. *257*, 30-42.

Hansen,J.C. (2002). Conformational dynamics of the chromatin fiber in solution: determinants, mechanisms, and functions. Annu. Rev. Biophys. Biomol. Struct. *31*, 361-392.

Harp,J.M., Hanson,B.L., Timm,D.E., and Bunick,G.J. (2000). Asymmetries in the nucleosome core particle at 2.5 A resolution. Acta Crystallogr. D. Biol. Crystallogr. *56 Pt 12*, 1513-1534.

Harp,J.M., Uberbacher,E.C., Roberson,A.E., Palmer,E.L., Gewiess,A., and Bunick,G.J. (1996). X-ray diffraction analysis of crystals containing twofold symmetric nucleosome core particles. Acta Crystallogr. D. Biol. Crystallogr. *52*, 283-288.

Harris,J.R. and Scheffler,D. (2002). Routine preparation of air-dried negatively stained and unstained specimens on holey carbon support films: a review of applications. Micron. *33*, 461-480.

Holst,M. (2001). Adaptive Numerical Treatment of Elliptic Systems on Manifolds. Advances in Computational Mathematics *15*, 139-191.

Horowitz-Scherer,R.A. and Woodcock,C.L. (2005). Organization of interphase chromatin. Chromosoma 1-14.

Jiang,W. and Ludtke,S.J. (2005). Electron cryomicroscopy of single particles at subnanometer resolution. Curr. Opin. Struct. Biol. *15*, 571-577.

Jones,S. and Thornton,J.M. (1995). Protein-protein interactions: a review of protein dimer structures. Prog. Biophys. Mol. Biol. *63*, 31-65.

Jones,S. and Thornton,J.M. (1996). Principles of protein-protein interactions. Proc. Natl. Acad. Sci. U. S. A *93*, 13-20.

Joyeux,L. and Penczek,P.A. (2002). Efficiency of 2D alignment methods. Ultramicroscopy *92*, 33-46.

Klug,A., Crick,F.H.C., and Wyckoff,H.W. (1958). Diffraction by helical structures. Acta Cryst. *11*, 199-213.

Klug,A., Rhodes,D., Smith,J., Finch,J.T., and Thomas,J.O. (1980). A low resolution structure for the histone core of the nucleosome. Nature *287*, 509-516.

Kornberg,R.D. (1977). Structure of chromatin. Annu. Rev. Biochem. *46*, 931-954.

Kornberg,R.D. and Lorch,Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell *98*, 285-294.

Kossel,A. (1928). The Protamines and Histones. Longmans, London).

Laemmli,U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

Langst,G. and Becker,P.B. (2004). Nucleosome remodeling: one mechanism, many phenomena? Biochim. Biophys. Acta *1677*, 58-63.

Lowary,P.T. and Widom,J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J. Mol. Biol. *276*, 19-42.

Ludtke,S.J., Baldwin,P.R., and Chiu,W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. J. Struct. Biol. *128*, 82-97.

Luger,K. and Hansen,J.C. (2005). Nucleosome and chromatin fiber dynamics. Curr. Opin. Struct. Biol. *15*, 188-196.

Luger,K., Mader,A.W., Richmond,R.K., Sargent,D.F., and Richmond,T.J. (1997a). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Luger,K., Mader,A.W., Sargent,D.F., and Richmond,R.K. (1998). The Atomic Structure of Chromatin's Nucleosome Core Particle at 2 Å Resolution. ESRF Newsletter *31*, 9-11.

Luger,K., Rechsteiner,T.J., Flaus,A.J., Waye,M.M., and Richmond,T.J. (1997b). Characterization of nucleosome core particles containing histone proteins made in bacteria. J. Mol. Biol. *272*, 301-311.

Meersseman,G., Pennings,S., and Bradbury,E.M. (1992). Mobile nucleosomes--a general behavior. EMBO J. *11*, 2951-2959.

Miller,O.L., Jr. and Beatty,B.R. (1969). Visualization of nucleolar genes. Science *164*, 955-957.

Mirzabekov,A.D., Shick,V.V., Belyavsky,A.V., and Bavykin,S.G. (1978). Primary organization of nucleosome core particle of chromatin: sequence of histone arrangement along DNA. Proc. Natl. Acad. Sci. U. S. A *75*, 4184-4188.

Moody,M.F. (1967). Structure of the sheath of bacteriophage T4. I. Structure of the contracted sheath and polysheath. J. Mol. Biol. *25*, 167-200.

Moody,M.F. (1971). Application of optical diffraction to helical structures in the bacteriophage tail. Philos. Trans. R. Soc. Lond B Biol. Sci. *261*, 181-195.

Moudrianakis,E.N. and Arents,G. (1993). Structure of the histone octamer core of the nucleosome and its potential interactions with DNA. Cold Spring Harb. Symp. Quant. Biol. *58*, 273-279.

Muthurajan,U.M., Bao,Y., Forsberg,L.J., Edayathumangalam,R.S., Dyer,P.N., White,C.L., and Luger,K. (2004). Crystal structures of histone Sin mutant nucleosomes reveal altered protein-DNA interactions. EMBO J. *23*, 260-271.

Olins,A.L. and Olins,D.E. (1974). Spheroid chromatin units (v bodies). Science *183*, 330-332.

Olins,D.E. and Olins,A.L. (2003). Chromatin history: our view from the bridge. Nat. Rev. Mol. Cell Biol. *4*, 809-814.

Orlova,E.V., Dube,P., Harris,J.R., Beckman,E., Zemlin,F., Markl,J., and van Heel,M. (1997). Structure of keyhole limpet hemocyanin type 1 (KLH1) at 15 A resolution by electron cryomicroscopy and angular reconstitution. J. Mol. Biol. *271*, 417-437.

Oudet,P., Gross-Bellard,M., and Chambon,P. (1975). Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. Cell *4*, 281-300.

Pardon,J.F. and Wilkins,M.H. (1972). A super-coil model for nucleohistone. J. Mol. Biol. *68*, 115-124.

Penczek,P., Radermacher,M., and Frank,J. (1992). Three-dimensional reconstruction of single particles embedded in ice. Ultramicroscopy *40*, 33-53.

Penczek,P.A., Grassucci,R.A., and Frank,J. (1994). The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. Ultramicroscopy *53*, 251-270.

Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C., and Ferrin,T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605-1612.

Polak,E. and Ribiére,G. (1969). Note sur la convergence de directions conjugées. Rev. Francaise Informat Recherche Operationalle *16*, 35-43.

Quillin,M.L. and Matthews,B.W. (2000). Accurate calculation of the density of proteins. Acta Crystallogr. D. Biol. Crystallogr. *56 (Pt 7)*, 791-794.

Radermacher,M. (1988). Three-dimensional reconstruction of single particles from random and nonrandom tilt series. J. Electron Microsc. Tech. *9*, 359-394.

Ramakrishnan,V. (1997). Histone structure and the organization of the nucleosome. Annu. Rev. Biophys. Biomol. Struct. *26*, 83-112.

Rhodes,D. (1985). Structural analysis of a triple complex between the histone octamer, a Xenopus gene for 5S RNA and transcription factor IIIA. EMBO J. *4*, 3473-3482.

Richmond,T.J. and Davey,C.A. (2003). The structure of DNA in the nucleosome core. Nature *423*, 145-150.

Richmond,T.J., Finch,J.T., Rushton,B., Rhodes,D., and Klug,A. (1984). Structure of the nucleosome core particle at 7 A resolution. Nature *311*, 532-537.

Richmond,T.J., Rechsteiner,T., and Luger,K. (1993). Studies of nucleosome structure. Cold Spring Harb. Symp. Quant. Biol. *58*, 265-272.

Robinson,P.J., Fairall,L., Huynh,V.A., and Rhodes,D. (2006). EM measurements define the dimensions of the "30-nm" chromatin fiber: Evidence for a compact, interdigitated structure. Proc. Natl. Acad. Sci. U. S. A *103*, 6506-6511.

Robinson,P.J. and Rhodes,D. (2006). Structure of the '30nm' chromatin fibre: A key role for the linker histone. Curr. Opin. Struct. Biol. *16*, 336-343.

Sanner,M.F. (1999). Python: a programming language for software integration and development. J. Mol. Graph. Model. *17*, 57-61.

Sanner,M.F., Olson,A.J., and Spehner,J.C. (1996). Reduced surface: an efficient way to compute molecular surfaces. Biopolymers *38*, 305-320.

Schalch,T., Duda,S., Sargent,D.F., and Richmond,T.J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature *436*, 138-141.

Shannon,C.E. (1949). Communication in the presence of noise. Proceedings of the Institute of Radio Engineering *37*, 10-22.

Simon,R.H. and Felsenfeld,G. (1979). A new procedure for purifying histone pairs H2A + H2B and H3 + H4 from chromatin using hydroxylapatite. Nucleic Acids Res. *6*, 689-696.

Simpson,R.T. and Stafford,D.W. (1983). Structural features of a phased nucleosome core particle. Proc. Natl. Acad. Sci. U. S. A *80*, 51-55.

Strahl,B.D. and Allis,C.D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Suto,R.K., Clarkson,M.J., Tremethick,D.J., and Luger,K. (2000). Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. Nat. Struct. Biol. *7*, 1121-1124.

Suto,R.K., Edayathumangalam,R.S., White,C.L., Melander,C., Gottesfeld,J.M., Dervan,P.B., and Luger,K. (2003). Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands. J. Mol. Biol. *326*, 371-380.

Thastrom,A., Bingham,L.M., and Widom,J. (2004). Nucleosomal locations of dominant DNA sequence motifs for histone-DNA interactions and nucleosome positioning. J. Mol. Biol. *338*, 695-709.

Thoma,F., Koller,T., and Klug,A. (1979). Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. J. Cell Biol. *83*, 403-427.

Tietz,H.R. (1986). On-line image processing in real time for transmission electron microscopy. Journal de Microscopie et de Spectroscopie Electroniques, *11*.

Trus,B.L. and Steven,A.C. (1984). Diffraction patterns from stained and unstained helices: consistency or contradiction? Ultramicroscopy *15*, 325-335.

Tsunaka,Y., Kajimura,N., Tate,S., and Morikawa,K. (2005). Alteration of the nucleosomal DNA path in the crystal structure of a human nucleosome core particle. Nucleic Acids Res. *33*, 3424-3434.

Uberbacher,E.C., Harp,J.M., and Bunick,G.J. (1988). DNA sequence patterns in precisely positioned nucleosomes. J. Biomol. Struct. Dyn. *6*, 105-120.

Unser,M., Sorzano,C.O., Thevenaz,P., Jonic,S., El Bez,C., De Carlo,S., Conway,J.F., and Trus,B.L. (2005). Spectral signal-to-noise ratio and resolution assessment of 3D reconstructions. J. Struct. Biol. *149*, 243-255.

Unwin,P.N. and Henderson,R. (1975). Molecular structure determination by electron microscopy of unstained crystalline specimens. J. Mol. Biol. *94*, 425-440.

van Heel,M. and Frank,J. (1981). Use of multivariate statistics in analysing the images of biological macromolecules. Ultramicroscopy *6*, 187-194.

van Heel,M. and Harauz,G. (1986). resolution criteria for three dimensional reconstruction. Optik *73*, 119-122.

van Heel,M., Harauz,G., Orlova,E.V., Schmidt,R., and Schatz,M. (1996). A new generation of the IMAGIC image processing system. J. Struct. Biol. *116*, 17-24.

von Holt,C., Brandt,W.F., Greyling,H.J., Lindsey,G.G., Retief,J.D., Rodrigues,J.D., Schwager,S., and Sewell,B.T. (1989). Isolation and characterization of histones. Methods Enzymol. *170*, 431-523.

Wang,Y.A., Yu,X., Yip,C., Strynadka,N.C., and Egelman,E.H. (2006). Structural polymorphism in bacterial EspA filaments revealed by cryo-EM and an improved approach to helical reconstruction. Structure. *14*, 1189-1196.

Wedemann,G. and Langowski,J. (2002). Computer simulation of the 30-nanometer chromatin fiber. Biophys. J. *82*, 2847-2859.

White,C.L., Suto,R.K., and Luger,K. (2001). Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. EMBO J. *20*, 5207-5218.

Widom,J. (1998). Structure, dynamics, and function of chromatin in vitro. Annu. Rev. Biophys. Biomol. Struct. *27*, 285-327.

Williams,R.C. and Fisher,H.W. (1970). Electron microscopy of tobacco mosaic virus under conditions of minimal beam exposure. J. Mol. Biol. *52*, 121-123.

Williams,S.P., Athey,B.D., Muglia,L.J., Schappe,R.S., Gough,A.H., and Langmore,J.P. (1986). Chromatin fibers are left-handed double helices with diameter and mass per unit length that depend on linker length. Biophys. J. *49*, 233-248.

Wolffe,A. (1999). Chromatin: Structure and Function. Academic Press).

Wood,C.M., Nicholson,J.M., Lambert,S.J., Chantalat,L., Reynolds,C.D., and Baldwin,J.P. (2005). High-resolution structure of the native histone octamer. Acta Crystallographica Section F *61*, 541-545.

Woodcock,C.L. (2006). Chromatin architecture. Curr. Opin. Struct. Biol. *16*, 213-220.

Woodcock,C.L. and Dimitrov,S. (2001). Higher-order structure of chromatin and chromosomes. Curr. Opin. Genet. Dev. *11*, 130-135.

Woodcock,C.L., Frado,L.L., Hatch,C.L., and Ricciardiello,L. (1976). Fine structure of active ribosomal genes. Chromosoma *58*, 33-39.

Woodcock,C.L., Frado,L.L., and Rattner,J.B. (1984). The higher-order structure of chromatin: evidence for a helical ribbon arrangement. J. Cell Biol. *99*, 42-52.

Woodcock,C.L., Grigoryev,S.A., Horowitz,R.A., and Whitaker,N. (1993). A chromatin folding model that incorporates linker variability generates fibers resembling the native structures. Proc. Natl. Acad. Sci. U. S. A *90*, 9021-9025.

Woodcock,C.L. and Horowitz,R.A. (1995). Chromatin organization re-viewed. Trends Cell Biol. *5*, 272-277.

Woodcock,C.L. and Horowitz,R.A. (1997). Electron microscopy of chromatin. Methods *12*, 84-95.

Woodcock,C.L., McEwen,B.F., and Frank,J. (1991). Ultrastructure of chromatin. II. Three-dimensional reconstruction of isolated fibers. J. Cell Sci. *99 (Pt 1)*, 107-114.

Worcel,A., Strogatz,S., and Riley,D. (1981). Structure of chromatin and the linking number of DNA. Proc. Natl. Acad. Sci. U. S. A *78*, 1461-1465.

Wriggers,W., Milligan,R.A., and McCammon,J.A. (1999). Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. J. Struct. Biol. *125*, 185-195.

Yao,J., Lowary,P.T., and Widom,J. (1991). Linker DNA bending induced by the core histones of chromatin. Biochemistry *30*, 8408-8414.

Yu,X., Jacobs,S.A., West,S.C., Ogawa,T., and Egelman,E.H. (2001). Domain structure and dynamics in the helical filaments formed by RecA and Rad51 on DNA. Proc. Natl. Acad. Sci. U. S. A *98*, 8419-8424.

Zhao,F.Q. and Craig,R. (2003). Capturing time-resolved changes in molecular structure by negative staining. J. Struct. Biol. *141*, 43-52.

Zlatanova,J. and Leuba,S.H. (2003). Chromatin fibers, one-at-a-time. J. Mol. Biol. *331*, 1-19.

Zlatanova,J., Leuba,S.H., and van Holde,K. (1999). Chromatin structure revisited. Crit Rev. Eukaryot. Gene Expr. *9*, 245-255.