To the Graduate Council:

I am submitting herewith a thesis written by Joel Monroe Harp entitled "Structural Studies of the Nucleosome Core Particle." I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Biomedical Sciences.

Gerard J. Bunick, Major Professor

We have read this thesis
and recommend its acceptance:

Elizabeth E Howell

Eugene Serpersu

Jeffrey M Becker

Accepted for the Council:

Interim Vice Provost and

Dean of The Graduate School

STRUCTURAL STUDIES OF THE NUCLEOSOME CORE PARTICLE

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Joel Monroe Harp

December 2000

Acknowledgments

The work reported here is most certainly not all my own. Research is not a solitary enterprise. It is necessary to recognize the contributions of all those who contributed as well as those who have influenced it. I must begin with the names of Dr. Gerard J. Bunick and Dr. Edward C. Uberbacher. They gave me a wonderful opportunity that has made all the difference. Particular thanks go to my mentor, Gerry Bunick, who has allowed me to mature as a scientist and to pursue a degree while working in his laboratory. Ed Uberbacher has been a most valuable influence on my development as a scientist and should also be acknowledged for pioneering work on the use of a defined sequence DNA palindrome to improve the quality of diffraction from nucleosome core particle crystals. The years of working at the bench with him and Dr. Arthur Roberson were pure fun.

There are many who have come through this laboratory during the years that it took to achieve the nucleosome core particle structure. They have all made a contribution.

They are: Christopher Bunick, Lucia Cacheiro, Ryan Carrick, Matthew Davis, Andreas Geweiss, Leif Hanson, Elise Palmer, Steve Plichta, Arthur Roberson, Elizabeth Singley, David Timm, Everline Wright, and Melissa York. Chris Bunick has been an enthusiastic and energetic contributor and tireless advocate. Leif Hanson, fellow reformed invertebrate zoologist, worked long and hard on fitting and refinement and is responsible for some elegant work with the difficult histone tails and identified the cacodylate ion which plays such a prominent role in crystal packing. Elise Palmer was responsible for working through the large-scale production of the DNA palindrome. Steve Plichta will always be remembered for his good humor, adequate singing voice, and his hard work. Dr. David Timm, who came to us from the prestigious laboratory of Sir Thomas Blundell, is widely known as a brilliant and aggressive scientist. The time he spent at Oak Ridge was certainly of great benefit to me. All have been enjoyable and valuable colleagues.

I thank the members of the dissertation committee, Dr. Jeffrey Becker, Dr. Elizabeth Howell, and Dr. Engin Serpersu for their willingness to serve and for their time and attention. I also wish to give special thanks to Elizabeth Singley for her generosity in reading and editing the manuscript.

Clyde Mellon and James Hall of the Cherokee Poultry Unit of the University of Tennessee School of Agriculture were unfailingly helpful in supplying all of the chickens required for preparation of the large quantities of histones used in this work.

All of the folks in the graduate school have made a difference. The late Barbara Moorman will always be fondly remembered. I am particularly grateful to Gaynelle Russell and Kay Gardner who, in addition to the usual headaches, have some how managed to work out all of the administrative problems associated with my position as a part-time student and research associate.

Thanks are due to Dr. Daniel Carter and present and former colleagues at NASA's Marshall Space Flight Center and New Century Pharmaceuticals for making the microgravity environment available for crystallization of the nucleosome core particle.

Diffraction data for this study were collected at Brookhaven National Laboratory in the Biology Department single-crystal diffraction facility at beamline X12-C in the National Synchrotron Light Source. This facility is supported by the United States Department of Energy Offices of Health and Environmental Research and of Basic Energy Sciences under prime contract DE-AC02-98CH10886, by the National Science Foundation, and by National Institutes of Health Grant 1P41 RR12408-01A1.

Thanks are due to Dr. Nic Steussy, Dr. Thomas Hurley, and Dr. Sangita Sinha for permission to use unpublished results.

My wonderful wife, Ruth, and son, Wesley, have been understanding and supportive through it all. I thank my parents for being supportive of my eccentricities. And, finally, I wish to acknowledge the contribution of Wiley F. Thurman who showed the

adventure of science to a barefooted kid in a dusty West Texas yard. Thanks, Wiley, wherever you are.

Abstract

The 2.5 Å X-ray crystal structure of the nucleosome core particle presented here provides significant additions to the understanding of the nucleosome, the fundamental unit of chromatin structure. Extensions are made to the structure of the N-terminal histone tails and details provided on hydration and ion binding. The structure is composed of 2-fold symmetric molecules, native chicken histone octamer, and the DNA palindrome, which were expected to form a perfectly 2-fold symmetric nucleosome core particle. In fact, the result is asymmetric due to the binding of the DNA to the protein surface and to the packing of the particles in the crystal lattice. An analysis is made of the asymmetries by comparisons both within the nucleosome core particle and to the structure of the histone octamer core of the nucleosome.

Details of the discovery and investigation of macromolecular crystal annealing, M-CA, are also presented. This discovery was a direct consequence of studies directed toward obtaining high-resolution diffraction from nucleosome core particle crystals.

The existence of the phenomenon was completely unexpected and thus no documentation of its mechanism or application to macromolecular crystals was available. MCA has already proven to be an important contribution to the field of macromolecular cryocrystallography.

Research sponsored by grants from NIH (GM-29818), NASA (NAG8-1568), the Office of Biological and Environmental Research, U.S. Department of Energy, and the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

Contents

1 Introduction **1**

1.1 A short history of the nucleosome. 5

1.1.1 In the beginning. 7

1.1.2 The structure of the histone core of the nucleosome 10

1.1.3 Early nucleosome crystal structures 14

1.1.4 Crystals of NCPs containing defined sequence DNA 22

1.2 Macromolecular crystal annealing 27

1.2.1 What is mosaicity? 29

1.3 Glossary 33

2 Methods **36**

2.1 Preparation of histones from chicken erythrocyte nuclei. 36

2.1.1 Buffers 37

2.1.2 Detailed protocol for preparation of soluble chromatin 39

2.1.3 Hydroxyapatite chromatography 42

2.1.4 Final purification on Superose 12 43

2.2 Expression and purification of recombinant histones. 44

2.3 Assembly of recombinant octamer. 44

2.4 Identification of candidate DNA sequences. 45

2.5 Footprinting analysis of phasing of the ASN DNA on histone octamer cores. 46

2.5.1 Indirect end labeling 47

2.5.2 Construction of marker fragments by guanine base-specific chemical cleavage 49

2.5.3 DNase I digestion of naked and bound DNA 51

2.5.4 Analysis of the protection pattern 52

2.6 Large scale purification of 146 bp α-satellite DNA palindrome fragments 53

2.7 Large scale production of ASP DNA by ligations of half palindromes 54

2.7.1 Cloning of 1/2ASP for large scale production 54

2.7.2 Large scale production of plasmid containing 1/2ASP DNA fragments 56

2.7.3 Construction of 146 bp ASP palindrome DNA from plasmid containing 1/2ASP DNA fragments 59

2.8 Reconstitution of nucleosome core particles 62

2.9 Purification of correctly phased nucleosome core particles 64

2.9.1 Micrococcal nuclease analysis of reconstituted nucleosome core particles 65

2.9.2 Preparative polyacrylamide gel electrophoresis 66

2.10 Microgravity crystallization of nucleosome core particles 70

2.11 X-ray diffraction data collection 76

2.11.1 Preparation and mounting of crystals for low-temperature data collection 76

2.11.2 Macromolecular crystal annealing 77

2.11.3 In-house data collection 79

2.11.4 Synchrotron data collection 79

2.11.5 Data reduction and quality analysis 80

2.12 Phasing of the data 80

2.12.1 Isomorphous replacement 81

2.12.2 Molecular replacement 85

2.13 Fitting and refinement of model coordinates 86

2.14 Model analysis 87

2.14.1 Analysis of DNA helical parameters 87

2.14.2 Analysis of the molecular models 88

2.14.3 Preparation of figures 90

3 Results **92**

3.1 Sequencing of histone clones 92

3.2 Large-scale preparation of ASP DNA 92

3.3 Microgravity crystallization using the DCAM 99

3.4 Space group determination 105

3.5 Cryogenic data collection and macromolecular crystal annealing . . . 105

3.5.1 Does the cryoprotectant have an effect on the results of the annealing treatment? 110

3.5.2 Can any crystal be annealed? 111

3.6 Diffraction analysis of nucleosome core particle crystals 116

3.7 Fitting and refinement of the model 117

3.8 Hydration and ion binding 119

3.8.1 Ions 119

3.8.2 Structural water 122

3.9 Structure of the nucleosome core particle at 2.5 Å 124

3.9.1 General features of the molecular structure 124

3.9.2 The histone fold and the handshake motif 129

3.9.3 Structure of the $(H3{\cdot}H4)_2$ tetramer 132

3.9.4 Structure of the individual histone proteins 132

3.9.5 DNA binding to the octamer 143

3.9.6 Folding of the DNA into the nucleosome core particle 160

3.9.7 Stacking interactions between histone faces 167

4 Discussion **172**

4.1 Microgravity crystallization: Was it worth it? 172

4.2 Macromolecular crystal annealing. 175

4.2.1 The definition of success 176

4.3 Structure of the nucleosome core particle as seen in the 2.5 Å model . 179

4.3.1 Influence of the solution environment on the crystal structure 179

4.3.2 Applications of the NCP structure to questions of function . . 185

4.3.3 Future directions 190

Bibliography **195**

Vita **210**

List of Figures

2.1 Sequence of the 1/2ASP DNA fragment 54

2.2 Diagram of the preparative electrophoresis elution buffer flow 67

2.3 Diagram of the DCAM apparatus 71

2.4 Color code for identification of histone subunits in figures 91

3.1 Sequence of chicken histone H2A gene with the amino acid translation of the DNA sequence. 93

3.2 Sequence of chicken histone H2B gene with the amino acid translation of the DNA sequence. 94

3.3 Sequence of chicken histone H3 gene with the amino acid translation of the DNA sequence. 95

3.4 Sequence of chicken histone H4 gene with the amino acid translation of the DNA sequence. 96

3.5 Equilibration of $MnCl_2$ concentration in a DCAM 100

3.6 A crystal grown on microgravity mission, USML-2, which shows an hexagonal cross-section suggesting that the crystal was grown under microgravity conditions. 102

3.7 A very large nucleosome core particle crystal grown in a DCAM aboard the Mir space station. 102

3.8 Comparison of microgravity and ground-based diffraction data 104

3.9 Comparison of diffraction images before and after macromolecular crystal annealing. 107

3.10 Demonstration of the identity of the cacodylate ion 123

3.11 Views of the nucleosome core particle. 126

3.12 Overlap of histone structure in the NCP on the histone octamer without DNA 127

3.13 Overlap of the two halves of the histone core of the NCP 128

3.14 Structure of the histone fold motif and handshake interaction of histones131

3.15 Structure of the $(H3{\cdot}H4)_2$ tetramer 133

3.16 Secondary structure alignment with amino acid sequence of histone H2A 135

3.17 Secondary structure alignment with amino acid sequence of histone H2B 136

3.18 Secondary structure alignment with amino acid sequence of histone H3 136

3.19 Secondary structure alignment with amino acid sequence of histone H4 137

3.20 Interaction of the N-terminal tails of H2A between symmetry neighbors 138

3.21 Stereo view of the interaction of N-terminal tails of H2A between symmetry neighbors 139

3.22 C-terminal domains of the dimer histone subunits 141

3.23 Mapping of minor groove-in, mGI, positions 145

3.24 Plot of temperature factors refined for phosphorus atoms along the backbone of the DNA 146

3.25 Graphical representation of temperature factors for DNA 146

3.26 Stereo view of the structure of the αmGI binding site at dorsal mGI 5 148

3.27 The αmGI motif found at mGI 2 149

3.28 Comparison of dorsal mGI 1 in the octamer and in the NCP structure 152

3.29 Direct binding at the N-terminus of the mH helix of H3 to DNA at mGI 3 154

3.30 The structure of the β-bridge at mGI 4 156

3.31 The structure of dorsal mGI 6 158

3.32 Base stacking is continuous between neighboring NCPs 159

3.33 Overlapping of the phosphodiester backbone traces 163

3.34 Stereo packing diagram showing the continuous path of DNA from the NCP to symmetry neighbor 164

3.35 Plot of helical twist expressed as base pairs per turn over the length of the DNA. 166

3.36 Electrostatic surface potential map of the histone octamer 167

3.37 Stacking interactions between histone faces of neighboring NCPs . . . 169

3.38 Enlarged view of stacking interactions between histone faces 170

3.39 GRASP electrostatic potential surface laid over a model of the dimers 171

4.1 An example of the use of macromolecular crystal annealing to recover a mishandled crystal 178

4.2 Stereo view of the DNA phosphodiester backbone trace from the 8 Å structure of the NCP 183

4.3 Homology models of *C. elegans* H3-like protein D6H3 187

4.4 Positions of arginines 128, 129, 131, and 134 on the C-terminus of H3 189

List of Tables

3.1 List of microgravity experiments flown. 100

3.2 Crystal systems used for the study of macromolecular crystal annealing. 112

3.3 Crystal systems that have been successfully annealed in addition to those used in this study. 113

3.4 Summary of the results of annealing treatments 115

3.5 Diffraction and phasing statistics 117

3.6 Refinement statistics 118

3.7 Tabulation of Mn^{2+} ions in the NCP structure 120

3.8 Tabulation of potassium ions in the structure 121

3.9 Tabulation of minor groove-in, mGI, positions 144

Chapter 1

Introduction

> Adhere to your own act, and congratulate yourself if you have done something strange and extravagant, and broken the monotony of a decorous age.
> – Ralph Waldo Emerson

The research presented here is the culmination of 15 years of work resulting in the 2.5 Å structure of the nucleosome core particle (PDB ID code 1EQZ). The atomic resolution model provides a pivot point in the investigation of chromatin structure. Studies of structural issues relating to the nucleosome itself as well as higher-order chromatin structure are now feasible by using lessons learned in the production of high-quality crystals of the nucleosome core particle (NCP) and by using the structural model to phase diffraction data for new structures. These studies can immediately attack such vital issues as the function of nucleosome structure in gene control. Details of

chromosomal architecture can be examined through structures of nucleosome variants occurring in the centromere. Larger scale studies can utilize oligonucleosomes forming engineered chromatin fragments complexed with proteins known to modulate chromatin structure and function. This is a time of great advances in the study of chromatin due to the emergence of evidence linking chromatin structure as an essential factor in the regulation of chromatin function (Widom, 1998). The structure presented here opens a new world of opportunities for investigation of the exquisite design of what is perhaps the most complex and powerful command and control structure ever known, the cell nucleus.

The quest for a high resolution structure of the nucleosome has seen three important milestones: the 3.1 Å structure of the histone core of the nucleosome (Arents *et al.*, 1991), the development by this laboratory of the human X-chromosome α-satellite DNA palindrome to extend the two fold symmetry to the NCP and avoid potential packing disorder in the crystals (Harp *et al.*, 1996), and the use of recombinant histones to facilitate heavy atom incorporation in the NCP (Luger *et al.*, 1997).

The NCP is a substructure of the canonical nucleosome defined by the DNA protection pattern of histone octamer in nuclease digestion of chromatin. It contains the histone octamer core with 2 each of the core histones, H2A, H2B, H3, and H4, and 145 to 147 base pairs of DNA wrapped in about 1.75 superhelical turns around the histone core. The octamer forms a superhelical ramp capable of bending duplex DNA into a compact structure. The NCP lacks the DNA that links adjacent nucleosomes in

chromatin, the linker DNA, and a linker histone such as H1. The DNA and protein constituents of the NCP contribute nearly equal molecular weight to the final 208 kDa complex.

The determination of the crystal structure of the NCP has consumed such a long period of time in large part due to the packaging function of the structure. The histone octamer core of the nucleosome does not bind DNA in a sequence specific fashion. High resolution crystallographic studies require that NCPs should possess a degree of uniformity in structure not commonly seen in nature. In those cases where nucleosomes are known to bind at precise locations determined by DNA sequence, those positions are generally involved in gene regulation or formation of nucleosome arrays essential to development of higher order structures. Much of the work presented here involves the development of the human X-chromosome α-satellite DNA palindrome as it was of critical importance to the production of high quality NCP crystals. The concept of using DNA palindromes to engineer NCP crystals was developed in this laboratory to solve two critical problems. The first was to locate or engineer a DNA fragment with sequence-dependent structural properties allowing it to bind symmetrically on the histone octamer core. The use of α-satellite sequences took advantage of a rich source of nucleosome positioning, or phasing, sequences designed by nature. Engineering of a DNA palindrome based on the nucleosome phasing sequences found in α-satellite DNA served to ameliorate the second critical problem, potential 2-fold disorder introduced into the crystal packing when the DNA on either side of the NCP

is different. The use of the DNA palindrome served to extend the 2-fold symmetry of the histone octamer core to the entire NCP.

The α-satellite DNA palindrome was incorporated into NCPs by reconstitution onto histone octamers purified from chicken erythrocyte chromatin. Histones purified from the inactive erythrocyte nuclei are homogeneous and unaltered by post-translational modifications such as acetylation or phosporylation and, as the purification involved only mild salt extraction, the folding is authentic and not subject to uncertainties involved in the use of recombinant material. The phasing of the even numbered 146 base pair DNA palindrome was predicted to result in placement of the 2-fold symmetry axis of the histone octamer core passing between base pairs 73 and 74, leaving 73 base pairs of DNA positioned on either side of the NCP. That is, the center of palindrome sequence symmetry and the 2-fold axis of symmetry of the histone core would coincide resulting in a perfectly 2-fold symmetric structure. In fact, the 2-fold axis of symmetry of the histone octamer, the dyad, passes through base 73, leaving 72 base pairs of DNA on one side and 73 base pairs of DNA on the other with one base pair lying directly on the dyad. The structure is asymmetric. This asymmetry may be the result of the necessity of mutating the central base pairs to introduce an *Eco*R I restriction site for construction of the palindrome. An analysis of asymmetries both in the DNA and in the protein moieties form a large and significant portion of the results and will be discussed in detail.

The crystal structure of the nucleosome, as well as the histone core of the nucleosome,

has been the goal of only a few groups worldwide for more than 20 years. Success has required solution of a number of difficult problems, not the least of which involved engineering NCPs capable of forming crystals that diffract X-rays to a high resolution. A brief recapitulation of the history of NCP structural studies may be useful to place the current structure within the context of a long and often contentious chapter in chromatin research.

1.1 A short history of the nucleosome.

> A eukaryotic chromosome made out of self-assembling 70 Å units, which could perhaps be made to crystallize, would necessitate rewriting our textbooks on cytology and genetics! I have never read such a naive paper purporting to be of such fundamental significance. Definitely it should not be published anywhere! – Anonymous review of paper submitted by C.F.L. Woodcock

We now take the existence of the nucleosome as the defining feature of chromatin very much for granted. It is therefore instructive to reflect that before 1971, the possibility of such a structure had not occurred to researchers at the forefront of the field. Before the discovery of the nucleosome, it was generally believed that the histones were uniformly distributed along the length of the DNA in the chromosome in what was known as the uniform supercoil model. The existence of the histones and the double helical structure of DNA were well known. There was no reason to suspect that the basic histone proteins did much more than coat the highly charged, acidic polyelectrolyte structure of the DNA as it packed into the nucleus. The discovery of

the nucleosome was nearly as revolutionary in its time as that of the DNA double helix. Even now, as the structure of the nucleosome is revealed in exquisite detail by X-ray crystallography, the revolution is still unfolding as the significance of the nucleosome is seen to expand into every facet of the storage and management of the genetic information contained in the nucleus. Looking back over the history of the struggle for the nucleosome structure, it seems most remarkable that the story has been so very long in the making. Thus, by way of introduction to the subject, a brief historical sketch will be presented.

An historical perspective of the nucleosome is particularly appropriate here as the high-resolution X-ray crystallographic studies reported are the culmination of more than 15 years of highly focused effort. The quote above was taken from the chapter on the development of the nucleosome model in the definitive history of its time (van Holde, 1988). The anonymous reviewer's comments concerned a manuscript containing some of the first electron micrographs of nucleosomes, that were taken by Dr. Christopher Woodcock. That work was done independently of and simultaneously with the work done here at Oak Ridge National Laboratory by Drs. Ada and Donald Olins, who had borrowed a technique from Oscar Miller, Jr., for centrifuging swollen nuclei through a sucrose cushion onto a carbon film on an electron microscope sample grid. Both groups, unaware of the other, presented their micrographs at the annual Cell Biology Meeting in 1973 (Olins & Olins, 1973; Woodcock, 1973). Fortunately for the Olinses, the reviews of their manuscript were less skeptical than that with which

Woodcock had to contend, and their paper (Olins & Olins, 1974) became a citation classic (Olins & Olins, 1983). In the realization that one picture is truly worth a thousand gels, the Olinses brought in a post-doctoral researcher, Dr. Gerard Bunick, later joined by Dr. Edward Uberbacher, to begin work on X-ray crystallographic studies of the nucleosome at Oak Ridge National Laboratory. Bunick was made a staff scientist at ORNL and together with Uberbacher began the studies that have culminated in the structure of the nucleosome core particle that forms the subject of this thesis. The author began working with the Olinses as an electron microscopist in 1983 and in 1985 was given the great opportunity of becoming a part of the nucleosome project with Drs. Bunick and Uberbacher.

1.1.1 In the beginning.

The first half of the decade of the 1970s saw a remarkable chapter in chromatin research. The uniform supercoil model of chromatin was suddenly and quite unexpectedly overthrown and replaced by the nucleosomal model which we now take for granted. The overthrow came about through two primary lines of evidence involving nuclear DNA on the one hand and, on the other, the self-association and stoichiometry of the core histones, H2A, H2B, H3, and H4.

The first clue to the existence of a repeating unit in chromatin came as the result of nuclease digestion that produced a ladder of bands on non-denaturing gel elec-

trophoresis (Hewish & Burgoyne, 1973). The nucleosomal ladder was discovered in South Australia by Hewish and Burgoyne, working with preparations of isolated rat liver nuclei. That work had been preceded by a couple of papers which had hinted at a regularly repeating subunit structure in chromatin (Williamson, 1970; Clark & Felsenfeld, 1971), but the Australian experiments were convincing for two reasons. First, they were clearly working with nuclear DNA which was gradually degraded by an endogenous Ca-Mg endonuclease to give a population of DNA fragments with sizes in integral multiples of a single minimum size, about 200 base pairs. And second, preparations of the endogenous nuclease digested naked, protein-free, DNA clearly demonstrating that it was the protection of DNA from nuclease digestion by the histones that resulted in the observed ladder of bands from nuclei. Hewish and Burgoyne were the first to unequivocally state the hypothesis that chromatin possessed a regularly repeating subunit structure.

A number of groups were examining the association of histone proteins in the absence of DNA. It was known previously that individual histones tended to form dimers. However, it was becoming apparent that mixtures of histones behaved peculiarly. A number of papers were published detailing the behaviors of nearly all possible combinations of the 4 histones but the most careful and complete studies were done in a series of papers by D'Anna and Isenberg at Corvallis, Oregon (D'Anna & Isenberg, 1973; D'Anna & Isenberg, 1974a; D'Anna & Isenberg, 1974b; D'Anna & Isenberg, 1974c; D'Anna & Isenberg, 1974d). The interaction that proved most useful in eluci-

dating the nucleosome model was that of H3 and H4 which existed as an H3/H4 dimer that could be reversibly associated to form a stable tetramer. This was accomplished at the Medical Research Council Laboratory at Cambridge where Roger Kornberg was able to synthesize an exposition of the nucleosome based on the three pieces of evidence: the ladder of DNA bands resulting from nuclease digestion with a fundamental repeat of about 200 base pairs; the self-association of histones H3 and H4 into a tetramer which in itself provided a basis for the assumption of some sort of subunit structure in chromatin; and the equimolar stoichiometry of the 4 histones (Kornberg, 1974). If, as he hypothesized, the four histones, H2A, H2B, H3, and H4 associated to form an octamer with two units of each histone, then it would be natural for the 200 base pairs to wrap around the protein to form a structure which he then dubbed, the nucleosome, that would display all of the physical properties measured by analytical ultracentrifugation and other biophysical techniques. Of course, there was no direct evidence for the path of the DNA through the nucleosome structure. It was for later biophysical experiments to verify that aspect of the model.

The Cold Spring Harbor Symposium held in June of 1977 seems to have served as the turning point in chromatin research which established the nucleosome model as true and correct and presented the beginnings of research aimed at elucidation of the many questions raised by the model of 8 histone subunits and 200 bp of DNA. In his summary of the proceedings, Chambon (Chambon, 1978) referred to the "coming-of-age" of chromatin research. Chromatin was no longer thought of as simply a more

complicated, and messy, version of the bacterial system. It was organized according to an algorithm for the maintenance and management of genomes whose beauty and complexity we are just beginning to fully appreciate.

1.1.2 The structure of the histone core of the nucleosome

One on the more entertaining episodes in the history of chromatin research involves the work of Evangelos Moudrianakis and his coworkers who were the first to crystallize the histone octamer (Burlingame *et al.*, 1984). However, the overall dimensions of the 3.3 Å structure (Burlingame *et al.*, 1985) did not appear to fit within the known X-ray structures of the nucleosome core particle (Richmond *et al.*, 1984). The octamer model was presented as a prolate ellipsoid, a "Rugby ball," with a maximum chord of about 110 Å. The nucleosome core particle appeared to be a flattened cylinder, like a cat food can, with a diameter of about 110 Å and a height of 57 Å as modeled from small-angle scattering, X-ray crystallography and neutron diffraction studies (Finch *et al.*, 1980; Klug *et al.*, 1980; Bentley *et al.*, 1981; Finch *et al.*, 1981; Bentley *et al.*, 1984). The unexpected conformation of the histone octamer was claimed to be more biologically relevant than the nucleosome since the nucleosome is crystallized without the linker histone, H1, and the linker DNA – a peculiar objection since the octamer was without even the core DNA and was crystallized in high salt.

A vigorous exchange in the pages of *Science* (Klug *et al.*, 1985; Uberbacher & Bunick,

1985b; Moudrianakis *et al.*, 1985) followed the publication of the histone structure and the assertion by Moudrianakis that the Rugby ball conformation was the correct physiological condition and that the nucleosome model was an artifact! There were two main proposals put forward from other groups to explain the discrepancy. The Oak Ridge response did not dispute the correctness of the crystallography and gave the Moudrianakis group the benefit of the doubt. The Oak Ridge response suggested a possible misinterpretation of the structure on the part of the Moudrianakis group and went to some length to square the reported structure with the canonical nucleosome (Uberbacher & Bunick, 1985b). Uberbacher showed that it might be possible that an alternative docking of the histone core with nucleosomal DNA, as hypothesized by Moudrianakis, could rotate the putative structure of the histone core into the canonical structure of the nucleosome. The dimensions of the histone octamer might also have been a consequence of the high salt of the crystallization medium, about 70% saturated ammonium sulfate. We subsequently demonstrated that the radius of gyration of the histone octamer varied considerably with ionic strength as measured by small-angle neutron scattering (Uberbacher *et al.*, 1986). The high salt environment of the histone octamer crystals had certainly produced a more extended structure than would be seen in the histone core bound to DNA in the nucleosome crystals under much lower ionic strength. It was unfortunate that a miscalculation of the molarity of the ammonium sulfate solution based on percent saturation was incorrectly reported in that paper as it only served to cloud the issue (see (van Holde, 1988) page 167)

and had no real bearing on the validity of the results.

Another explanation of Moudrianakis' result was put forward by Sir Aaron Klug and his colleagues at the Medical Research Council Laboratory in Cambridge, England (Klug *et al.*, 1985). They astutely reasoned that in the single isomorphous replacement with anomalous scattering (SIRAS) solution (discussed below), the proximity of the heavy atom binding site to a special position in the unit cell could have produced a spurious solution. The only cysteine in the histone octamer, at residue 110 in H3, caused the mercury atom to be positioned within angstroms of a special position at fractional coordinates of (1/3, 1/3, 0). Had the mercury been directly on the special position, it would have had zero phasing power. The solution of the crystal structure of the histone octamer was one of the first applications of the method of SIRAS as developed by Bi-Cheng Wang (Wang, 1985) and remains one of the more impressive examples of that technique, which uses density modification and real space noise filtering to solve the phase ambiguity presented by a single heavy atom derivative. To the Klug group, the correctness of the phasing was suspect because of both the proximity of the special position and the lack of additional phase information from multiple heavy atom derivatives.

The answer to the question of the validity of the structure and its interpretation required another ten years. No further structural papers were published based on the original electron density maps because it was impossible to interpret all of the density. This prompted Wang and John Rose to rethink the phasing. In a fascinating

and elegant analysis (Wang *et al.*, 1994) they discovered that the original phasing of the histone octamer data suffered from a misplacement of the heavy atom site by 2.7 Å. The proximity of the site to the special position had, as Klug and his colleagues had suggested, reduced the phasing power of the mercury atom, and had further caused a peculiar condition in the final electron density map in which the true structure had superimposed on it two additional copies. Each image of the octamer was rotated by 120 degrees about the three-fold axis. This left certain regions of the dimer clear and easily interpretable as helices while the bulk of the true structure was obscured by overlap with the two spurious images. This accounted for the original misinterpretation of the shape of the histone octamer which produced the controversy.

A number of papers were published using the 3.1 Å structure of the histone core of the nucleosome once the phasing problem had been solved (Arents *et al.*, 1991; Arents & Moudrianakis, 1993; Arents & Moudrianakis, 1995; Baxevanis *et al.*, 1995; Moudrianakis & Arents, 1993) and the coordinates of the structure were deposited in the PDB in 1991. However, since the coordinates were not released until 1998, a great deal of effort was expended toward solving the octamer structure at Oak Ridge. Those efforts were largely unsuccessful as were the efforts made in the laboratory of Timothy Richmond at the Eidgenossiche Technische Hochschule in Zurich (Toldo, 1995). The solution of the original histone octamer structure was, thus, made even more remarkable by the lack of success in two laboratories to repeat the studies using nearly identical methods. With the coordinates of the histone octamer now in hand,

we are able to verify that the structure of the histone octamer is quite close to the structure of the octamer core of the nucleosome as seen in the reported structures from both the Richmond and Bunick groups.

1.1.3 Early nucleosome crystal structures

The first crystals of the nucleosome core particle were prepared in Russia (Bakayev *et al.*, 1975). These were prepared from nucleosome core particles isolated from mouse Erlich ascites tumor cells and were used to produce X-ray powder diffraction data. The powder diffraction was capable only of recording a strong ring at about 51 Å with weaker data for higher spacings.

Klug's group at Cambridge formed what would become the leading program for the nucleosome crystal structure of the nucleosome and whose first efforts were reported in 1977 (Finch *et al.*, 1977; Finch & Klug, 1978). The first single crystal data were collected from crystals of nucleosome core particles isolated from rat liver chromatin (Lutter, 1978). The crystals diffracted to 20 Å and displayed unit cell dimensions of a=110 Å b=192 Å and c=340 Å in space group $P2_12_12_1$ with 3 nucleosomes in the asymmetric unit. Phasing of the diffraction data was solved using optical transforms of electron micrographs of the crystals (Crowther & Klug, 1975). It was found that the crystals contained proteolyzed histones. SDS gels showed the H2A and H3 bands had been replaced with lower molecular weight degradation products consistent with

the presence of a contaminating serine protease. However, sedimentation constants were unchanged for the proteolyzed particles indicating that the overall structure had not been adversely affected by the protease digestion. The degradation did have an impact on the packing of the particles resulting in a much longer c-axis and a sinusoidal wave in the pattern of molecules seen in electron micrographs viewing the plane containing the c-axis. Interestingly, the proteolyzed particles crystallized more easily and formed larger crystals than intact particles. This c-axis corresponded to the face-to-face packing of nucleosome cores more or less along the superhelical axis of the DNA. The path of the DNA around the histone core was modeled into the structure since, at the resolution limit of 20 Å, DNA density could not be distinguished from protein. The electron density map was sufficient to visualize for the first time the wedge shape characteristic of NCPs when seen along the axis orthogonal to the dyad and superhelical axes of the particle which roughly corresponds to the a-axis of the unit cell. The model proposed was correct, but it was for the neutron diffraction studies to provide solid evidence.

A number of small-angle neutron scattering studies had been done on the nucleosome by the time of the Cold Spring Harbor Symposium in 1977 (Pardon *et al.*, 1978; Hjelm *et al.*, 1977). However, the solution studies suffered from the limitations of the technique. The primary limitation is that the structures are rotationally averaged. Analysis of the scattering data involved comparison of calculated scattering curves from hypothetical models with the experimental data. This approach cannot

presume that the correct model has been proposed since there is not necessarily a unique solution. The limitation of resolution to 20 Å made the scattering studies largely inconclusive. By contrast, single crystal neutron diffraction studies were quite well suited to the problem presented by the nucleosome as the data did not suffer from rotational averaging of the solution studies. And, with the approximately equal masses of the DNA and protein in the NCP, the technique of solvent contrast was particularly informative. The scattering cross section of the atomic nuclei of hydrogen and deuterium are sufficiently different that by varying the D_2O and H_2O proportions of the solvent, the average scattering of the solvent can be made to match that of the protein at 39% D_2O and that of the DNA at 65% D_2O. When the average scattering of the solvent matches the average scattering of the protein component of the nucleosome, then the neutron diffraction data is dominated by the scattering of the DNA and the protein, in effect, disappears. The obverse is then true when the average scattering from the solvent matches that of the DNA and the DNA effectively disappears. By collecting data from crystals equilibrated in varying solvent contrasts, the true position of the DNA with respect to the protein core could be determined. It is interesting to note that the phasing of the neutron data did not depend on knowledge of the X-ray structure. Phasing was accomplished by collecting data at four different contrasts. The magnitude of the phase difference between contrasts could be calculated and the sign determined by trial-and-error. The DNA was found to wind in 1.8 turns of a superhelix with a pitch of 27.5 Å and a radius of 42 Å. The

neutron diffraction studies employing contrast variation were quite revealing, and the 16 Å structure (Bentley *et al.*, 1984) provided details unavailable by X-ray diffraction. Not only were the DNA and protein separable by contrast variation but regions of differing hydrophobicity in the protein could be distinguished. The neutron structure clearly showed four separate regions formed by histone dimers. These dimers were interpreted as being (H2A·H2B) or (H3·H4) dimers. The interaction between histones in a dimer, what we now know as the histone fold, was revealed as strongly hydrophobic in nature while the interaction between dimers was seen as weaker or perhaps hydrophilic. The octamer was seen as being helical in structure and in contact with the DNA. Contacts were also seen extending out from the histone core. In particular, H3 contacts with the termini of the DNA represent N-terminal tails of the H3 histone visible at even such a low resolution. The analysis of the structure showed that it possessed a non-crystallographic dyad shared by both the protein and DNA moieties. The symmetry was, however, not complete and the first indication of interparticle contacts in the crystal lattice was demonstrated in the asymmetry of the (H2A·H2B) dimers. Density from one of the (H2A·H2B) dimers was seen to extend beyond the DNA to interact with an adjacent particle.

Subsequent work with material isolated from a variety of sources and with improved crystallization conditions resulted in X-ray diffraction data for unproteolyzed NCP crystals (Finch *et al.*, 1981). These crystals were quite closely related to the original crystals. The space group remained $P2_12_12_1$ but the cell dimensions were now a=111

Å b=193-203 Å and c=111 Å. The c-axis had been reduced from the earlier 340 Å and the asymmetric unit contained a single nucleosome rather than 3 as before, greatly simplifying the crystallography. The variation in the b-axis dimension was reported as being apparently related to the source of the NCPs. We now know that the b-axis, which roughly corresponds to the dyad axis of the particles and is routinely modified by dehydration using polyalcohols to stabilize the crystal and improve diffraction quality. The length of the b-axis is directly proportional to the length of the DNA as will be discussed in detail later in the crystal structure analysis. The analysis of the unit cell showed that the dimensions of the a and b-axes were related by the $\sqrt{3}$ indicating an approximate hexagonal packing along the z-axis.

Concurrent with the work done in Cambridge, an effort was underway at Oak Ridge to solve the structure of the NCP. Bunick and Uberbacher were working with NCPs prepared from chicken erythrocyte nuclei. The nucleosomes were prepared from chromatin solubilized by micrococcal nuclease digestion of whole nuclei and then trimmed to core particles using the same enzyme. The distribution of DNA lengths was narrowed using preparative polyacrylamide gel electrophoresis (Uberbacher & Bunick, 1985a). It was found that crystals formed from NCPs taken from the center of the eluted peak which corresponded to DNA length of 146 bp. It was evident that interaction between the DNA ends on the NCPs was important in the crystal packing.

Crystals of NCPs were obtained using spermine and acetate buffer maintained in a very narrow pH range, ±0.02 pH units, around a pH of 5.21. It is interesting to note

now that the recent structure (Luger *et al.*, 1997) from the Richmond laboratory uses the same pH 5.2 buffer while the structure reported by this laboratory was crystallized at pH 6.0. Crystals formed in the monoclinic space group $P2_1$ with unit cell dimensions of a=168.2 Å b=216.5 Å c=67.4 Å and β= 95.8^o. There were two nucleosomes in the asymmetric unit which in actuality comprised 4 unique half-nucleosomes. The crystals could reversibly convert to a second crystal form by soaking in solutions containing 12% 1,6-hexanediol for about two weeks. The new crystal form was in the monoclinic space group C2 with unit cell dimensions of a=158.0 Å b=108.0 Å c=67.0 Å and β= 95.9^o with only a single half-nucleosome in the asymmetric unit. A remarkable feature of the C2 form was that the dyad axis of the nucleosome lay directly along the c-axis. This has interesting implications for the nature of the symmetry in the NCP, as will be seen later in the Discussion. Complete diffraction data from the NCP crystals were obtained to a resolution of 8 Å using the multi-wire area detector at the University of California, San Diego.

Phasing of the Oak Ridge structure was done entirely by molecular replacement and an interesting application of the then new method of density modification in which a mask of the molecule in the unit cell was determined and all density outside of the model, that is, the solvent, was set to zero. A 25 Å model was constructed using the best information available at the time which was, of course, the structural information already obtained by the Cambridge group. A systematic search of model parameters was made to refine the model on not only positional parameters but also

on features such as DNA diameter and superhelical pitch. Positional refinement was made feasible due to restraints placed on the positions of the nucleosomes within the cell owing to the interconvertibility of the $P2_1$ and C2 crystal forms. Refinement of the model resulted in a nucleosome with a DNA superhelical radius of 43.0 Å and a superhelical pitch of 28.0 Å with 78 base pairs of DNA per superhelical turn. Density modification was used to further refine the model against the data to a resolution of 15 Å. The procedure was very similar to that developed by Bi-Cheng Wang for determining single isomorphous replacement phases (Wang, 1985).

The phasing method engendered some controversy over the correctness of the resulting model. The Cambridge group had used experimental phases which were not subject to model bias. However, Uberbacher and Bunick argued that, given the the level of disorder in the crystals at that time, the accuracy of the measurements of the differences in structure factor amplitudes between native and derivative crystals was not sufficient to warrant the use of derivatives. In particular, as the NCP was a large structure of about 205,000 Daltons, the difference in scattering using even heavy metal cluster compounds was small. This was even more significant for the $P2_1$ data which was used for the Oak Ridge structure solution since there were two nucleosomes in the asymmetric unit making the need for precise measurements of isomorphous differences even more pressing. In the molecular replacement method, the accuracy of the native diffraction could be used to best advantage. The real problem in the case of a molecular replacement solution comes in the very real possibility of model

bias where phases calculated from a hypothetical model overwhelm the information contained in the experimentally obtained amplitudes. The correctness of the model can then be disputed as being corrupted by model bias. It now appears that model bias was not a significant feature of the 16 Å structure as there were several very telling differences between the Oak Ridge structure and the Cambridge structure. Foremost among the differences was the extension of protein density between the DNA gyres which was not seen in the Cambridge X-ray structure but which was evident in the Oak Ridge map as well as the previously discussed neutron map at 16 Å (Bentley *et al.*, 1984). Also, the Cambridge structure contained density associated with histone H2A extending out between the gyres of DNA which would apparently block the exit of the DNA from the particle. This density was not seen in the Oak Ridge structure.

Further refinement of the NCP structure to 8 Å resolution at Oak Ridge (Uberbacher & Bunick, 1989) was accomplished by phase extension, again using methods related to those developed by Wang for single isomorphous replacement. The details of the structure revealed other differences with the Richmond 7 Å structure as well as similarities. The path of the DNA around the histone core was clearly seen and found to be relatively uniformly bent with the exception of positions about 1.5 and 4.5 helical turns from the center of the nucleosomal DNA in either direction, as was seen in the Richmond structure. The minor groove-in positions, where the minor groove contacted the histone core, were seen as regions where there was significant

compression of the minor groove. The most significant feature of the Oak Ridge model was an extension of protein density protruding between the DNA gyres which bears a similarity to density seen in the 16 Å neutron structure but which was not apparent in the Richmond model.

1.1.4 Crystals of NCPs containing defined sequence DNA

The resolution of the early models of the NCP was severely limited by disorder in the crystals. The disorder was presumed to derive from both the random sequences of the DNA and from heterogeneity of the histone proteins caused by variability in post-translational modification of the native proteins. It was obvious that the DNA was important to the quality of the diffraction from NCP crystals but the role of histone heterogeneity was not so clear. While both groups worked to develop NCPs containing defined sequence DNA, the focus of the two groups diverged at this point. Two publications mark this phase of the history of the NCP, one from each of the two groups (Richmond *et al.*, 1988; Harp *et al.*, 1996). While the two were published nine years apart, they represent research programs which were very much in competition, and mark an unpleasant episode in chromatin history.

The European group, now led by Timothy Richmond who had moved from Cambridge to the ETH in Zurich, felt that protein heterogeneity was important and pursued a program of cloning and expression of the core histones in bacteria. Recombinant

histones were also believed to be important by the Richmond group as a means of producing new heavy atom derivatives. The random sequence DNA was replaced by a DNA fragment of a 5S RNA gene from the sea urchin, *Lytechinus variegatus*, which had been developed in the laboratory of Robert Simpson at NIH (Simpson & Stafford, 1983). Diffraction from crystals containing the 5S RNA DNA fragment and native chicken erythrocyte histones was reported to be anisotropic with reflections extending to 3.0 Å on the c-axis and only 5.0 Å on the b-axis (Richmond *et al.*, 1988). The reasons for the anisotropy will be discussed later and will by important in understanding the differences in the results coming from the two competing laboratories.

The Oak Ridge group felt that histones derived from chicken erythrocyte nuclei were reasonably homogeneous and chose instead to pursue a strategy of improving the DNA. As the mass of protein and the mass of the DNA in the NCP is roughly equal, the choice was important. The improvements in the DNA were 2-fold. First, the phasing of the DNA on the histone core had to be very precise. A strong effort was made to understand the DNA sequence effects on nucleosome phasing (Uberbacher *et al.*, 1988). And second, the Oak Ridge group developed the concept of using a DNA palindrome to extend the 2-fold symmetry of the histone core to the DNA which would result in a 2-fold symmetric nucleosome. This would eliminate the 2-fold disorder caused by the indeterminacy of packing of an asymmetric particle into the crystal lattice. Much of the development of NCP crystals containing a defined sequence DNA palindrome forms the subject of the present work and will be discussed

in detail later. This portion of the work required several years of effort to identify candidate sequences through an understanding of the effect of DNA sequence on the phasing of the DNA on the histone core and confirmation of the phasing of cloned DNA in reconstituted NCPs by DNAse I footprinting. A palindrome based on one-half of the primary candidate sequence was constructed and methods were developed to produce the palindrome fragment in large quantities for reconstitution of NCPs.

The sequence chosen for the production of the DNA palindrome was obtained fortuitously. A post-doctoral researcher, Dr. Loren Hauser, had entered the laboratory of Donald Olins at Oak Ridge. Hauser had come from the laboratory of Dr. Barbara Hamkalo at the University of California, Irvine, and had worked on α-satellite DNA sequences from a variety of primates including humans. As part of his dissertation work (Hauser, 1985), he had sequenced a *Bam*H I repeat of α-satellite DNA from the human X chromosome (Yang *et al.*, 1982). Hauser communicated this work to the Bunick group and the sequence was examined using the computer algorithm developed by the Oak Ridge group to predict phasing of nucleosomes from the distribution of dinucleotides and the symmetry of the dinucleotide distribution about putative phasing centers (Uberbacher *et al.*, 1988). The computer results predicted 12 equally strong phasing positions, which agreed with mapping results done by Hauser and a clone of the *Bam*H I repeat was obtained from Hamkalo's laboratory. At that time, the technology of PCR cloning was not readily available. The palindrome had to be constructed using standard cloning technology. Subcloning of the palindrome made

use of an *Alu* I site occurring near one of the predicted nucleosome centers. The *Alu* I cleaved half-nucleosome DNA fragment was ligated to a commercially available *Eco*R I linker. The resulting half-palindrome fragment could then be ligated to itself to form a perfect 146 bp DNA palindrome. An important aspect of this is that the nucleosome phasing sequence used was chosen from 12 possible phasing sites. The resulting α-satellite DNA palindrome was thus 1 of 24 possibilities and was chosen based on the existence of an *Alu* I site in an appropriate location for sub-cloning.

Once the α-satellite DNA palindrome was constructed and cloned it had to be produced in large quantities. This aspect of the project consumed about two years of work. A 146 bp DNA palindrome is essentially a metastable intermediate in solution with the 73 bp hairpin form being the energetically preferred conformation. This meant that purification protocols had to be modified to prevent "snap-back" of the halves of the DNA duplex to form hair pin structures. Heroic efforts were made to clone multiple copies of the DNA palindrome into a vector for large-scale production in bacteria with no success. Early production efforts were forced to make use of very large-scale production of the single copy plasmid. One such effort made use of pilot plant facilities in the Chemical Technology Division of Oak Ridge National Laboratory. A 400 liter fermentation run was performed using the plasmid containing a single copy of the DNA palindrome. This production run failed to produce even a fraction of the predicted yield of DNA palindrome. Subsequent experiments determined that production of DNA palindrome in bacteria was poisoned by the transient formation

of cruciform structure in the plasmid during replication as discussed in detail in the Results chapter. Systematic efforts were then made to find an alternative method for large-scale production of the DNA palindrome. The successful protocol was developed by Elise Palmer (Palmer *et al.*, 1996), who cloned multiple copies of the half-palindrome fragment into a vector. The plasmid containing as many as 32 copies of the half-palindrome was replicated in bacteria without the problems associated with the palindrome and the half-palindrome fragment could be produced in large quantity. The half-palindrome was then ligated in very large reactions to provide the required quantities of α-satellite DNA palindrome.

Reconstitution of NCPs using the α-satellite DNA palindrome was accomplished easily but was found to still suffer from the existence of multiple phases. The symmetrically phased form was identified on polyacrylamide gels as the resistant band in a time course of micrococcal nuclease cleavage. The symmetrical phasing meant that both DNA termini were protected from nuclease attack while the so called degenerate phases possessed asymmetrically bound DNA with one terminus exposed to varying degrees. Since the degenerate phases could be separated from the correctly phased form on non-denaturing polyacrylamide gels, a strategy for purification of the correctly phased NCPs was developed (Harp *et al.*, 1995) using preparative polyacrylamide gel electrophoresis based on earlier work done in the Oak Ridge laboratory (Uberbacher *et al.*, 1986). The preparation of large quantities of purified NCPs containing only precisely symmetrically bound α-satellite DNA palindrome was thus the culmi-

nation of many years of single-minded effort on the part of all of the members of the Oak Ridge group.

Crystallization of the purified NCPs containing α-satellite DNA palindrome and chicken erythrocyte histones resulted in crystals diffracting isotropically to 3.0 Å with reflections seen to 2.8 Å using the rotating anode X-ray source available in-house. The strategy to produce improved diffraction from crystals of NCP crystals by concentrating on improvement of the DNA moiety was successful (Harp *et al.*, 1996). The Oak Ridge group has now completed the structure of the NCP to 2.5 Å (Harp *et al.*, 2000a). At the time of this writing, the structure is the most complete and highest resolution picture of the NCP available and forms the central subject of this dissertation.

1.2 Macromolecular crystal annealing

An entirely unexpected discovery was made during the course of this research and was given the name, "macromolecular crystal annealing." The discovery was important to technical issues involved in the collection of diffraction data from nucleosome core particle crystals and so is germane to this thesis. The significance of the discovery to the field of crystallography was deemed sufficient to warrant inclusion here of an investigation and discussion of the phenomenon of annealing and its application to a wide variety of crystal systems. This elaboration is useful as there was no hint of

the existence of this phenomenon prior to this work and thus no precedent to which the reader may be referred for understanding its physical basis and the reasons for its use.

Crystals of the nucleosome core particle are quite sensitive to radiation damage at ambient temperature, making cryogenic data collection essential. The development of techniques for cryogenic data collection has been an important advance in the field of macromolecular crystallography (Garman & Schneider, 1997; Rodgers, 1997; Garman, 1999). Protection of macromolecular crystals from radiation damage at cryogenic temperatures allows collection of complete and even multiple complete data sets from a single crystal, and is especially important for collection of multiple wavelength anomalous dispersion (MAD) data. The low temperature of the flash-cooled crystal can also reduce temperature factors, B-factors, depending on the static disorder of the molecules in the crystal and can thereby increase the intensity of diffraction and the quality of the resulting molecular model. The only significant disadvantage of cryocrystallography has been increased mosaicity in some crystals after flash-cooling. The scope of this problem was illustrated in a study of 19 crystal systems (Rodgers, 1994). In that study, only 8 of the 19 crystal types exhibited little or no mosaicity increase after flash-cooling, while 8 of the remaining 11 were found to exhibit a doubling of rocking curve widths (mosaicity) after flash-cooling. In most cases the advantages of cryogenic data collection far outweigh even a doubling of mosaicity. However, a substantial increase in mosaicity coupled with large unit cell dimensions can cause

overlapping of Bragg reflections and result in unusable diffraction data. Even when the mosaicity is within acceptable limits, lowering mosaicity will improve diffraction data quality. An increase in mosaicity increases the width of a Bragg reflection thereby making integration of the peaks in diffraction data less accurate by reducing the signal-to-noise ratio.

1.2.1 What is mosaicity?

Mosaicity is the name associated with a diffraction data quality statistic which relates the dimensions of a reciprocal lattice point to the structure and perfection of the crystalline lattice. With that said, it will be necessary to review the fundamentals of diffraction from a crystal lattice. All of X-ray crystallography is built on the first principle of Bragg's Law, which describes the conditions necessary to the development of a diffraction peak, or Bragg reflection, that can then be recorded. A peak of intensity in an X-ray diffraction image from a crystal comes about as the result of constructive interference of wavelets scattered by electrons arranged in a regular, repeating pattern within the lattice. A diffraction peak may be found at the angle, with respect to the planes defined by the repeating pattern, at which the reflected wavelets leave the crystal in phase. This corresponds to the angle for which the difference in pathlength through the crystal for any two wavelets is an integral multiple of the wavelength. Thus the distance between two diffracting planes, d, in Ångtrom

units, is related to the angle of diffraction, θ, by the classical equation of Bragg's Law

$$2d \sin\theta = n\lambda$$

where λ is the wavelength of the radiation and n is any integer. The Bragg peaks are then related to the angle of a series of regularly repeating, parallel diffracting planes to the incident X-ray beam by the inverse of the distance separating the planes.

An important construct for the understanding of diffraction data is the reciprocal lattice, which is defined as a lattice of points in reciprocal space (which is itself sometimes facetiously defined as 1 divided by the real world). While the construction of the reciprocal lattice may appear at first to be a mathematical trick making calculations easier, it is in fact a real and useful way to visualize diffraction. If the incident X-ray beam is represented as the wave vector, $\mathbf{s}_0$, and the diffracted wave, $\mathbf{s}$, each with magnitude $1/\lambda$, is regarded as having been reflected at an angle θ, then the resultant wave vector, $\mathbf{S}$, is given as

$$\mathbf{S} = \mathbf{s} - \mathbf{s}_0$$

The direct space unit cell with axes a, b, and c is transformed in the reciprocal space unit cell with axes a*, b*, and c* first by noting the properties of $\mathbf{S}$. A discussion of these may be obtained from basic texts (Drenth, 1994). The vector $\mathbf{S}$ is perpendicular to the diffracting planes and for the basic sets of planes defining the unit cell, (100), (010), and (001) we have the the vectors $\mathbf{S}(100)\mathbf{S}(010)$, and $\mathbf{S}(001)$, which then correspond to the axes of the reciprocal space unit cell. That is, for the special set of

planes (100) with Miller indices h=1, k=0, and l=0, the vector **S**(100) is perpendicular to the (100) planes with a magnitude of 1/d, which is the reciprocal of the d spacing of the direct space planes. The same construction can be made for any given set of diffracting planes and the points representing the ends of the **S**(hkl) vectors are arranged in a lattice, the reciprocal space lattice. A further construct of worth is the Ewald sphere which can be used to determine when a given reciprocal lattice point is in the diffracting condition. The Ewald sphere is constructed as a sphere of radius $1/\lambda$ with the origin of the reciprocal lattice on its surface and the center of the sphere on the incident beam. Any reciprocal space point in contact with the surface of the Ewald sphere is in diffracting condition. Thus, as the crystal moves, the reciprocal space lattice moves and reciprocal lattice points move into and out of diffracting condition. For the rotation method of data collection used throughout this study, the construction just described is quite useful as it provides a description of the diffraction image in all 3 dimensions. In rotation photography, a crystal in the X-ray beam is rotated about an axis, ϕ, through a given angular range while the data is collected on the image plate. Thus, the rotation photograph records all reciprocal space lattice points falling on the Ewald sphere as the reciprocal lattice was rotated, along with the crystal, through the ϕ angle defining the rotation. The two dimensions parallel to the plane of the image plate area detector are easily recorded in the image. However, the other dimension perpendicular to the image plate can only be recorded as the ϕ angle through which the reflection remains recordable in the image.

Mosaicity can affect all three dimensions of a reciprocal space lattice point by broadening of the point in two basic ways. One way is sometimes referred to as a micromosaic contribution and is directly related to the quality of the packing of the macromolecules in the crystal lattice. That is, if there is some underlying disorder, such as tail regions that interfere with the regularity of the packing of the molecules, this would affect the regularity of the diffracting structures within the lattice and contribute, in general, a small amount, to the broadening of the reciprocal lattice point. The more classical treatment of mosaicity involves a macromosaic contribution. In this mode, the crystal is considered to be composed of many smaller perfect crystals or mosaic blocks. Each mosaic block may be composed of hundreds or thousands of unit cells and the micromosaic of each block may be quite small or even zero. However, the mosaic blocks are misaligned with respect to each other to some degree. The degree of the average misalignment then is the macromosaic contribution to the mosaicity and may represent 0.2 to 0.5 degree of mosaicity in a normal crystal (Drenth, 1994). Mosaicity can then be conceived of as the inflation of a reciprocal space lattice point from the theoretically ideal dimensionless point to an ellipsoid of dimensions determined by such effects as mosaicity. Unfortunately, the measurement of mosaicity is complicated by such factors as the divergence of the X-ray beam used in the diffraction experiment and the efficiency and resolution of the detector. Correct integration of diffraction peaks from image plate data using the rotation method depends on the correct estimation of the mosaicity of the crystal in order to calculate the precise

dimension of a reciprocal lattice point in terms of its ϕ range. Larger mosaic spread in a crystal resulting in excessive inflation of a reciprocal lattice point means that it may be recorded over multiple images, reducing the accuracy of the integration for what are termed "partially recorded reflections."

In summary, mosaicity of a crystal can be defined as the sum of those factors which result in inflation of the dimensions of a reciprocal space lattice point. This means that the point remains in reflecting condition longer as the point passes through the Ewald sphere and has as its primary feature a reduction of the signal to noise ratio of the data. The primary contributor to the mosaicity of a macromolecular crystal is the misorientation of mosaic blocks with the crystal. These blocks may be the result of any number of defects in the crystal resulting from such things as the inclusion of impurities or packing faults caused by changes in solution conditions or temperature during crystal growth. It is this contribution to the mosaicity of the crystal which is believed to be impacted in the phenomenon of macromolecular crystal annealing.

1.3 Glossary

- *CaH* Accessory helix C-terminal to the histone fold domain
- *CH* The C-terminal α helix of the histone fold motif
- *DCAM* Acronym for Diffusion-controlled Crystallization Apparatus for Microgravity

- *(H2A·H2B)* Heterodimer formed by the association of one subunit of histone H2A and one subunit of histone H2B.
- *$(H3{\cdot}H4)_2$* Heterotetramer formed by the association of 2 subunits each of histones H3 and H4.
- *MPD* 2-methyl,2,4-pentanediol
- *NaH* Accessory helix N-terminal to the histone fold domain
- *NCP* Acronym for the nucleosome core particle.
- *NH* The N-terminal α helix of the histone fold motif
- *NL* Loop connecting the NH and mH α helices of the histone fold motif
- *PIP* di-μ-iodobis(ethylenediamine)diplatinum
- *dorsal* That side of the NCP containing the DNA half with 72 base pairs lateral to the central base pair.
- *mH* The median α helix of the histone fold motif
- *nucleosome*: The defining structure of chromatin. The nucleosome is composed of histone proteins and DNA in a compact structure. The structure is divided into a histone octamer core, a core region of DNA of about 146 base pairs, a variable lengthlinker region of DNA, and a lysine-rich linker histone. The histone octamer core is composed of 2 subunits each of the 4 core histones, H2A, H2B, H3, and H4.

- *nucleosome core particle*: This substructure of the nucleosome is defined by protection of the DNA from digestion with miccrococcal nuclease by the histone octamer core. It consists of the histone octamer core and 145-147 base pairs of DNA wound in about 1.75 left-handed superhelical turns around the histone octamer. The structure lacks the linker region of the DNA and the lysine-rich linker histone.

- *mGI* Acronym for minor groove-in position.

- *minor groove-in position*: A point at which the DNA minor groove faces the surface of the histone protein core of the nucleosome is termed a minor groove-in position, mGI. These positions are numbered 1 to 7 symmetrically around the nucleosome from the pseudodyad

- *ventral* That side of the NCP containing the DNA half with 73 base pairs lateral to the central base pair.

Chapter 2

Methods

2.1 Preparation of histones from chicken erythrocyte nuclei.

In general terms, histone octamer was prepared by gentle salt extraction from soluble chromatin prepared from chicken erythrocyte nuclei. Core histones were separated from DNA using a linear NaCl gradient in hydroxyapatite chromatography. Stable octamer was then reassembled by dialysis against 2 M NaCl and then purified by gel filtration on Superose 12.

2.1.1 Buffers

All buffers were prepared the day before the preparation and the pH of each adjusted at the required temperature. A fresh stock solution of 100 mM phenylmethylsulfonyl flouride, PMSF, was prepared in anhydrous isopropanol for each preparation. PMSF was not added to the buffer solution until immediately prior to use since the enzyme inhibitor is unstable in aqueous solution.

Wash buffer, 4 liters:

- 15 mM Tris-Cl
- 15 mM NaCl
- 60 mM KCl
- 0.5 mM spermidine
- 0.15 mM spermine
- 0.34 M sucrose
- 2 mM EDTA
- 0.5 mM EGTA
- 15 mM β-mercaptoethanol
- 1 mM benzamidine

- 0.2 mM PMSF

The pH was adjusted to 7.4 with HCl at 4°C and β-mercaptoethanol and PMSF were added just before use.

Lysis buffer, 4 liters:

The lysis buffer was the same as the wash buffer with the addition of 0.5% Nonidet P-40 detergent.

Digestion buffer, 0.5 liter at 4°C and 0.1 liter at 37°C:

- 50 mM Tris-Cl, pH 7.4
- 25 mM KCl
- 4 mM $MgCl_{(}2)$
- 1 mM $CaCl_{(}2)$
- 1 mM benzamidine
- 0.2 mM PMSF

Extraction buffer, 2 liters, 4°C:

- 10 mM Tris-Cl, pH 8.0
- 0.25 mM EDTA

- 1 mM benzamidine
- 0.2 mM PMSF

Finish buffer, 2 liters, 4°C:

- 10 mM Tris-Cl, pH 8.0
- 1 mM benzamidine
- 0.2 mM PMSF

2.1.2 Detailed protocol for preparation of soluble chromatin

Soluble chromatin was prepared in one day. Chickens were obtained from the Cherokee Poultry Unit of the University of Tennessee School of Agriculture. Chickens were isolated in holding cages on the previous night and transported to the laboratory on the morning of the preparation. Two hens were used for each preparation. Each chicken was anesthetized with Metofane and the neck feathers removed. A window of skin was excised to exposed the jugular vein. The fascia covering the vein was then carefully cut using a number 11 scalpel blade, the vein lifted and completely exposed. The chicken was then held with the vein positioned over a plastic beaker containing 5,000 units of heparin in 2 mL of Wash buffer and 0.5 mL of 100 mM PMSF in isopropanol. An assistant was required to hold the body of the chicken. It was important to position the head of the chicken away from the beaker such that

any regurgitation did not contaminate the collected blood. When correctly and stably positioned, the vein was severed using a scissors and the blood allowed to flow into the beaker. The blood was then filtered through sterilized cheese cloth, to remove any clots, into 40 mL Oak Ridge tubes on ice. The blood was then centrifuged at 5000 rpm for 10 minutes in a Sorvall SS34 rotor or equivalent. The tubes were carefully removed from the rotor to prevent shifting of the packed cells. A critical point in the procedure was removal of the " buffy coat" of white cells which cover the packed erythrocytes. The buffy coat was carefully and completely removed by aspiration. For this, a 9 inch glass pasteur pipette was attached with tubing to a side arm flask connected to house vacuum with mild suction applied. The erythrocytes were then suspended in Wash buffer and transferred to 500 mL polycarbonate bottles fitting the Sorvall GSA rotor. The cells were aliquoted into six bottles with Wash buffer and spun at 5000 rpm for 10 minutes. The supernatant was decanted and the cells resuspended in Wash buffer on ice and spun. The wash was repeated twice for a total of three washes.

The washed cells were then suspended in Lysis buffer and spun at 5000 rpm for 10 minutes at 4°C. The nuclear pellet often required treatment at low speed using a Tissuemizer but could often be resuspended by pipetting up and down with a 10 mL plastic pipet. The nuclear pellet was washed twice with Lysis buffer spinning each time at 5000 rpm for 10 minutes. The number of bottles used to contain the material was reduced by 2 at each wash until at the last wash only 2 bottles are

used. The nuclear pellet was then suspended in Digestion buffer on ice and spun at 5000 rpm for 10 minutes. The nuclei were then resuspended in Digestion buffer and consolidated into a single bottle and spun. The supernatant was completely decanted and the nuclei resuspended in 20 mL of Digestion buffer adjusted to pH 7.4 at 37 C. The quantity of DNA in the pellet was approximated by dissolving 10 μL of 0.1 N NaOH and measuring absorbance at 260 nm. The reading was multiplied by 100 for the dilution factor and then by 33 mg/mL for single stranded DNA. Total DNA was then determined by multiplying the concentration determined by spectrophotometry by the total volume, 20 mL. The amount of micrococcal nuclease required was 40 units for each mg of DNA. Micrococcal nuclease from Worthington Biochemicals was prepared by addition of 1 mL of 1 mg/mL of bovine serum albumin to the lyophilized enzyme. The correct amount of enzyme was added to the nuclei in the bottle and incubated at 37°C with gentle stirring for 30 minutes. The reaction was stopped by addition of 0.4 mL of 0.5 M EDTA and the bottle spun at 5000 rpm for 10 minutes. The supernatant was decanted and the nuclei transferred to a 15 mL Dounce homogenizer on ice. The nuclei were then suspended in 15 mL of Extraction buffer using the B pestle and the suspension transferred to a 40 mm Spectapore 1 dialysis tube. The suspension was dialyzed overnight at 4°C against 2 liters of Extraction buffer. The suspension was further dialyzed against 2 L of Finish buffer to remove EDTA in preparation for hydroxyapatite chromatography. Soluble chromatin was harvested by centrifugation of the suspension in a 40 mL teflon Oak Ridge tube at

10000 rpm for 10 minutes. The soluble chromatin supernatant was decanted into a sterile tube and the quantity of chromatin was assessed by absorption at 260 nm using an extinction of 1 AU for 10 mg/mL of chromatin. The quality of the preparation was determined by SDS-PAGE. Coomassie Blue-stained gels were scanned on a Molecular Dynamics Computing Densitometer Model 300A and the stoichiometry of the histone preparation was compared to a standard prepared by solubilizing total protein from whole chicken erythrocyte nuclei. Peaks were fitted using ImageQuant (Molecular Dynamics) and PeakFit (Jandel Scientific) software.

2.1.3 Hydroxyapatite chromatography

Octamer was separated from DNA using ceramic hydroxyapatite (Bio-Rad Macro-Prep Type II, 20 μm particle size) packed in a 16 mm x 10 cm column (Pharmacia-LKB HR16/10). The column was equilibrated with 10 mM Na phosphate, pH 7.4. The column was loaded with approximately 100 mg of chromatin and washed with 10 mM Na phosphate buffer at a flow rate of 3 mL/min. Very lysine rich histones, H5, H1, and H1^o, were eluted with 0.7 M NaCl, 10 mM Na phosphate. A second peak containing (H2A·H2B) dimer was obtained after the linker histones and was collected. Octamer was then eluted using 3 M NaCl, 10 mM Na Phosphate in a bimodal peak. The column was regenerated using 0.4 M Na phosphate, pH 6.8, to remove all DNA. The soluble chromatin was processed in multiple runs and all octamer peaks collected and pooled. The quality of the material was again assessed by

SDS-PAGE. The octamer was then concentrated using Centriprep 10 and Centricon 10 microconcentrators to 20 mg/mL as determined using an extinction of 0.43 AU at 280 nm. The octamer was reassembled by dialysis against 2 M NaCl, 20 mM Tris-Cl, 0.1 mM EDTA, 0.1 mM PMSF, pH 8.0, in preparation for final purification by gel filtration on Superose 12.

2.1.4 Final purification on Superose 12

The final step of purification was gel filtration on a prepacked Superose 12 in an HR 10/30 column (Pharmacia-LKB) equilibrated in 2 M NaCl, 20 mM Tris-Cl, 0.1 mM EDTA, 0.1 mM PMSF, pH 8.0. The octamer peak was collected and pooled from multiple runs then concentrated using Centriprep 10 and Centricon 10 microconcentrators to 20 mg/mL. Only a single pass of the octamer over the Superose 12 column was done. A dynamic equilibrium exists between the octamer and a dimer-hexamer pair (Stein & Page, 1980) which can result in dimer-depleted histone stocks if gel filtration chromatography is used for purification. A final check for possible proteolysis and assessment of stoichiometry was done on the finished product. The histone octamer stocks were then stored at -20°C until used for reconstitution of nucleosomes or for crystallization of the histones.

2.2 Expression and purification of recombinant histones.

Clones of each of the four core histones were obtained from Dr. Venkatraman Ramakrishnan of the University of Utah (now at the Medical Research Council, Cambridge, England). The histone genes from chicken had been cloned into the *Nde* I/*Bam*H I site of a vector related to pET13. The clones were sequenced by transforming *E. coli* DH5α cells and growing the plasmid in 1 liter cultures using LB broth and 50 μg/mL kanamycin. The plasmid was purified from the cell pellet using a standard alkaline lysis protocol, and final purification of the plasmid by ion exchange chromatography on a Mono-Q column (Pharmacia-LKB). The cloned histone genes were then sequenced by PCR using a T7 promoter primer and large dye terminator sequencing reagents according to the manufacturer's protocol (ABI Prism). The sequencing reaction products were run on an ABI Prism 377 DNA Sequencer. The DNA sequences of the histones were translated using GCG software and the amino acid sequences compared to published sequences (van Holde, 1988).

2.3 Assembly of recombinant octamer.

Lyophilized histones were dissolved in 8 M guanidine, 20 mM K acetate, pH 5.2, and concentrations determined by UV absorbance at 280 nm using extinction coeffi-

cients taken from van Holde (van Holde, 1988). Histones were combined in equimolar amounts to result in approximately 1 mg/mL final concentration of histone octamer. Stoichiometry of the histones was checked by SDS-PAGE. The mixture was then dialyzed extensively against 3 liters of 2 M NaCl, 20 mM Tris-Cl, 1 mM EDTA, 1 mM β-mercaptoethanol, 0.2 mM PMSF with 3 changes of buffer at 4°C. The octamer was then harvested and concentrated to approximately 20 mg/mL using Centricon 10 microconcentrators. The assembled octamers were then purified using Superose 12 in an HR 10/30 column equilibrated with 2 M NaCl, 20 mM Tris-Cl, 1 mM EDTA, 1 mM β-mercaptoethanol, 0.2 mM PMSF. Peak fractions were pooled and concentrated to 20 mg/mL using Centricon 10 microconcentrators and stored at -20°C until used.

2.4 Identification of candidate DNA sequences.

A *Bam*H I repeat taken from the centromere region of the human X chromosome was provided by Dr. Loren Hauser. The approximately 2000 bp fragment contained 12 nucleosome phasing sequences in repeats of about 170 bp which were initially detected by restriction enzyme analysis (Yang *et al.*, 1982) and later mapped and sequenced (Hauser, 1985). The full sequence was subjected to computer analysis using software developed to examine likelihood of nucleosomal phasing based on correlation of dinucleotide distribution to a matrix of dinucleotide frequencies in known nucleosomal

sequences. The algorithm also examined symmetry in dinucleotide distribution about the center of a putative nucleosomal site (Uberbacher *et al.*, 1988). Each of the 12 nucleosomal sites were predicted by computer analysis and corresponded to mapped positions. The strength of correlation to the dinucleotide distribution matrix and the degree of symmetry was roughly equal for each of the nucleosomal positions.

2.5 Footprinting analysis of phasing of the ASN DNA on histone octamer cores.

The binding of histone octamer to the α-satellite native (ASN) DNA fragment was verified by DNase I footprinting. In general terms, footprint analysis examines the protection of DNA by bound protein from digestion with the endonuclease DNase I. This protection pattern is revealed by digesting a protein-DNA complex with enzyme under conditions in which each DNA molecule is cut, on average, only once. The frequency of cutting at each position on the DNA molecule is determined by high-resolution PAGE. Indirect end labeling allowed determination of the length and identity of each band from a single ^{32}P labeled end by alignment with a lane of the same DNA cut at guanine residues using chemical scission. Although DNase I is generally a non-specific endonuclease, it does exhibit some sequence preferences due to the variable presentation of phosphodiester linkage in A/T- versus G/C-rich fragments. It is thus necessary to subtract the cutting pattern of a control lane of DNase I

digest of naked DNA from the DNase I digest of the protein-DNA complex to provide a precise protection pattern revealing the position of the bound protein on the DNA.

2.5.1 Indirect end labeling

Indirect end labeling involves the use of two different restriction enzyme recognition sequences to define the ends of the DNA fragment to be analyzed. The DNA restriction fragment is first labeled using T4 polynucleotide kinase to transfer a ^{32}P label onto both 5′ ends of the fragment. A restriction enzyme cut positioned adjacent to one end of the DNA fragment is then used to remove one end label leaving the remaining portion of the fragment labeled with a single ^{32}P at the other end. This provides the advantage of quantifiable autoradiography to analyze the resulting pattern since each fragment is labeled with one and only one radioactive atom at a known position.

For the DNase I footprint analysis of the ASN nucleosome core particles, the plasmid pνAJE containing a single copy of the ASN sequence was cut adjacent to the nucleosomal DNA in the multiple cloning region using *Bam*H I and then treated using bacterial alkaline phosphatase. Labeling was accomplished by adding 0.2-0.5 μg of DNA in 3 μL of glass distilled water. The reaction was then assembled as follows:

- 3 μL DNA
- 2 μL 5X kinase buffer
- 1 μL 0.1 M ATP

- 1 μL 10 mM spermidine
- 2 μL γ-^{32}P ATP (ICN crude fraction)
- 1 μL T4 polynucleotide kinase

5x kinase buffer

- 350 mM Tris-Cl, pH 7.6
- 20 mM $MgCl_2$
- 10 mM KCl
- 4 mM dithiothreitol

The reaction was incubated at 37°C for 30 minutes and stopped by precipitation with two volumes of ethanol. Incorporation of label was checked by spotting the reaction mixture on DE-81 paper (Whatman). Two samples were spotted; one for total counts and one for incorporated counts. The sample used for measurement of total counts incorporated into the DNA was washed with 0.5 N NaOH and then dried with ethanol. Both samples were placed in scintillation vials with 10 mL Scintiverse scintillation cocktail and counted in a scintillation counter using the appropriate discriminator window settings for ^{32}P. The DNA fragment was then digested with *Bst*N I restriction enzyme to remove the unwanted labeled end. Addition of labeled DNA in further reactions was followed using a hand-held Geiger-Muller counter.

2.5.2 Construction of marker fragments by guanine base-specific chemical cleavage

Markers for identification of fragments resulting from DNase I digestion were constructed by limited chemical cleavage of the nucleosomal DNA fragment at guanine residues (Maxam & Gilbert, 1980). In this reaction, base specificity is provided by dimethyl sulfate (DMS) which methylates guanines at N7 thereby opening the purine ring. Piperidine releases the ring fragments from the ribose moiety and then catalyzes β-elimination of phosphates from empty sugars to provide strand scission. As the reaction is not provided in commercially available kits, the full protocol is given here.

- Place 5 μL of ^{32}P end labeled DNA in glass distilled water in a silanized Eppendorf tube.
- Add 200 μL of DMS buffer and incubate on ice for 5 minutes.
- Add 1 μL of DMS and mix.
- Incubate reaction at 20°C for 5 minutes.
- Add 50 μL of DMS stop buffer and mix.
- Precipitate with 750 μL of ethanol and mix by inverting tube 4 times.

- Incubate at -80°C for 5 minutes and spin down DNA in Eppendorf centrifuge at 4°C for 10 minutes.
- Decant supernatant into 5M NaOH and dry pellet in vacuum.
- Dissolve pellet in 225 μL glass distilled water and add 25 μL 3M Na acetate, pH 5.5, and 750 μL ethanol, and mix by inverting tube 4 times.
- Incubate at -80°C for 5 minutes and spin down DNA in Eppendorf centrifuge at 4°C for 10 minutes.
- Decant supernatant into 5 M NaOH and dry pellet in vacuum.
- Dilute 100 μL of piperidine with 900 μL of glass distilled water.
- Add 100 μL of diluted piperidine to the DNA pellet, vortex, and quick-spin in an Eppendorf centrifuge.
- Place a piece of teflon tape over the opening of the Eppendorf tube and close the cap. Clamp the cap in place.
- Incubate the reaction at 90°C for 30 minutes.
- Freeze the reaction in an ethanol/dry ice bath and dry on a SpeedVac.
- Dissolve the pellet with 10 μL glass distilled water and dry on a SpeedVac. Repeat.
- Dissolve the pellet in gel loading buffer.

Buffers

DMS buffer

- 50 mM sodium cacodylate, pH 8.0
- 1 mM EDTA

DMS stop buffer

- 1.5 M sodium acetate, pH 7.0
- 1.0 M β-mercaptoethanol
- 100 μg/mL tRNA

2.5.3 DNase I digestion of naked and bound DNA

Nucleosome core particles containing the ^{32}P end labeled DNA were reconstituted using cold DNA to bring the DNA concentration to the appropriate level and assembled by a stepped salt gradient using plastic crystallization buttons covered with Spectrapore 1 dialysis membrane. The assembled nucleosome core particles were separated from labeled, free DNA by gel filtration chromatography using Sephacryl S-400HR in an HR5/5 column. The pooled peak fractions were concentrated using Centricon 30 microconcentrators.

DNase I reactions were performed in 50 μL volumes with 1 mM $MgCl_2$ and an empirically determined dilution of enzyme for each substrate. Unlabeled nucleosomal DNA was used to bring the amount of DNA to 4 μg for the free DNA reaction. The reactions were incubated at 37°C for 30 minutes, stopped by addition of 1 μL 0.3 M EDTA, 6 μL 10% sodium dodecyl sulfate, 3 μL 1% proteinase K, and then incubated for an additional 40 minutes at 37°C. The reaction products were precipitated with 3 volumes of cold ethanol overnight. After centrifugation and drying, ethanol pellets were dissolved in gel loading buffer for denaturing 8% polyacrylamide sequencing gels (Maxam & Gilbert, 1980).

2.5.4 Analysis of the protection pattern

The DNase I reaction products were analyzed by high-resolution PAGE on 8% denaturing polyacrylamide gels along with the marker fragments. The gels were then dried and autoradiographed. The autoradiographs were scanned and digitized using a Molecular Dynamics Model 300A computing densitometer.

2.6 Large scale purification of 146 bp α-satellite DNA palindrome fragments

Large-scale fermentation was done using a 14 L New Brunswick benchtop fermentor. Fermentation was done in 10 L batches with Terrific broth (TB) (Sambrook *et al.*, 1989) using 100 μg/mL ampicillin for antibiotic selection. All batches were started from colonies plated from the same stock of frozen cells which had been transformed with plasmid in which the palindrome had been verified by sequencing. The host cells used for all cloning and preparation procedures were Epicurian Coli ® SURE® cells (Stratagene Cloning Systems, La Jolla CA). These cells were necessary to prevent loss of the large palindrome since they are deficient in homologous recombination with the genotype mcrCB-hsdSMR-mrr)171 endA1 supE44 thi-1 gyrA96 relA1 lac recB recJ sbcC. Fermentation runs were begun with 1 L of overnight culture and allowed to grow until the turbidity of the culture reached an absorbance of 10-12 at 600 nm. Cells were harvested in an Heraeus 300 ml capacity continuous flow rotor (Heraeus Instruments, South Plainfield NJ) operated at 12,800 rpm at 5°C. The cell paste was resuspended in 5 L of cold 10 mM Tris-Cl, 50 mM EDTA, pH 8.0. Cells were then lysed by addition of 5 L of 1% SDS, 0.2 N NaOH with gently stirring. The solution was then neutralized by addition of 5 L of cold 3 M potassium acetate, pH 5.5. The lysate was cleared using the Heraeus continuous-flow centrifuge rotor operated at 12,800 rpm and 5°C. The cleared lysate was precipitated with 0.7 volumes

```
EcoR V                                                                  EcoR I
ATCAATATCCACCTGCAGATTCTACCAAAAGTGTATTTGGAAACTGCTCCATCAAAAGGCATGTTCAGCGG
TAGTTATAGGTGGACGTCTAAGATGGTTTTCACATAAACCTTTGACGAGGTAGTTTTCCGTACAAGTCGCCTTAA
```

Figure 2.1: Sequence of the *Eco*R V/*Eco*R I fragment, 1/2ASP. Ligation of the 1/2ASP fragment at the *Eco*R I site resulted in construction of the 146 bp ASP DNA palindrome.

of isopropanol.

2.7 Large scale production of ASP DNA by ligations of half palindromes

Large scale production of the 146 bp ASP DNA palindrome was accomplished by first producing large quantities of the the *Eco*R I/*Eco*R V fragment, 1/2ASP, representing half of the full length palindrome (see Figure 2.1). The 1/2ASP fragment was then ligated in a large reaction, and the resulting DNA recut with *Eco*R V to result in the 146 bp palindrome.

2.7.1 Cloning of 1/2ASP for large scale production

The cloning of the 1/2ASP restriction fragment was done using pBluescript® (Stratagene Cloning Systems, La Jolla CA) taking advantage of the multiple cloning site which provided adjacent *Eco*R I and *Eco*R V cutting sites for cloning of a single copy of the 1/2ASP fragment. Cloning was done using standard protocols (Sambrook

et al., 1989) using T4 DNA ligase to ligate the fragment into prepared vector DNA and to transform *E. coli* host cells for maintenance and production of the DNA. As the goal of the cloning strategy was to construct a plasmid containing many copies of the 1/2ASP fragment in multiple direct repeats, the Epicurian Coli Sure® strain (Strategene Cloning Systems) was used due its deficiency in homologous recombination. After cloning of the single copy of 1/2ASP, three other adjacent restriction sites were utilized to increase number of copies of 1/2ASP in the same plasmid. Two restriction enzymes, *Sma* I and *Hin*C II, were chosen which produce blunt ends, one on either side of the 1/2ASP insert. Another enzyme, *Xho* I, located distal of the 1/2ASP insert from the *Hin*C II site was also used. The strategy for cloning the multiple copy 1/2ASP plasmid then involved preparation of the 1/2ASP plasmid and separation of the plasmid material into 2 streams for subsequent manipulations (Palmer *et al.*, 1996). One aliquot of the plasmid was used to excise the 1/2ASP fragment with *Sma* I and *Xho* I. The other aliquot was used to prepare vector for the *Sma* I/*Xho* I fragment, by cutting with the other blunt cutter, *Hin*C II, as well as *Xho* I and removing the resulting 4 bp fragment. The *Sma* I/*Xho*I fragment was then ligated into the *Hin*C II/*Xho* I opened plasmid which already contained the 1/2ASP fragment. This ligation resulted in the destruction of the *Hin*C II site in the original plasmid construction while carrying a fresh copy of it along with the *Sma* I/*Xho* I insert. Because of this, the *Sma* I/*Xho* I restriction fragment represented a cassette which could be doubled by following this cloning procedure and thereby double the

number of copies of the 1/2ASP fragment carried in the plasmid in each iteration of the procedure.

2.7.2 Large scale production of plasmid containing 1/2ASP DNA fragments

Bacterial fermentation was done in 1 L of TB containing 100 μg/mL ampicillin in 2 L volume baffle flasks. Plasmids containing either 16 or 32 copies of the 1/2ASP fragment were used to transform Epicurian Coli Sure competent cells and an inoculum was prepared by incubation of the cells in TB containing 100 μg/mL ampicillin overnight at 37°C overnight. Each 1 L culture was inoculated with 2 mL of overnight culture and allowed to ferment with shaking at 37°C until reaching stationary phase. The bacteria were then harvested by centrifugation at 4°C in a Sorvall GS3 rotor at 5000 rpm for 15 minutes. The bacterial pellet from each 1 L of culture was then subjected to a standard alkaline lysis procedure as follows:

Buffers

Buffer P1:

- 25 mM glucose
- 10 mM EDTA
- 25 mM Tris-Cl, pH 8.0

Buffer P2:

- 1% sodium dodecyl sulfate
- 0.2 N NaOH

Buffer P3:

- 3 M potassium acetate, pH 4.8

Protocol

- The pellet was suspended in 100 mL of buffer P1 in a 400 mL Sorvall centrifuge bottle.
- 100 mL of buffer P2 was added and the bottle gently inverted 5 times to mix the suspension without shearing of chromosomal DNA.
- 100 mL of buffer P3 was added and mixed by gently inverting the bottle several times and then allowed to incubate at 4°C for 30 minutes.
- Cell debris was removed by straining the lysate through several folds of cheese-cloth.
- The lysate was then treated with 0.7 volumes of isopropanol to precipitate plasmid DNA.

- The precipitate was harvested from lysate by centrifugation in a Sorvall GS3 rotor at 7000 rpm for 30 minutes. Centrifugation was done at room temperature to prevent precipitation of salts.

- Cleared lysate was discarded and the pellets dried under vacuum.

- The plasmid pellets were resuspended in in TE and all material from six 1 L cultures was collected into a total volume of 50 mL.

- The plasmid was precipitated by addition of 33 mL of 20% PEG 8000, 1.8 M NaCl with continuous stirring. The precipitation was incubated overnight at 4^oC.

- The plasmid was harvested from the precipitation by centrifugation in a Sorvall GS3 rotor at 7000 rpm for 1 hour at 4^oC and the pellet was resuspended in 20 mL of TE buffer and transferred to a 50 mL teflon Oak Ridge tube.

- The PEG precipitated plasmid was then treated to remove impurities and residual PEG 8000 by extracting twice with 1/2 volume of buffered phenol and then by extracting twice with 1 volume of chloroform.

- The plasmid was precipitated with 1/10 volume of 3 M sodium acetate, pH 5.5, and 2 volumes of ethanol.

- The plasmid was harvested by centrifugation in a Sorvall GS3 rotor at 7000 rpm for 1 hour at 4^oC and dried under vacuum.

- The final pellet was dissolved in TE and the yield of plasmid determined by UV spectrophotometry and the quality of the plasmid checked by electrophoresis in 1% agarose.

2.7.3 Construction of 146 bp ASP palindrome DNA from plasmid containing 1/2ASP DNA fragments

Plasmid from a total of 18 L of bacterial culture was pooled for each preparation of 1/2ASP DNA fragments. The 1/2ASP fragment was prepared from the plasmid by excision from the plasmid with *Eco*R I and *Eco*R V restriction enzyme reactions and the 1/2ASP fragment separated from vector DNA by precipitation with PEG 8000. The 1/2ASP fragment was then separated from linker DNA fragments by FPLC using ion exchange chromatography. Because of the amounts of material involved in each reaction, pilot reactions were performed prior to full-scale enzymatic treatment in order to minimize the cost of production. The completeness of each reaction was verified by electrophoresis before proceeding to the next step in the procedure. In order to release the *Eco*R I/*Eco*R V fragment, the plasmid was first cut with *Eco*R I due to the inefficiency of *Eco*R V enzyme when cutting supercoiled plasmid. Details of the procedure were as follows:

- Purified plasmid from 18 L of culture was pooled.
- Pilot reactions were performed to establish minimum amount of *Eco*R I enzyme

needed to completely cut plasmid.

- Large scale reaction was performed with *Eco*R I and completeness verified by electrophoresis in 1% agarose gel using 1X TBE running buffer.

- Pilot reactions were performed to establish minimum amount of *Eco*R V enzyme needed to completely cut *Eco*R I linearized plasmid.

- Large-scale reaction was performed with *Eco*R I and completeness verified by electrophoresis in 12% polyacylamide slab gels using 1X TBE running buffer.

- Vector DNA was precipitated by drop-wise addition of 20% PEG 8000, 1.8 M NaCl to a final concentration of about 5% PEG 8000, and centrifugation in a Sorvall GS3 rotor at 7000 rpm for one hour at 4°C. The precipitation was checked by electrophoresis of supernatant and precipitated DNA to verify that all 1/2ASP DNA remained in solution.

- The 1/2ASP DNA was precipitated with 3 volumes of ethanol and harvested by centrifugation in a Sorvall GS3 rotor at 7000 rpm for one hour at 4°C. The DNA pellet was dissolved in 10 mM Tris-Cl, pH 8.0, 1 mM EDTA and residual PEG removed by extracting twice with an equal volume of chloroform.

- The DNA was then precipitated with 1/10 volume of 3 M sodium acetate, pH 5.5, and 3 volumes of ethanol. The pellet was collected by centrifugation and dried under vacuum.

- The pellet was then dissolved in 0.4 M NaCl, 20 mM Tris-Cl, pH 8.0 and the 1/2ASP DNA fragment purified by FPLC ion exchange chromatography on a MonoQ HR16/10 column using a linear gradient from 0.4 M to 1.0 M NaCl.

- Peak fractions from the linear gradient were pooled and the 1/2ASP DNA precipitated using 3 volumes of ethanol.

- The quality and yield of the final 1/2ASP DNA product were checked by UV absorbance spectroscopy and by polyacrylamide gel electrophoresis.

- The amount of T4 DNA ligase required to completely ligate the 1/2ASP DNA was established by pilot reactions and the total yield of 1/2ASP fragment was ligated in a single reaction.

- The completeness of the ligation was verified by polyacrylamide gel electrophoresis.

- The 146 bp DNA palindrome was then produced by cutting the ligated DNA chains with *Eco*R V and the palindrome DNA was subjected to a final purification by FPLC on a MonoQ HR16/10 column using a linear gradient from 0.4 M to 1.0 M NaCl.

- Peak fractions from the linear gradient were pooled and precipitated with 3 volumes of ethanol.

- The yield and quality of the final ASP palindrome DNA was checked by by UV

absorbance spectroscopy and by PAGE.

2.8 Reconstitution of nucleosome core particles

Reconstitution of nucleosome core particles was initiated by adding together purified nucleosomal DNA and purified histone octamer in approximately equimolar amounts in a high salt solution. The ionic strength of the solution was then reduced to allow the DNA to bind to the histone octamer core. Input ratios for the two reactants were first approximated by UV absorbance spectroscopy to provide a mid-point for a series of small scale pilot reactions. The reactions were performed in Eppendorf tubes by dilution of a solution of DNA and histone octamer in a 20 μL volume containing 1 M NaCl to a final volume of 200 μL. In each set of reactions, the concentration of either the protein or the DNA was held constant and the concentration of the other reactant varied around the calculated midpoint representing equimolar stoichiometry. A 10 μL aliquot of each reaction in a series was combined with 3 μL of 40% sucrose and loaded onto a 6% polyacrylamide gel made in 0.1X TBE buffer on a Bio-Rad miniProtean electrophoresis apparatus. The gels were electrophoresed at a constant 100V and then stained with ethidium bromide to visualize the DNA by UV fluorescence. The final input ratio was determined using the point in the series of reactions at which the free DNA band on a stained gel became visually undetectable. The ideal input ratio provided a slight excess of DNA over protein. Pilot reactions were done using the

reactants from the exact production batches to be used for large scale reconstitutions in order to eliminate batch to batch variation in the behavior of the reconstitution reactions. This step was necessary to maximize yield of nucleosome core particles from materials which were costly to produce in terms of time and effort.

Large-scale reconstitutions were assembled, based on the results of the pilot reactions, to yield between 10 and 30 mg of nucleosome core particles at a final concentration of 1 mg/mL. The reactants were equilibrated overnight at room temperature by dialysis against a starting buffer of 2 M KCl, 20 mM Tris-HCl, 1 mM EDTA, 1 mM β-mercaptoethanol, 0.04% NaN_3, 0.1 mM PMSF, 0.01% Nonidet P-40, pH 8.0. The reconstitution was then done by a modification of the slow salt gradient dialysis method (Richmond *et al.*, 1988). The slow salt gradient was produced by pumping an ending buffer of 0.3 M KCl, 20 mM Tris-Cl, 1 mM EDTA, 1 mM β-mercaptoethanol, 0.04% NaN_3, 0.1 mM PMSF, and 0.01% Nonidet P-40, pH 8.0, in the starting dialysis buffer with stirring using a Pharmacia LKB peristaltic pump P-1. A constant volume was maintained in the dialysis vessel by pumping the ending buffer into the bottom of a 1 L side-arm Erlenmeyer flask and allowing the excess solution to flow out through the side arm. The rate at which the ending buffer was added to the dialysis chamber was adjusted to lower the salt concentration to 1 M KCl in one day and to complete the dialysis over a additional period of 4 days at 29°C. When using 0.01% Nonidet P-40 in the solution, it was necessary to keep the reaction at room temperature until the concentration of KCl was 1 M or less. At 2 M KCl the cloud point of the detergent is

exceeded at 29°C and the solution undergoes a liquid-liquid binary phase separation. The volume of the starting buffer was 1 L and a total of 3 L of ending buffer were used for the salt gradient. At the end of the salt gradient dialysis, the reconstitution mixture was dialyzed against a final buffer of 50 mM KCl, 20 mM Tris-Cl, 0.04% NaN_3, 1.0 mM EDTA, 0.1 mM PMSF.

The nucleosome core particles were then harvested from the dialysis and analyzed by non-denaturing PAGE on 6% gels with 0.1X TBE buffer. The yield of nucleosome core particles was determined by UV absorbance spectroscopy using an extinction of 10 AU at 260 nm for 1 mg/mL NCPs. This extinction value for nucleosome core particles is the same as for average B-form DNA and assumed that one half of the mass is DNA and that the absorbance of histone was negligible at 260 nm. The core particles were then concentrated using Centriprep 30 and Centricon 30 microconcentrators to 10 mg/mL prior to purification of correctly phased nucleosome core particles by preparative PAGE.

2.9 Purification of correctly phased nucleosome core particles

The steps required for purification of correctly phased NCPs have been published (Harp *et al.*, 1995) but will be repeated here for completeness and to update the

procedures to reflect current practice. The separation was based on differential electrophoretic mobility of NCP with the histone core bound at different positions (phases) on the DNA. The correctly phased NCPs that had no DNA terminus unbound to protein migrated most rapidly.

2.9.1 Micrococcal nuclease analysis of reconstituted nucleosome core particles

Nucleosome core particles for micrococcal (streptococcal) nuclease (Worthington Biochemical Corp., Freehold, NJ) analysis were prepared by dilution of 20 μL DNA and histone octamer in 0.4 M NaCl to 40 mM NaCl in equal steps by addition of 0.1X TBE, and 0.1% Nonidet P-40 at 4°C. The final DNA concentration was 0.1 mg/mL and the concentration of nucleosome core particles was approximately 0.2 mg/mL. The nucleosome core particles were treated with micrococcal nuclease by aliquoting 25 μL of the reconstitution into separate Eppendorf tubes and bringing the solution to 1 mM $CaCl_2$ by addition of 10 mM $CaCl_2$. Reactions used varying amounts of enzyme - 0.1, 0.06, 0.03, 0.01 units/μg of DNA - and were done at 37°C for 30 minutes. Reactions were stopped by addition of EDTA to 2 mM. The reactions were analyzed by non-denaturing polyacrylamide gel electrophoresis in a 6% gel with 0.1X TBE. The gels were stained with ethidium bromide and photographed using UV fluorescence. The photographic negative was scanned on a Molecular Dynamics Model

300A computing densitometer to quantify changes in band densities between nuclease treatments and an untreated control lane.

2.9.2 Preparative polyacrylamide gel electrophoresis

The symmetrically phased nucleosome core particles were separated from degenerately phased particles by preparative non-denaturing polyacrylamide gel electrophoresis using the Bio-Rad Model 491 Prep Cell apparatus. In this apparatus, molecules were separated in a toroidal gel until, upon exiting the gel, they were eluted in a continuous flow radially inward and then upward through a central core (see Figure 2.2). The flow was then passed through a Pharmacia LKB UV-1 monitor connected to a chart recorder and was then to a Pharmacia LKB FRAC-100 fraction collector. The central core also contained a cooling core, surrounding the elution path, in a buffer recirculation circuit which included the lower buffer chamber and a heat exchanging coil in a controlled-temperature water bath. Continuous flow of the elution buffer was maintained using a Pharmacia LKB peristaltic pump P-1 set for a 1 mL/min flow and proceeded from a reservoir built into the top of the apparatus through a flow distribution baffle surrounding the elution chamber and then radially into the center and out through the elution port in the central cooling core. The elution chamber was constructed using a plastic frit in contact with the lower surface of the gel underlain with a membrane to contain the particles as they were eluted. A second plastic frit formed the floor of the elution chamber maintaining continuity with the lower

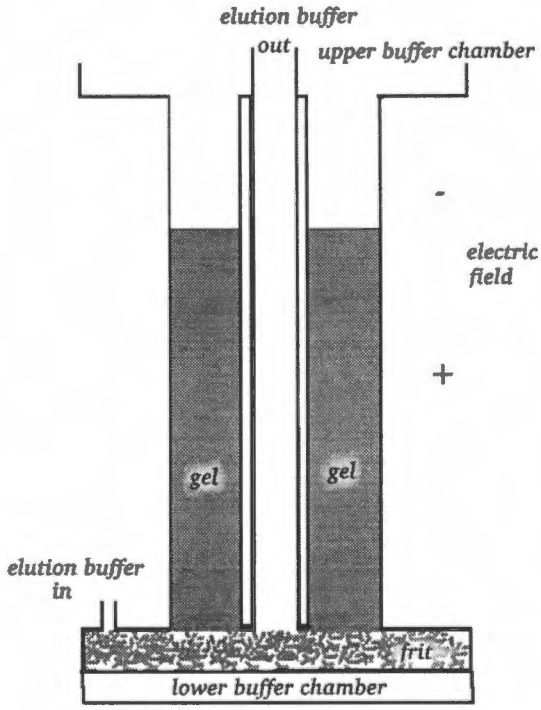

Figure 2.2: Longitudinal cross-section of the apparatus used for preparative electrophoresis shows the input port for elution buffer flow entering a baffled frit to produce a radial flow inward toward a tube extending upward through the center of the gel. Molecules eluting from the bottom of the gel were swept upward through the central tube and carried to a fraction collector. The cooling water circuit included in the central core is not shown.

buffer. The buffers used were fractional dilutions of 1X TBE (89 mM Tris base, 89 mM boric acid, and 10 mM EDTA). The pH of the buffer was adjusted to pH 8.0 after dilution from a 5X TBE stock solution. The upper buffer chamber contained 0.1X TBE, 0.01% Nonidet P-40 while the lower buffer chamber was filled with 0.5X TBE. The elution buffer was 0.5X TBE, 0.01% Nonidet P-40, 1 mM benzamidine, and 0.1 mM PMSF.

The gel was cast in the 37 mm ID tube using the manufacturer's gel casting stand. Approximately 10 mm of water-saturated 2-butanol was placed in the bottom of the tube mounted on the casting stand and a sufficient quantity of degassed 6% acrylamide (29:1 acrylamide: N,N′-methylenebisacrylamide) solution in 0.1X TBE to make approximately 12.5 cm of gel was injected beneath the 2-butanol overlayer. The casting stand was then carefully leveled, and a syringe fitted with a thin plastic tube was used to draw the level of the acrylamide down to 11.5 to 12.0 cm by aspirating material from the acrylamide/2-butanol interface. This treatment resulted in a smooth and level gel surface which was critical to prevent broadening of eluted peaks caused when particle zones do not elute precisely parallel with the elution chamber. The polymerized gel interface formed within 1-2 hours, at which time the 2-butanol was removed and replaced by several washes of degassed 0.1X TBE and left overnight for complete polymerization.

The apparatus was then assembled and placed on a leveling table. Particular care was taken to ensure that no air bubbles were left in the elution chamber after assembly.

This was accomplished by attaching a 10 mL syringe filled with degassed elution buffer to the elution port at the top of the cooling core and gently pumping buffer in and out of the elution chamber while allowing bubbles to escape from the lower elution buffer port of the flow distribution baffle. Flow was begun on both the elution path and the lower buffer cooling circuit and the system allowed to equilibrate for from several hours to overnight. The gel was pre-electrophoresed at either a constant 50 V overnight or 250 V for 30 minutes. The nucleosome core particle solution was made to 9-10% sucrose by addition of 40% sucrose, and loads of 5-7 mg of nucleosome core particles in a total volume of 0.25 to 0.5 mL were used. No tracking dye was used. The apparatus was leveled and the nucleosome core particle solution was carefully layered onto the gel surface. Electrophoresis was done using a constant power of about 3.5 W which produced an initial voltage of about 250 V and a final voltage of about 300 V. Correctly phased nucleosome core particles eluted within 5-6 hours and were collected using 1-3 mL fractions.

The fractions were analyzed using 6% non-denaturing gel electrophoresis with 0.1X TBE running buffer on a Bio-Rad mini-Protean apparatus. The nucleosome core particles were visualized by UV fluorescence of ethidium bromide stain and the peak fractions pooled accordingly. The pooled fractions were concentrated using Centricon 30 microconcentrators and passed through Sephacryl S-400 HR in an HR 10/30 column equilibrated with 50 mM KCl, 10 mM potassium cacodylate, and 0.1 mM EDTA, pH 6.0 for final purification, and then reconcentrated with Centricon 30 mi-

croconcentrators prior to use in crystallization experiments. Final quality checks of the nucleosome core particles were made by non-denaturing polyacylamide gel electrophoresis to verify that only a single band was present and by SDS-PAGE to verify that no proteolysis had occurred during the lengthy procedures.

2.10 Microgravity crystallization of nucleosome core particles

Ground-based crystallization experiments suffered from a tendency of the NCP crystals to grow rapidly in one direction resulting in hollow crystal morphology. In order to ameliorate this deficiency, microgravity crystallization using Diffusion Controlled Apparatus for Microgravity, DCAM (Carter *et al.*, 1999b), was used to produce the best possible quality of nucleosome core particle crystals. The DCAM consists of a body containing two chambers, one large and one small separated by a septum through which passes a "fuse" which contains an agarose plug to prevent mixing of buffer in the two chambers (see Figure 2.3). End plates cap the chambers externally and one of the end plates possesses a button in which the nucleosome solution is placed and covered by a dialysis membrane secured with an O-ring. For the nucleosome core particle experiments, specially prepared buttons were provided on request by NASA. These buttons had been vapor polished by exposure to methylene chloride fumes. The polished buttons were required to prevent nucleation of crystals on a

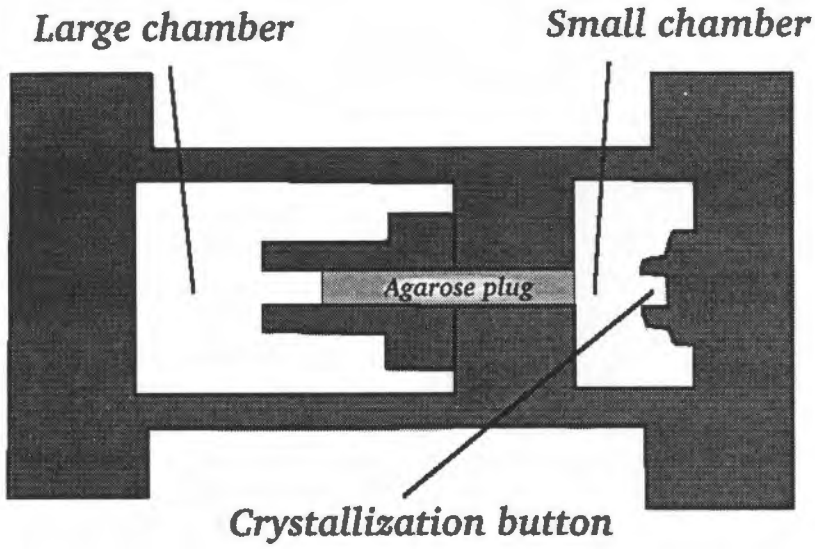

Figure 2.3: Diagram of the DCAM in longitudinal section shows the two buffer chambers separated by an agarose plug. The macromolecule to be crystallized is placed in the crystallization button and covered with a dialysis membrane.

rough surface. For ground-based experiments, buttons were polished using a 0.3 μm grit alumina.

A DCAM was measured using a micrometer to provide the volumes of the two chambers with and without the fuse. It was only possible to install the fuse in the large chamber although the crystallization button could be installed in either chamber to allow for a greater variety of equilibration trajectories. The measurements used for calculation of concentrations necessary to provide a given equilibration point were as follows:

small chamber with button	2.11 mL
small chamber without button	2.36 mL
large chamber with button	7.07 mL
large chamber without button	6.82 mL

Loading of a DCAM experiment was accomplished by first preparing the fuse, which consisted of a hollow cylinder threaded at one end to fit a threaded hole in the septum between the large and small chambers. A collar was provided on the fuse to hold an O-ring which prevented leakage of buffer through the threaded connector. A 0.6% solution of Seakem GTG (FMC Corporation) agarose in buffer was prepared by first adding dry agarose and 10 mL of buffer to a 50 mL Erlenmeyer flask and weighing the flask and contents. The agarose was then melted in a microwave oven and the flask and contents brought back to original weight by addition of glass distilled wa-

ter. The warm agarose solution was then pipetted into clean fuses which had been sealed with label tape on the unthreaded end. A row of fuses could be sealed and held upright on a bench top by pressing the unthreaded end of each onto the sticky side of a strip of tape which had itself been taped, sticky side up, on the bench top. Once the agarose had cooled and gelled, the fuses were installed in the DCAM septum. The nucleosome core particle solution was then pipetted into the crystallization buttons, that held approximately 45 μL of solution and a small piece of Spectropore #1 dialysis membrane was carefully fitted over the button to avoid trapping bubbles. The membrane was then secured with an O-ring and excess membrane trimmed away with a scissors. The chambers were filled with appropriate solutions for the given experiment and the end caps fitted. Bubbles in the DCAM chambers were removed using the debubbling ports provided in the end caps and a small screw was fitted to seal the chamber. Care was exercised to prevent build up of pressure when fitting these screws to prevent dislodging of the agarose plug in the fuse assembly. For microgravity experiments, end caps and debubbling port screws were tightened with a torque wrench to NASA specifications for the flight hardware.

The rate of change in the concentration of precipitant in the chamber containing the crystallization button could be modified by changing the slope of the gradient between the two chambers; and by changing the size and diameter of the agarose plug fuse between the two chambers by placing the crystallization button in either the large or the small chamber and, for ground-based experiments, by varying the

orientation of the DCAM with respect to the Earth's gravitational field. Trajectories of the concentration of the Mn^{++} in the DCAM chambers were determined using a conductivity meter (Orion Instruments). For these measurements, time points were taken by gently drawing 25 μL samples from a chamber through the fill port. The sample was added to exactly 5 mL of glass distilled water in a Falcon 2057 culture tube. The tube was capped and vortexed for 10 seconds and the conductivity probe placed directly into the Falcon tube containing the diluted samples. A series of ground-based trajectories was made to establish appropriate initial conditions for microgravity experiments.

Because of the time delay between loading of a DCAM experiment and arrival of the apparatus in the microgravity environment on orbit, initial conditions were selected to provide a trajectory of Mn^{++} concentration such that nucleation of crystals would commence in 5-6 days. Because the DCAM experiments began at the moment that the unit was assembled, the unit had to be loaded onto the space shuttle and launched into the microgravity environment of space before crystals begin to nucleate. The vagaries of the NASA schedule and delays due to any number of problems that might be encountered prior to launch required that the first experiments be repeatedly reloaded. What was needed was a way to increase the time between loading and the first nucleation of crystals while still providing a satisfactory equilibration trajectory while in orbit. This was accomplished by installation of a "time delay fuse." This technique used a sufficiently high concentration, 200 mM, of $MnCl_2$ in the fuse's

agarose plug to require a number of days before an actual gradient between the two chambers was established.

All materials and solutions for a microgravity experiment were packaged and carefully transported by car to the point of assembly. The DCAM experiments were assembled at Kennedy Space Center, Marshall Space Flight Center, or New Century Pharmaceuticals in Huntsville, Alabama. After assembly, DCAMs were turned over to NASA personnel for transportation to the launch site at Kennedy Space Center and loaded on the shuttle. Upon return to Earth, DCAMs were retrieved at either Marshall Space Flight Center or New Century Pharmaceuticals and examined and photographed while still in the DCAM before returning to the laboratory. Once back at the laboratory, DCAMs were disassembled and the crystals were harvested into artificial mother liquor as described in the next section. The crystals were photographed once more after harvest to obtain images and measurements undistorted by the plastic housing of the DCAM.

2.11 X-ray diffraction data collection

2.11.1 Preparation and mounting of crystals for low-temperature data collection

Crystals were harvested from the crystallization medium using an artificial mother liquor, AML, containing 30 mM $MnCl_2$, 30 mM KCl, 10 mM K cacodylate, pH 6.0, containing 2% MPD. The NCP crystals were stable indefinitely in this solution. The crystals were then dehydrated through a graded series of AML containing increasing concentrations of 2-methyl-2,4 pentane diol, MPD. This was done by dialysis in buttons over a period of days. The crystals in a dialysis button were placed in a scintillation vial containing exactly 5 mL of AML with 2% MPD. The MPD concentration was increased in 2% steps by addition of AML with 50% MPD and allowed to equilibrate for a minimum of 8 hours at each step. This was continued until the concentration of MPD reached 22.5%. The exact level was measured by refractometry and the final solution adjusted to provide a refractive index of 1.3602 which was determined using an empirically derived regression equation, $\nu = 0.0012(\%MPD) + 1.3332$.

Crystals were mounted using rayon loops in Hampton magnetic goniometer caps (Hampton Research, Laguna Hills CA) using 18 mm pins. Loop sizes were chosen to match the crystal size with a loop only slightly smaller than the crystal being favored. Loops larger than the crystal result in a substantial quantity of the cryoprotectant

solution being carried with the crystal. Excess solution could be removed, if necessary, using absorbent paper dental points. Flash-cooling was achieved by either quickly placing the cap onto the magnetic goniometer head thereby placing the crystal on the loop directly into the cold nitrogen gas stream, or by blocking the stream with a plastic paddle while placing the magnetic cap onto the goniometer and then quickly removing the paddle to release the cold gas stream to flow over the crystal. Blocking the cold gas stream provided time for careful placement of the cap and crystal before flash-cooling, and prevented snapping of the cap onto the goniometer which could result in loss of the crystal from the loop.

2.11.2 Macromolecular crystal annealing

The general applications for the protocol for macromolecular crystal annealing have been published (Harp *et al.*, 1998; Harp *et al.*, 1999). The protocol for annealing a crystal assumes that an adequate cryoprotectant is available or that the crystal can be flash-cooled using a hydrocarbon oil. NCP crystals were annealed using either the AML with MPD or Paratone N, which is a highly refined hydrocarbon oil whose primary use is as a viscosity additive for motor oil. Paratone N used for mounting of crystals was kept dry under vacuum until just prior to use. The crystal to be annealed was first flash-cooled. The crystal was then removed from the cold nitrogen gas stream and quickly transferred to a 0.3 mL drop of cryoprotectant or oil at room temperature on a silanized well slide. The crystal was allowed to drift off of the loop

and then left to warm in the drop for at least 3 minutes. The drop was covered to prevent evaporation using silanized cover slips to cover the drop. Silanization is important to prevent wicking of the buffer out of the well. The crystal was then remounted onto a loop and flash-cooled to complete the annealing protocol. If an NCP crystal had received a substantial radiation exposure prior to annealing, the crystal was aligned to expose a fresh area of the crystal to avoid degradation of data quality after annealing.

Summary of annealing protocol

- Mount crystal in cryoprotectant or oil onto a loop on a magnetic goniometer cap.
- Flash-cool.
- Transfer crystal from the loop quickly into a 0.3 mL or larger drop of cryoprotectant or oil.
- Incubate for a minimum of 3 minutes at crystal growth temperature.
- Remount crystal onto loop.
- Flash-cool.

2.11.3 In-house data collection

X-ray diffraction data were collected in-house using an 18 cm Mar image plate area detector (X-ray Research Gmbh) with double-focusing mirrors (Charles Supper Co., Natick MA) mounted on a Rigaku RU200HB rotating anode X-ray generator using a copper anode with a 0.3 mm focal spot. The X-ray generator was operated at 50 kV and 100 mA. The cryostat was a transfer tube (Molecular Structure Corporation, The Woodlands TX) delivering nitrogen gas at between -183°C and -163°C (90 K to 110 K). The in-house data collection facility was used to screen crystals for diffraction quality and for collection of complete data sets. For collection of complete data sets, the program Strategy (Ravelli *et al.*, 1997) was used to optimize completeness and redundancy of the data by choosing the ϕ angle that would sample the largest unique region of reciprocal space in the shortest rotation range. Rotation angles were generally set to 0.25 or 0.5 degrees and exposure times ranged from 10 to 40 minutes.

2.11.4 Synchrotron data collection

Synchrotron data were collected at beamline X12C at the National Synchrotron Light Source at Brookhaven National Laboratory. The beamline provided a facility for tunable wavelength using a highly accurate monochromator as well as a high intensity synchrotron beam for collection of high quality native diffraction data. X-ray detectors used included a 30 cm Mar image plate area detector and a CCD area detector

with either one module, B1, or four modules, the B4. Each detector consisted of a CCD chip bonded to the small end of a 1:4 fibre-optic taper with a phosphor-coated front surface approximately 100 mm square. The detectors were constructed at Brandeis University. The cryostat was an Oxford Cryosystems (Oxford, UK) Cryostream operated at -163°C (100 K). The beamline used a Nonius goniometer with a kappa cradle providing both ϕ and ω scans.

2.11.5 Data reduction and quality analysis

All area detector image data were integrated and scaled using the HKLSuite programs, denzo and scalepack (Otwinowski & Minor, 1997).

2.12 Phasing of the data

The solution of a crystal structure has as its first goal the production of a three dimensional electron density map which is an accurate representation of the electron distribution in the unit cell. In general, the electron density of a point in the unit cell with coordinates, x, y, and z is calculated by the Fourier summation

$$\rho(x\ y\ z) = \frac{1}{V}\Sigma \mid F(h\ k\ l) \mid e^{-2\pi(hx+ky+lz)+i\alpha(h\ k\ l)} \tag{2.1}$$

where h, k, and l are the Miller indices of the diffracting plane, $\mid F(h\ k\ l) \mid$, is the amplitude of the diffracted wave and α is the phase angle of the diffracted wave.

The diffraction data collected from a crystal represents the intensity of the diffracted wave, I(h k l) from which the amplitude is easily obtained by

$$I(h\ k\ l) = \mid F(h\ k\ l) \mid^2 \tag{2.2}$$

However, all phase information is unavailable. Thus, the first major task is the estimation of the correct phase angle for each reflection.

Two primary methods were available for obtaining phase information, isomorphous replacement and molecular replacement. Molecular replacement is the most rapid method available but requires the availability of accurate molecular models which are sufficiently homologous to the unknown structure. As noted in the Introduction section, the structure of the histone octamer had been published but the coordinates became only recently available. This would certainly have allowed solution of the structure in a timely manner as the histone octamer core of the nucleosome core particle represents nearly one half of the mass of the NCP and is largely unchanged by the binding of DNA. The unavailability of this vital information necessitated a search for heavy atom compounds which would bind to the NCP without significantly affecting the structure to result in an isomorphous heavy atom derivative.

2.12.1 Isomorphous replacement

Isomorphous replacement phasing is accomplished by adding a heavy atom compound to the molecules in the crystal which adds scattering from a small number

of electron-rich atoms to the diffraction due to the macromolecular structure itself. Determination of structure factors of the heavy atoms alone in the unit cell, $\mathbf{F}_H$, allows for solution of structure factors of the macromolecular atoms alone in the unit cell $\mathbf{F}_P$. This is possible if the heavy atom derivative of the macromolecular crystal is sufficiently isomorphous to the native crystal using the relationship $\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H$. In practice, isomorphous replacement involves soaking macromolecular crystals in solutions containing the heavy atom compound in an artificial mother liquor in which the crystal is stable. The usual result is cracking of the crystal or degradation of diffraction from the crystal. However, in this work, if the heavy atom compound did not destroy the crystal and the heavy atom bound with high occupancy at defined positions without altering the structure or positions of the macromolecule in the unit cell, then a derivative diffraction data set was collected which would scale to the native diffraction data with an r_{merge} between about 5% and 15%.

Once a potential isomorphous derivative was available, it was necessary to locate the position of the heavy atoms in the unit cell. The maps used for locating the heavy atoms were produced by subtracting the native data, $|\mathbf{F}|_P$, from the derivative data, $|\mathbf{F}|_{PH}$, to produce a set of isomophous differences, $\Delta|\mathbf{F}|_{iso}$. To improve the quality of the isomorphous differences, the native and derivative data were scaled together using local scaling (Matthews & Czerwinski, 1975) which partitions reciprocal space into blocks and applies a separate scaling factor for reflections within each block. Local scaling was applied to the data using a Fortran program supplied by Robert

Blessing (personal communication) at the Hauptmann-Woodward Institute, Buffalo, New York. Location of the heavy atoms was determined from the isomorphous difference data by application of the Patterson methods which required Fourier transformation of the data using the Patterson synthesis, P(u v w), which is essentially the modulus of the usual Fourier summation, Eq 2.3.

$$P(u\ v\ w) = \frac{1}{V}\Sigma\Delta \mid F(h\ k\ l) \mid^2 cos[2\pi(hu + kv + lw)] \tag{2.3}$$

where u, v, and w are fractional coordinates within the unit cell. This drops out the imaginary part of the structure factor and thereby removes the phase angle. This has the very important advantage of allowing visualization of atomic positions in the absence of phase information. The Patterson synthesis also transforms the data from real space to vector space where peaks represent interatomic distances. This means that a peak in the Patterson map at position (u v w) occurs because real cell atoms occur in the unit cell at positions (x y z) and (x+u y+v z+w) or (x-u y-v z-w). If N atoms exist in the unit cell, then N^2 peaks exist in the Patterson map. But, since N vectors will represent the vector distance between an atom and itself, then N vectors will have length 0. Therefore, the highest peak in the Patterson map occurs at the origin. The origin peak is routinely removed as it may obscure heavy atom peaks. The vector distance between atoms is known but not yet the position.

Real space positions of heavy atoms whose positions are related by space group symmetry produce vectors concentrated in particular locations in the Patterson map known as Harker lines and planes. For instance, for space group $P2_1$, vectors between

the symmetry related positions (x y z) and ($\overline{x}$ y+1/2 $\overline{z}$) will produce the Patterson vectors of the form 2x, 1/2, 2z. Thus, peaks on the plane v=1/2 can be easily related to real space heavy atom positions. To improve the heavy atom parameters and to establish quality and phasing power of the derivative, the heavy atom positions were refined by least-squares using the program MLPHARE in the CCP4 suite (Otwinowski & Minor, 1997).

Phasing by isomorphous replacement is often accomplished using multiple heavy atom derivatives since even when the correct real space location of a heavy atom is known there is still a fundamental ambiguity in the assignment of a phase angle. The use of a single heavy atom derivative results in two equiprobable phase angles, only one of which is correct while the other represents noise. A map calculated using these phases will not be interpretable. A second derivative would remove noise by elimination of much of the ambiguity but it is possible to refine the single isomorphous derivative data by filtering out the noise and improving phases using a method called solvent flattening (Wang, 1985). Solvent flattening was applied by knowledge of the mass of the NCP in the unit cell which provided an estimate of the solvent content of the unit cell. From this information, a threshold value for the electron density was determined which was then used to set all density at or below the cutoff level to zero for the solvent. The method was applied to the isomorphous difference data using the PHASES program package (Furey, 1997).

2.12.2 Molecular replacement

Phasing of experimental diffraction data was greatly simplified through the availability of refined structures homologous to the NCP. Models for molecular replacement were the histone octamer, PDB ID code 2HIO (Arents *et al.*, 1991), and the nucleosome core particle, PDB ID code 1AOI (Luger *et al.*, 1997). Calculation of initial phases from the phasing models required that the model be positioned in the unit cell precisely the same as the new structure. This positioning was accomplished in 2 stages using the CNS package (Brunger *et al.*, 1998). The first stage involved a search of all possible orientations of the phasing model by first translating the center of mass of the model to the unit cell origin. The rotation search was simplified through the use of Patterson maps of the native diffraction data and a Patterson synthesis based on the phasing model. Maximizing the similarity of the two Patterson maps produced a list of possible rotations which were then input to a translation search in the second stage of the process. The procedure was then monitored using the standard R factor relating the calculated structure factors and the measured intensities of the diffraction data. The very high homology of the phasing models made the molecular replacement method very attractive but not to the exclusion of the single isomorphous method since the phases obtained by molecular replacement may well suffer from phase bias due to the use of a phasing model. The single isomorphous replacement phases were used to check the results of molecular replacement phasing.

2.13 Fitting and refinement of model coordinates

Model fitting was performed using Turbo-FRODO and XtalView (McRee, 1999). Either the XtalView or CNS software packages were used to calculate sigmaa weighted electron density maps with Fourier coefficients $2mF_o$-DF_c or mF_o-F_c. SIR and MR phases were combined using the sigmaa program (Read, 1986) in the CCP4 suite (4, 1994). Refinements were performed using CNS with cross-validation throughout (Brunger, 1992). The simulated annealing protocol used the maximum likelihood target (Adams *et al.*, 1997) and bulk solvent correction along with individual temperature-factor refinement between each cycle of fitting. Water peaks were identified and evaluated using the DDQ program (van den Akker & Hol, 1999). Water molecules which refined to an individual temperature-factor significantly lower than the surrounding protein or DNA atoms were replaced as ions by consideration of chemical environment and evaluated by refinement to ascertain changes in temperature-factor and real space R factor. Secondary structure was evaluated using the program DSSP (Kabsch & Sander, 1983) and by manual inspection of hydrogen bonding patterns and interatomic distances.

2.14 Model analysis

2.14.1 Analysis of DNA helical parameters

Analysis of DNA helical parameters in the final molecular model of the nucleosome core particle was accomplished using a program devised and written at Rutgers University (Babcock *et al.*, 1993; Babcock & Olson, 1993; Babcock *et al.*, 1994) in response to the guidelines established by an European Molecular Biology Organization workshop on DNA curvature and bending held in 1988 (Dickerson *et al.*, 1989). The program, called rna for "Run Nucleic Acid," analyzes nucleic acid structure using a database of canonical base structures consisting of all ring atoms and the C1' atom of the ribose moiety. An additional coordinate, called the pivot point, is attached to each of the canonical reference structures and is used for application of rotation operations to the base atoms. The pivot point serves as the origin of a coordinate frame constructed on each base position and allows for direct comparisons between a variety of nucleic acid structures. The first step in the analysis requires least squares fitting of the reference structures to each of the experimental bases along the primary strand. To obtain rotational and translation parameter between adjacent bases or between complementary bases in a base paired structure, the program first applies a rotation to the coordinate frame of the secondary base, the adjacent or complementary base, until the axes coincide with the directions of the axes of the primary base. The coordinate frame of the secondary base is then rotated to an orientation midway

back to the original orientation. The primary base is then rotated to the same orientation and both coordinate frames are then translated to coincide at a point midway between the two original positions. This results in a derivative frame representing the pair of bases. The rna program then calculates and tabulates all relevant parameters describing the rotation and translation of bases as stacked along a strand and for base pairs stacking into a helix.

For comparison of DNA structure as found in the nucleosome to an energy-minimized structure, a 22 bp segment of the α-satellite DNA palindrome surrounding the dyad center at base 73 was chosen. The program was tested using a B-DNA crystal structure (Drew *et al.*, 1981). The results of the test runs were used to modify electrostatic force field simulation and to provide an accurate simulation of the structure of B-DNA in the crystal. The 22 base pair sequence was then subjected to energy minimization calculations using a program developed by Dr. Brian Hingerty and co-workers (Hingerty *et al.*, 1989).

2.14.2 Analysis of the molecular models

The quality of the molecular model was checked using a variety of programs and approaches. During refinement, the validity of calculated structure factors was monitored using cross-validation with the free R statistic (Brunger, 1992). For this purpose, a percentage of reflections, the test set, were excluded from refinement. The

closeness of agreement between the refined model and the diffraction data in the test set was assessed using the R factor, $R = \frac{\Sigma||F_o|-|F_c||}{|F_o|}$. The R factor calculated using the test set was then the free R and was compared to the same calculation made using the remaining data, the working set, which were used in the refinement. The R value can be made arbitrarily low by adding adjustable parameters and errors can be introduce by over-fitting of the model to the data. This is prevented by use of the free R statistic.

Stereochemistry, hydrogen bonding, nomenclature, and other important aspects of model correctness were examined using the WHAT_CHECK program version 4.99 (Hooft *et al.*, 1996) which contains the structure validation routines from the drug design and modeling package WHAT IF (Vriend, 1990). This analysis was used to help guide the final stages of model refinement and to prepare a final version of the model for deposition in the Protein Data Bank, PDB (Berman *et al.*, 2000).

While the WHAT_CHECK program does a thorough analysis of the model coordinates, it does not address the relative agreement of the structure with the experimental data. A difference density map was used to assess the local and global quality of the model utilizing an automatic algorithm implemented in the DDQ program (van den Akker & Hol, 1999). The difference density map was made using Fourier coefficients mF_o-DF_c calculated from the model with water purposely omitted and including all of the diffraction data. The difference density peaks were found and tabulated using the program peakmax from the CCP4 suite. The DDQ program

used the model coordinates in PDB format and the file containing the negative and positive difference density peaks as input. It assessed the peaks based on location relative to existing atoms in the model. A distinction was then made between peaks positioned within reasonable hydrogen bonding distance to existing atoms and those which were more reasonably interpreted as shift peaks (Stout & Jensen, 1989) due to either incorrect atomic position or incorrect assignment of B factors. A quality structure is one in which the difference density peaks due to ordered solvent are high and the shift peaks are small. The output of DDQ was then used to assess both the local and global quality of the structure. Beyond its design role for assessment of structure quality, the DDQ program was useful for finding waters in the map.

2.14.3 Preparation of figures

Figures were prepared using GRASP (Nicholls *et al.*, 1991), MOLSCRIPT (Kraulis, 1991), Raster3D (Merritt & Murphy, 1994), and the Raster3D interface from the xfit program in the XtalView package (McRee, 1999). A color code was prepared to aid recognition of specific histone subunits (Figure 2.4).

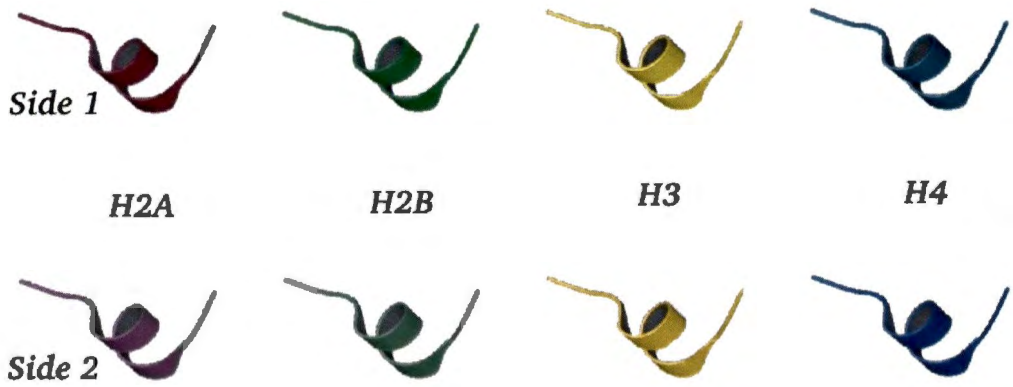

Figure 2.4: Color code for identification of histone subunits in figures. Unless otherwise indicated, all figures conform to this code. Note change in hue denoting the 2 sides of the histone octamer.

Chapter 3

Results

3.1 Sequencing of histone clones

All four of the histone clones were sequenced in single reactions and the published sequences verified. The nucleotide sequences of all the histone clones were translated to amino acid sequences using the GCG package and the results are presented in Figure 3.1 for H2A, Figure 3.2 for H2B, Figure 3.3 fot H3, and 3.4 for H4.

3.2 Large-scale preparation of ASP DNA

The ASP DNA palindrome was successfully cloned into the *Eco*R I site of the plasmid pBLUESCRIPTII, and maintained and produced in Epicurian Coli Sure cells, that

```
TTTTGTTTAACTTTAAGAAGGAGATATACATTGTCTGGTCGTGGTAAACAGGGCGGGAAG
---------+---------+---------+---------+---------+---------+
AAAACAAATTGAAATTCTTCCTCTATATGTAACAGACCAGCACCATTTGTCCCGCCCTTC
                                S  G  R  G  K  Q  G  G  K

GCGCGCGCCAAGGCCAAGTCGCGCTCGTCGCGGGCCGGGCTGCAGTTCCCCGTGGGCCGC
---------+---------+---------+---------+---------+---------+
CGCGCGCGGTTCCGGTTCAGCGCGAGCAGCGCCCGGCCCGACGTCAAGGGGCACCCGGCG
A  R  A  K  A  K  S  R  S  S  R  A  G  L  Q  F  P  V  G  R

GTGCACCGGCTGCTGCGCAAGGGCAACTACGCGGAGCGGGTGGGCGCCGGCGCCCCGGTG
---------+---------+---------+---------+---------+---------+
CACGTGGCCGACGACGCGTTCCCGTTGATGCGCCTCGCCCACCCGCGGCCGCGGGGCCAC
V  H  R  L  L  R  K  G  N  Y  A  E  R  V  G  A  G  A  P  V

TACCTGGCGGCCGTGCTGGAGTACCTGACGGCCGAGATCCTGGAGCTGGCGGGCAACGCG
---------+---------+---------+---------+---------+---------+
ATGGACCGCCGGCACGACCTCATGGACTGCCGGCTCTAGGACCTCGACCGCCCGTTGCGC
Y  L  A  A  V  L  E  Y  L  T  A  E  I  L  E  L  A  G  N  A

GCCCGCGACAACAAGAAGACGCGCATCATCCCCCGCCACCTGCAGCTGGCCATCCGCAAC
---------+---------+---------+---------+---------+---------+
CGGGCGCTGTTGTTCTTCTGCGCGTAGTAGGGGGCGGTGGACGTCGACCGGTAGGCGTTG
A  R  D  N  K  K  T  R  I  I  P  R  H  L  Q  L  A  I  R  N

GACGAGGAGCTCAACAAGCTGCTGGGCAAGGTGACCATCGCGCAGGGTGGGGTGCTGCCC
---------+---------+---------+---------+---------+---------+
CTGCTCCTCGAGTTGTTCGACGACCCGTTCCACTGGTAGCGCGTCCCACCCCACGACGGG
D  E  E  L  N  K  L  L  G  K  V  T  I  A  Q  G  G  V  L  P

AACATCCAGGCCGTGCTGCTGCCCAAGAAGACCGACAGCCACAAGGCCAAGGCTAAGTAA
---------+---------+---------+---------+---------+---------+
TTGTAGGTCCGGCACGACGACGGGTTCTTCTGGCTGTCGGTGTTCCGGTTCCGATTCATT
N  I  Q  A  V  L  L  P  K  K  T  D  S  H  K  A  K  A  K
```

Figure 3.1: Sequence of chicken histone H2A gene with the amino acid translation of the DNA sequence.

```
TTTTGTTTAACTTTAAGAAGGAGATATACATATGCCTGAACCAGCTAAATCCGCGCCCGC
---------+---------+---------+---------+---------+---------+
AAAACAAATTGAAATTCTTCCTCTATATGTATACGGACTTGGTCGATTTAGGCGCGGGCG
                                P  E  P  A  K  S  A  P  A

CCCCAAGAAGGGCTCTAAGAAGGCGGTCACCAAGACCCAGAAGAAGGGCGACAAGAAGCG
---------+---------+---------+---------+---------+---------+
GGGGTTCTTCCCGAGATTCTTCCGCCAGTGGTTCTGGGTCTTCTTCCCGCTGTTCTTCGC
 P  K  K  G  S  K  K  A  V  T  K  T  Q  K  K  G  D  K  K  R

CAAGAAGAGCCGCAAGGAGAGCTACTCGATCTACGTGTACAAGGTGCTGAAGCAGGTGCA
---------+---------+---------+---------+---------+---------+
GTTCTTCTCGGCGTTCCTCTCGATGAGCTAGATGCACATGTTCCACGACTTCGTCCACGT
 K  K  S  R  K  E  S  Y  S  I  Y  V  Y  K  V  L  K  Q  V  H

CCCCGACACGGGCATCTCGTCCAAGGCCATGGGCATCATGAACTCGTTCGTCAACGACAT
---------+---------+---------+---------+---------+---------+
GGGGCTGTGCCCGTAGAGCAGGTTCCGGTACCCGTAGTACTTGAGCAAGCAGTTGCTGTA
 P  D  T  G  I  S  S  K  A  M  G  I  M  N  S  F  V  N  D  I

CTTCGAGCGCATCGCCGGCGAGGCGTCGCGCCTGGCGCACTACAACAAGCGCTCGACCAT
---------+---------+---------+---------+---------+---------+
GAAGCTCGCGTAGCGGCCGCTCCGCAGCGCGGACCGCGTGATGTTGTTCGCGAGCTGGTA
 F  E  R  I  A  G  E  A  S  R  L  A  H  Y  N  K  R  S  T  I

CACGTCGCGGGAGATCCAGACGGCCGTGCGGCTGCTGCTGCCCGGCGAGCTGGCCAAGCA
---------+---------+---------+---------+---------+---------+
GTGCAGCGCCCTCTAGGTCTGCCGGCACGCCGACGACGACGGGCCGCTCGACCGGTTCGT
 T  S  R  E  I  Q  T  A  V  R  L  L  L  P  G  E  L  A  K  H

CGCGGTCTCGGAGGGCACCAAGGCGGTCACCAAGTACACTAGTTCCAAGTAAGGATCCGG
---------+---------+---------+---------+---------+---------+
GCGCCAGAGCCTCCCGTGGTTCCGCCAGTGGTTCATGTGATCAAGGTTCATTCCTAGGCC
 A  V  S  E  G  T  K  A  V  T  K  Y  T  S  S  K
```

Figure 3.2: Sequence of chicken histone H2B gene with the amino acid translation of the DNA sequence.

```
TTTGTTTAACTTTAAGAAGGAGATATACTATGGCTCGTACTAAACAAACCGCACGGAAGT
---------+---------+---------+---------+---------+---------+
AAACAAATTGAAATTCTTCCTCTATATGATACCGAGCATGATTTGTTTGGCGTGCCTTCA
                                A  R  T  K  Q  T  A  R  K  S

CCACCGGCGGGAAGGCGCCCCGCAAGCAGCTGGCCACCAAGGCGGCCCGCAAGAGCGCGC
---------+---------+---------+---------+---------+---------+
GGTGGCCGCCCTTCCGCGGGGCGTTCGTCGACCGGTGGTTCCGCCGGGCGTTCTCGCGCG
 T  G  G  K  A  P  R  K  Q  L  A  T  K  A  A  R  K  S  A  P

CGGCCACGGGCGGCGTGAAGAAGCCGCACCGCTACCGGCCCGGCACGGTGGCGCTGCGCG
---------+---------+---------+---------+---------+---------+
GCCGGTGCCCGCCGCACTTCTTCGGCGTGGCGATGGCCGGGCCGTGCCACCGCGACGCGC
 A  T  G  G  V  K  K  P  H  R  Y  R  P  G  T  V  A  L  R  E

AGATCCGGCGCTACCAGAAGTCCACGGAGCTGCTGATCCGCAAGCTGCCCTTCCAGCGCC
---------+---------+---------+---------+---------+---------+
TCTAGGCCGCGATGGTCTTCAGGTGCCTCGACGACTAGGCGTTCGACGGGAAGGTCGCGG
 I  R  R  Y  Q  K  S  T  E  L  L  I  R  K  L  P  F  Q  R  L

TGGTGCGCGAGATCGCGCAGGACTTCAAGACCGACCTGCGCTTCCAGAGCTCGGCCGTCA
---------+---------+---------+---------+---------+---------+
ACCACGCGCTCTAGCGCGTCCTGAAGTTCTGGCTGGACGCGAAGGTCTCGAGCCGGCAGT
 V  R  E  I  A  Q  D  F  K  T  D  L  R  F  Q  S  S  A  V  M

TGGCGCTGCAGGAGGCGAGCGAGGCCTACCTGGTGGGGCTCTTCGAGGACACCAACCTGT
---------+---------+---------+---------+---------+---------+
ACCGCGACGTCCTCCGCTCGCTCCGGATGGACCACCCCGAGAAGCTCCTGTGGTTGGACA
 A  L  Q  E  A  S  E  A  Y  L  V  G  L  F  E  D  T  N  L  C

GCGCCATCCACGCCAAGCGCGTCACCATCATGCCCAAGGACATCCAGCTCGCCCGCCGAA
---------+---------+---------+---------+---------+---------+
CGCGGTAGGTGCGGTTCGCGCAGTGGTAGTACGGGTTCCTGTAGGTCGAGCGGGCGGCTT
 A  I  H  A  K  R  V  T  I  M  P  K  D  I  Q  L  A  R  R  I

TTCGCGGTGAACGTGCATAAGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAATTG
---------+---------+---------+---------+---------+---------+
AAGCGCCACTTGCACGTATTCCTAGGCCGACGATTGTTTCGGGCTTTCCTTCGACTTAAC
 R  G  E  R  A
```

Figure 3.3: Sequence of chicken histone H3 gene with the amino acid translation of the DNA sequence.

```
AAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGTCTGGTCGTGGTAAAGGCG
---------+---------+---------+---------+---------+---------+
TTTTATTAAAACAAATTGAAATTCTTCCTCTATATGTATACAGACCAGCACCATTTCCGC
                                          S  G  R  G  K  G  G

GGAAGGGGCTCGGCAAGGGCGGCGCCAAGCGCCACCGCAAGGTGCTGCGCGACAACATCC
---------+---------+---------+---------+---------+---------+
CCTTCCCCGAGCCGTTCCCGCCGCGGTTCGCGGTGGCGTTCCACGACGCGCTGTTGTAGG
 K  G  L  G  K  G  G  A  K  R  H  R  K  V  L  R  D  N  I  Q

AGGGCATCACCAAGCCGGCCATCCGCCGCCTGGCGCGGCGCGGCGGCGTCAAGCGCATCT
---------+---------+---------+---------+---------+---------+
TCCCGTAGTGGTTCGGCCGGTAGGCGGCGGACCGCGCCGCGCCGCCGCAGTTCGCGTAGA
 G  I  T  K  P  A  I  R  R  L  A  R  R  G  G  V  K  R  I  S

CGGGGCTCATCTACGAGGAGACGCGCGGCGTGCTCAAGGTCTTCCTGGAGAACGTCATCC
---------+---------+---------+---------+---------+---------+
GCCCCGAGTAGATGCTCCTCTGCGCGCCGCACGAGTTCCAGAAGGACCTCTTGCAGTAGG
 G  L  I  Y  E  E  T  R  G  V  L  K  V  F  L  E  N  V  I  R

GCGACGCCGTCACCTACACCGAGCACGCCAAGAGGAAGACGGTCACGGCCATGGACGTGG
---------+---------+---------+---------+---------+---------+
CGCTGCGGCAGTGGATGTGGCTCGTGCGGTTCTCCTTCTGCCAGTGCCGGTACCTGCACC
 D  A  V  T  Y  T  E  H  A  K  R  K  T  V  T  A  M  D  V  V

TCTACGCGCTCAAGCGCCAGGGACGCACCCTCTACGGCTTCGGCGGTTAAGGATCCGGCT
---------+---------+---------+---------+---------+---------+
AGATGCGCGAGTTCGCGGTCCCTGCGTGGGAGATGCCGAAGCCGCCAATTCCTAGGCCGA
 Y  A  L  K  R  Q  G  R  T  L  Y  G  F  G  G
```

Figure 3.4: Sequence of chicken histone H4 gene with the amino acid translation of the DNA sequence.

are deficient in homologous recombination. Repeated attempts to produce a plasmid with multiple copies of the 146 bp ASP palindrome were unsuccessful, and the approach was abandoned as fruitless. Large-scale production of the ASP fragment was done using the single copy plasmid, pASP4. Unfortunately, DNA palindromes introduce significant problems into large-scale production schemes. The first problem encountered was the fact that duplex DNA palindromes are metastable, if sufficiently long, allowing folding and intra-strand base pairing into a stable free hairpin structure. This will reduce the yield of free duplex palindrome from bacterial fermentation if consideration is not given to protecting the DNA from destabilizing conditions such as high temperature and extraction with organic solvents during purification.

A second problem presented itself which was less obvious but ultimately more significant. The yield of ASP DNA from the plasmid preparations was found to be less than expected from large-scale fermentation. Yields were consistently 0.9-1.0 mg from a 10 L batch. Even a 400 L fermentation run using the 500 L New Brunswick fermentor produced very little ASP fragment. Strenuous attempts to clone multiple copies of the palindrome into a plasmid to increase yields were entirely unsuccessful. A clue to the reason for the low yields came from a restriction analysis of plasmid produced in large-scale fermentation. Restriction of the plasmid with *Eco*R V produced only the vector DNA and the expected 146 bp palindrome band. On the other hand, restriction of the plasmid with *Pvu* II, which cuts on either side of the multiple cloning site in pBLUESCRIPT, resulted in both the expected 594 bp band with the palin-

drome insert and another band at about 448 bp which indicated that the insert had been lost in some portion of the plasmid population. Further restriction of the 448 bp band with *Eco*R V was unsuccessful indicating that the *Eco*R V sites had been destroyed by the deletion of the palindrome insert. As the Epicurian Coli Sure host cell is deficient in homologous recombination, the most likely mechanism appeared to be "read through" by the bacterial DNA polymerase during plasmid replication.

It appears likely that the yield of DNA palindrome was affected by the formation of a transient cruciform in the center of the 146 bp palindrome during plasmid replication. The transient cruciform structure may affect production of the ASP fragment as the bacterial DNA polymerase encounters the Holiday junction by either allowing deletion of the palindrome by read through or by producing a pause site during plasmid replication. Plasmids in which the palindrome has been deleted are then favored during fermentation due to more efficient replication in the absence of the pause site. This would explain lowered yields in larger cultures requiring a greater number of cell generations. The cruciform structure must have been transient. An examination of the plasmids produced by large-scale fermentation performed by Mr. David Allison (personal communication) of Oak Ridge National Laboratory using a Nanoscope II AFM to perform atomic force microscopy, failed to detect the unusual structures.

Other strategies for large-scale production of the ASP DNA were examined including PCR and solid phase synthesis. They were found to be prohibitively expensive. The final strategy tried was to produce large quantities of one half of the palindrome

(1/2ASP), the *Eco*R I/*Eco*R V fragment, and assemble the palindrome by ligation of the *Eco*R I ends. This strategy worked remarkably well and is the standard protocol for production of palindrome DNA in the laboratory at this time. The details of the protocol have been published (Palmer *et al.*, 1996). The protocol is capable of routinely producing about 40 mg of ASP DNA from 18 L of culture in as short a time as two weeks.

3.3 Microgravity crystallization using the DCAM

The diffusion-controlled crystallization apparatus for microgravity, DCAM, utilizes microdialysis counterdiffusion and allows for relatively precise control of the rate of equilibration (Carter *et al.*, 1999b). Before microgravity experiments could be reasonably attempted, it was necessary to establish parameters which would produce ground-based crystals within the time frame needed to allow for nucleation of NCP crystals after the apparatus had arrived in the microgravity environment of either the NASA Space Shuttle in orbit or the Russian Mir Space Station. To establish these parameters, a set of DCAMs were obtained and the rate of equilibration of the Mn^{++} concentration within the 2 chambers of the device was examined using conductivity measurements. Results of a typical experiment are shown in Figure 3.5.

Microgravity experiments were flown on six separate missions as summarized in Table 3.1. Shuttle missions are listed only for those launches carrying DCAM into

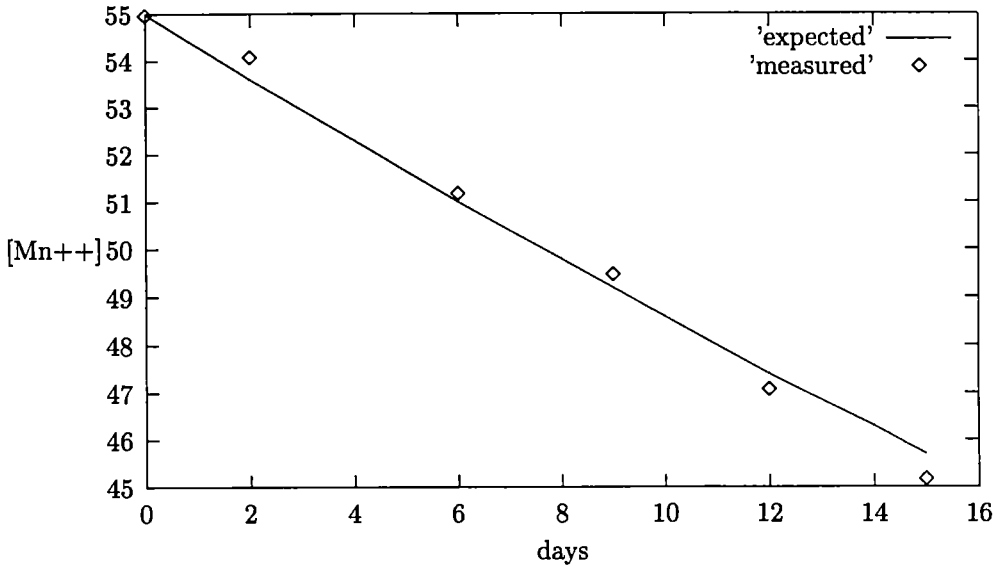

Figure 3.5: Equilibration of the $MnCl_2$ concentration in the small chamber of a DCAM. The DCAM was set to equilibrate at 40 mM $MnCl_2$. The solid line represents expected values assuming an equilibration rate constant of 0.0123.

Table 3.1: List of microgravity experiments flown.

Shuttle Mission	Launch Date	Mission	Number of DCAMs flown with NCP experiments
STS-73	10-20-95	USML-2	36
STS-76	3-22-96	Mir-3	36
STS-79	9-16-96	Mir-4	9
STS-81	1-12-97	Mir-5	36
STS-89	1-22-98	Mir-8	18

microgravity. Return missions such as STS-91 that only returned DCAMs are not shown. The microgravity experiments returned variable results primarily due to technical problems. The results from the first set of experiments, USML-2, provided only a few crystals that displayed a morphology indicating the crystals grew in microgravity (see Figure 3.6). In these instances, the cross-section of the rod-shaped crystals was nearly hexagonal. This morphology is not seen in ground-based experiments since the growing crystals either sedimented to the bottom of the crystallization chamber or nucleated on the surface of the chamber wall. Contact with a solid surface then prevented addition to the occluded surface of the crystal resulting in a flattened cross-section. In general, the microgravity environment did, in fact, solve the most significant problem presented by ground-based crystallization: hollow crystals. Figure 3.7 shows a particularly large microgravity crystal which is 0.6 mm across and shows no evidence of hollowness. Microgravity crystals of the nucleosome core particles tended to be solid and form as more nearly perfect hexagonal rods in contrast to the flattened cross section of ground-based crystals. The value of microgravity, given the cost of these experiments, is still questionable as crystals of equivalent size can now be grown in ground-based DCAM experiments. The true value of microgravity can still be found in the diffraction quality of the resulting crystals.

A crystal grown on the USML-2 mission was used to collect data at beamline X12C at the National Synchrotron Light Source. The resolution of the data extended to 2.5 Å and is the data used for the current high resolution structure. No ground-

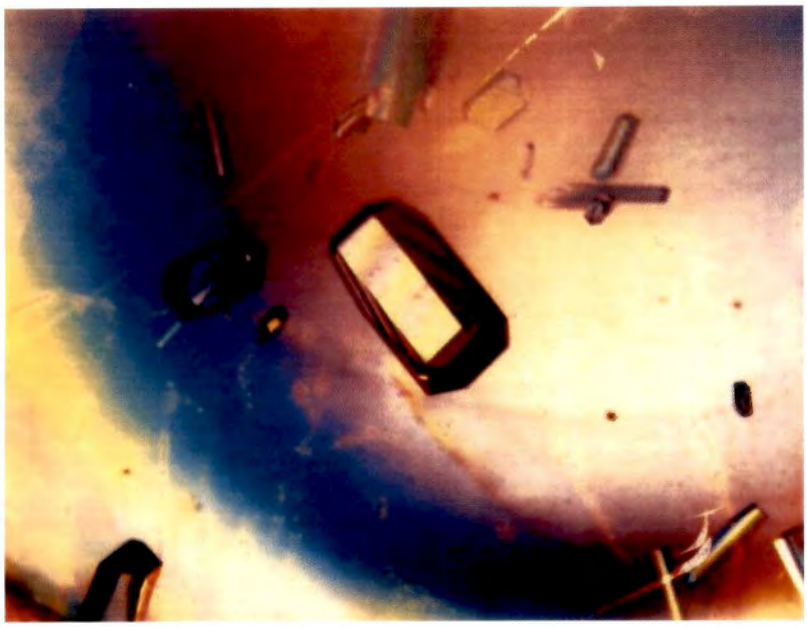

Figure 3.6: A crystal grown on microgravity mission, USML-2, which shows an hexagonal cross-section suggesting that the crystal was grown under microgravity conditions.

Figure 3.7: A very large nucleosome core particle crystal grown in a DCAM aboard the Mir space station.

based crystals have provided data of the same quality. This is seen primarily in the anisotropy of diffraction from ground-based crystals but is also evident in overall diffraction quality along any reciprocal space axis. Figure 3.8 provides a graphical illustration of the difference in microgravity versus ground-based crystals of the nucleosome core particle. The reason for increased diffraction quality for crystals grown in a microgravity environment is still somewhat controversial but may be due to the lack of convection in the microgravity environment as well as to the lack of sedimentation of crystals as they grow. A crystal nucleated in microgravity remains in place in the crystallization medium and does not sink to the bottom of the crystallization buttonm where addition of NCPs to the surface of the crystal in contact with the plastic surface of the button is inhibited, resulting in a flattened crystal morphology. Also, during crystal growth, a depletion zone develops around the crystal in which a lowered concentration of nucleosome core particles is found. This provides a barrier which reduces the rate of addition to the crystal lattice of larger aggregates with correspondingly lower diffusion coefficients. The depletion zone surrounding the growing crystal may also inhibit addition of impurities by slowing the rate of growth of the crystal thereby allowing more precise formation of the energy-minimized structure of the crystal lattice (Carter *et al.*, 1999a).

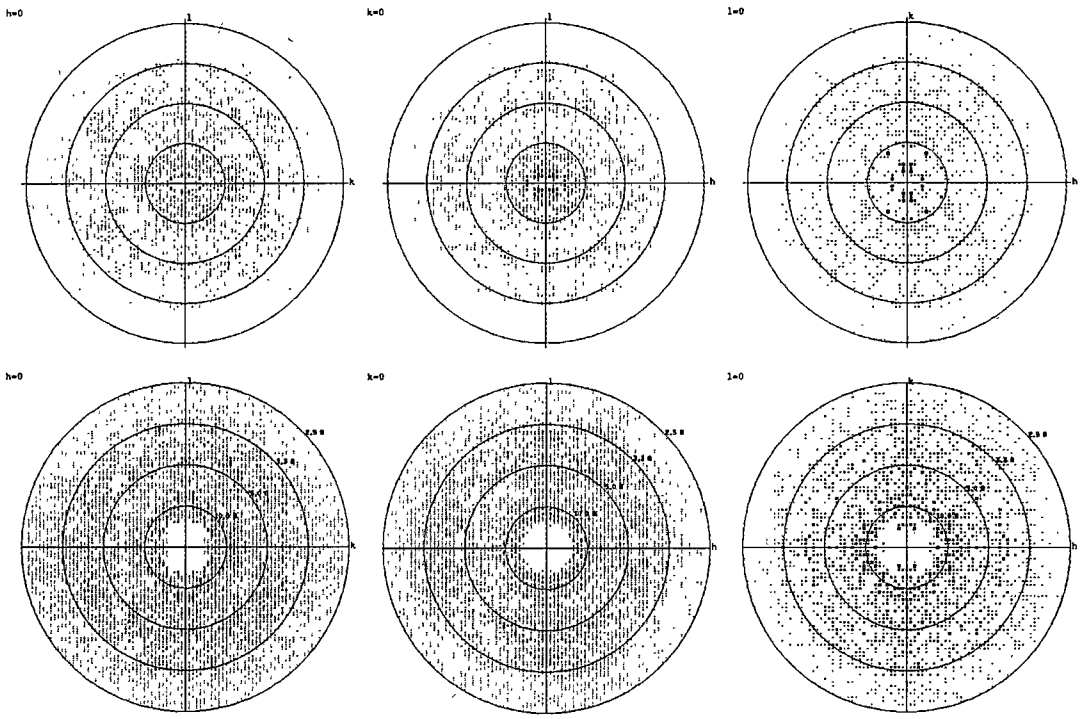

Figure 3.8: A comparison of microgravity versus ground-based diffraction data using pseudoprecession photographs to illustrate completeness and degree of anisotropy. The microgravity data (bottom row) were from an NCP crystal grown on the USML-2 mission and collected at beamline X12-C at NSLS. The ground-based data (top row) are from the best crystal grown on the ground and collected using the same synchrotron beamline.

3.4 Space group determination

The space group of the nucleosome core particle crystals has been previously determined to be the orthorhombic $P2_12_12_1$ space group (Harp *et al.*, 1996) which possesses three nonintersecting 2_1 screw axes parallel to each cell axis with the origin of the unit cell at the midpoint of the three screw axes (Hahn, 1995). In this work, the longest cell axis is defined as the c-axis, in contrast to the convention used by the European workers in which the longest axis is given as the b-axis. This then transposes the b- and c-axis of the European unit cell definition with the c- and b-axes, respectively, as used here. To avoid confusion, this fact should be borne in mind when examining differences between results of the two groups. The four general equivalent positions of the unit cell are generated by the three 2_1 symmetry operators resulting in the following list:

$$x, y, z \quad \overline{x}+\tfrac{1}{2}, \overline{y}, z+\tfrac{1}{2} \quad \overline{x}, y+\tfrac{1}{2}, \overline{z}+\tfrac{1}{2} \quad x+\tfrac{1}{2}, \overline{y}+\tfrac{1}{2}, \overline{z}$$

3.5 Cryogenic data collection and macromolecular crystal annealing

Flash-cooling of NCP crystals was possible using AML containing MPD as a cryoprotectant. However, flash-cooling results were uneven, with perhaps one in five attempts resulting in useful data as the crystals generally exhibited a flash-cooling induced in-

crease in mosaicity by a factor of two- to four-fold. Very few strategies were available to overcome this effect with the most common being to survey a variety of additives to the cryoprotectant solution and a variety of flash-cooling protocols such as liquid nitrogen or liquid propane quenching. During the survey of cryoprotectant additives, a crystal was found to exhibit unacceptably high mosaicity and was removed from the cold nitrogen gas stream of the cryostat. It was placed directly into a drop of the cryoprotectant solution for transfer to a light microscope for measurement and examination. The crystal remained intact and, surprisingly, appeared to be more transparent than before the flash-cooling treatment. The observation of improved visual quality prompted the examination of possible improvement in X-ray diffraction quality. The crystal was again flash-cooled and a diffraction image collected to compare with the original diffraction images. The subsequent diffraction immediately revealed that the quality of the crystal had been dramatically improved. This effect is illustrated in Figure 3.9. The reduction in mosaicity is evident from the decreased width of the lunes and the more sharply defined Bragg reflections in photographs taken over the same rotation angle.

Studies of the phenomenon of macromolecular crystal annealing were designed to answer the following questions:

- Does the cryoprotectant have an effect on the results of the annealing treatment?
- What parameters of the annealing protocol are most important to the outcome

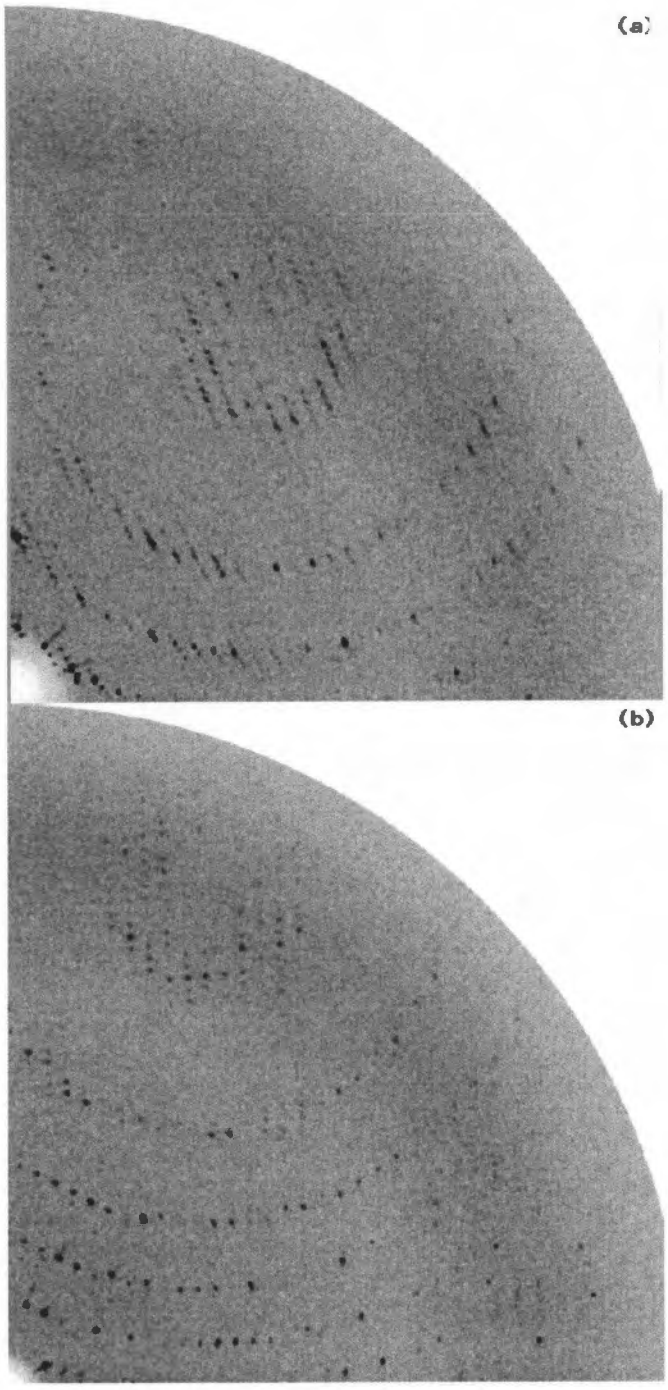

Figure 3.9: A comparison of diffraction from the same nucleosome core particle crystal demonstrating the effect of annealing. The two wedges are taken from images recorded on a Mar image plate area detector using the same rotation angle (0.25^o) and exposure time. Image (a) illustrates typical diffraction after initial flash-cooling. The same crystal (b) was then annealed, resulting in significantly improved diffraction.

of the procedure?

- Can any crystal be annealed or is annealing an effect peculiar to only a few types of macromolecular crystals?
- What is the effect of annealing on data quality at high versus low resolution? Can annealing improve resolution?
- Does the annealing treatment affect the structure of the molecule in the crystal lattice?
- What is the effect of stored radiation damage on an annealed crystal?

The answers to these questions will be discussed individually in later sections.

The first part of the study involved an intensive survey of the combinations and permutations of all the treatments which may have produced the annealing effect in the first, serendipitous annealing. The results of this survey were used to develop a canonical macromolecular annealing procedure which was capable of successfully annealing all crystals available for study. Fortunately, the two crystal systems which were immediately available, the NCP and crystals of the histone octamer core of the nucleosome (Burlingame *et al.*, 1984), proved to be appropriate candidates for the study as both exhibited increases in mosaicity after initial flash-cooling and both were quite sensitive to variations in the parameters of the annealing protocol. The annealing treatments involved either placing the crystals in varying volumes of a variety of

cryoprotectants or warming of the crystal in place on the loop by simply blocking the cold nitrogen gas stream for some time and then flash-cooling by releasing the stream to flow over the warmed crystal. This later variation was termed "annealing on the loop." Annealing on the loop produced inconsistent results and was abandoned in favor of placing the flash-cooled crystal directly into a drop of the cryoprotectant solution. The canonical protocol is given in the Methods but is repeated here.

Protocol for macromolecular crystal annealing:

- Mount crystal in cryoprotectant or oil onto a loop on a magnetic goniometer cap.
- Flash-cool.
- Transfer crystal from the loop quickly into a 0.3 mL or larger drop of cryoprotectant or oil.
- Incubate for a minimum of 3 minutes at crystal growth temperature.
- Remount crystal onto loop.
- Flash-cool.

3.5.1 Does the cryoprotectant have an effect on the results of the annealing treatment?

In the first studies of the annealing phenomenon, it was necessary to determine if the cryoprotective additives, such as the raffinose being surveyed, were responsible for the observed effect on data quality. Crystals were prepared for annealing from AML without additional additives by placing crystals in a drop of Paratone N oil and removing surface solution from the crystal. It was found that annealing was not affected by the presence or absence of cryoprotective additives so long as the crystal was stable in AML or the cryoprotectant. Glycerol, for instance, was not a suitable additive for NCP crystals as diffraction resolution was reduced proportionally to the length of time the crystal spent in a glycerol solution. The use of Paratone N oil, which is a highly purified hydrocarbon oil normally used to modify the viscosity of motor oils, was pioneered by Hope (Hope, 1990). Removing all surface liquid from a crystal immersed in oil prevents ice formation on the surface and, thus, prevents propagation of ice crystallization into the crystal. The technique essentially used the macromolecule as its own cryoprotectant. However, preliminary results using the Paratone oil were not consistent. The observation was made that the oil being used for cryoprotection appeared somewhat cloudy. The oil was placed into a vacuum oven and carefully dried. Annealing results using the dried oil were quite consistent and it appeared that all of the crystals examined could be flash-cooled in dry Paratone N. Dry oil provided the additional benefit of lowered background levels in diffraction

data over the wet oil. Paratone N is now routinely stored in a vacuum desiccator.

3.5.2 Can any crystal be annealed?

It was considered to be possible that the annealing phenomenon was a peculiar feature of the NCP and octamer crystals. To determine the general applicability of MCA, a number of crystals systems were examined (see Table 3.2). Results from other laboratories were also collected to extend the range of crystal systems examined. The results are shown in Table 3.3. With the notable exception of concanavalin A, all crystal systems for which a suitable cryoprotectant was available were successfully annealed.

The publication of a protocol termed "flash-annealing" (Yeh & Hol, 1999) prompted a reexamination of annealing protocols. In flash-annealing, the cold nitrogen gas stream of the cryostat is blocked for 1.5-2 seconds while the crystal remains on the loop. The gas stream is unblocked and the crystal allowed to recool for six seconds. The procedure is repeated twice for a total of three rounds of warming and flash-cooling. The study was based on results obtained using a single crystal system which showed dramatic improvement in data quality. The protocol resembled earlier attempts to understand annealing here in which the cold nitrogen gas stream was blocked while the crystal remained on the loop. The crystal was observed on a video monitor connected to a CCD camera mounted within the Mar image plate appara-

Table 3.2: Crystal systems used for the study of macromolecular crystal annealing.

Crystal System	Solvent fraction	Precipitant	Cryoprotectant
Concanavalin A I222 a=89.6,b=86.5,c=62.1 Å	0.49	MPD	MPD/glycerol Paratone N
Proteinase K $P4_32_12$ a=b=68.3,c=108.4	0.46		glycerol
Nucleosome core particle $P2_12_12_1$ a=105,b=110,c=180 Å	0.51	Mn^{++}	MPD/Paratone N
Chicken histone octamer $P3_221$ a=b=115,c=100 Å	0.65	NH_4SO_4	glycerol
Lysozyme (/+ N-acetyl glucosamine) $P4_32_12$ a=b=78.4,c=37 Å	0.37	NaCl	glycerol
Sperm whale myoglobin $P2_1$ a=64.7,b=30.8,c=35.1 Å	0.34		glycerol

Table 3.3: Crystal systems that have been successfully annealed in addition to those used in this study.

fumarylacetoacetate hydrolase (unpublished data)	
space group	$P2_1$
unit cell	a=64.0 Å b=11.0 Å c= 67.0 Å β=102.0^o
precipitant	PEG 8000
cryoprotectant	PEG 400
alcohol dehydrogenase (T. Hurley, pers. comm.)	
space group	$P2_1$
unit cell	a=55.7 Å b=100.2 Åc=69.0 Å β=104.9^o
precipitant	PEG 6000
cryoprotectant	PEG 200
aldehyde dehydrogenase (T. Hurley, pers. comm.)	
space group	$P2_1$
unit cell	a=102.0 Å b=177.0 Å c=101.0 Å β=94.6^o
precipitant	$MnCl_2$, MPD
cryoprotectant	ethylene glycol
crotonase (T. Hurley, pers. comm.)	
space group	$P3_1$
unit cell	a=76.5 Å b=76.5 Å c=214.6 Å
precipitant	PEG, NaCl
cryoprotectant	MPD
pyruvate dehydrogenase kinase (N. Steussy, pers. comm.)	
space group	P2
unit cell	a=72.0 Å b=109.0 Å c=73.0 Å β=102.0^o
cryoprotectant	Paratone N
purine operon repressor (S. Sinha, pers. comm.)	
space group	P1
unit cell	a=65.4 Å b=72.5 Å c=83.4 Å
	α=84.7^o β=84.6^o γ=67.5^o

tus. Crystals invariably became opaque when the cold gas stream was blocked and cleared as the crystal rewarmed on the loop. The transition from opaque to clear was observed and the cold gas stream immediately restored. The technique, which was termed annealing-on-the-loop, was found to be difficult to control and provided inconsistent results. Annealing-on-the-loop differs from flash-annealing in the timing of the warming period while the cold gas stream is blocked and in the number of iterations of the warming and flash-cooling steps. Annealing-on-the-loop does not place a time limit on warming but depends on observation of clearing of the crystal. The limitation of 2 seconds for warming is clearly not sufficient for most of the crystal tested. Flash-annealing requires three iterations of the warming and flash-cooling steps while annealing-on-the-loop requires only a single treatment. Examination of the results reported in Table 3.4 suggest that the success of the annealing-on-the-loop variation of macromolecular crystal annealing is correlated with the solvent content of the crystal. The highest solvent content crystals, the histone octamer, could not be annealed on the loop. The second highest solvent content crystal, the nucleosome core particle, demonstrated highly variable results. One of the greatest technical problems with crystals which require longer warming times for annealing is drying while exposed to ambient air conditions on the loop.

Table 3.4: Summary of the results of annealing treatments on the crystal systems used in this study. Annealing-on-the-loop was compared to standard macromolecular crystal annealing and related to the solvent content of the crystal unit cell. Results of annealing on the loop for the nucleosome core particle crystals required removal of surface liquid prior to flash-cooling for success. Success for an annealing protocol was defined as not producing a discernable deterioration in diffraction quality.

Molecule	solvent fraction	successful annealing	
		on the loop	MCA
sperm whale myoglobin	0.34	Yes	Yes
lysozyme(/+ N-acetyl glucosamine)	0.37	Yes	Yes
Proteinase K	0.46	Yes	Yes
Concanavalin A	0.49	No	No
nucleosome core particle	0.51	No/Yes	Yes
chicken histone octamer	0.65	No	Yes

3.6 Diffraction analysis of nucleosome core particle crystals

A crystal grown on the USML-2 mission was used to collect data at beamline X12C at the National Synchrotron Light Source. The resolution of the data extended to 2.5 Å. That data was used for the current high-resolution structure. As discussed previously, no ground-based crystals provided data of the same quality due primarily to anisotropy in ground-based crystals and in overall diffraction quality. Figure 3.8 provides a graphical illustration of the difference in microgravity versus ground-based crystals of the NCP using pseudoprecession photographs. The ground-based crystal was grown using a DCAM in the laboratory and provided the best quality data recorded from ground-based crystals during these studies. It was flash-cooled and annealed (Harp *et al.*, 1998; Harp *et al.*, 1999) and data collected using beam line X12C at the National Synchrotron Light Source. As is evident in Figure 3.8, the microgravity crystal does not exhibit the degree of anisotropy typically seen in ground-based crystals of the NCP which may exhibit reflections high resolution reflections in shells for which the data are largely incomplete and therefore useless. The microgravity crystal used in this study flash-cooled well and was used before the phenomenon of macromolecular crystal annealing had been discovered. Diffraction statistics for the microgravity crystal and for the di-μ-iodobis(ethylenediamine)diplatinum (PIP) heavy atom derivative crystal are provided in Table 3.5. The PIP derivative crystal

Table 3.5: Crystallographic diffraction and phasing statistics

	Native	PIP
Observations	605950	379171
Unique reflections	72200	32814
Completeness	98.7%	93.7%
Resolution	25-2.5 Å	40-3.3 Å
Linear R factor	0.045	0.095
Cullis R factor		0.80
Phasing power		2.00

data were collected in-house.

3.7 Fitting and refinement of the model

The experimental phases obtained from SIR and MLPHARE refinement were combined with molecular replacement phases using coordinates from the histone octamer core of the nucleosome (PDB ID code 2HIO) (Arents *et al.*, 1991) and the resulting map used to verify the placement of residues in the MR map. Additional phasing information for fitting of the DNA was provided by molecular replacement using coordinates from the nucleosome core particle structure (PDB ID code 1AOI). Solvent molecules were placed using the program DDQ (van den Akker & Hol, 1999) to locate

Table 3.6: Refinement statistics

Resolution range	25-2.5 Å
Observations	72200
Observations in test set	7303
Test set size	10.1%
R value	0.219
Free R value	0.282
estimated error of Free R value	0.003

positive difference density peaks in an mF_o-F_c electron density map calculated using the model with waters intentionally removed. A sigma cutoff of 3 was used and the program utilized hydrogen bonding distance restraints to differentiate putative solvent peaks from shift peaks produced by errors in placement of atoms in the model or from errors in temperature-factor refinement. Each of the putative water molecules were examined in both the hydrated mF_o-F_c and $2mF_o$-DF_c maps. Subsequent iterations of the protocol using mF_o-F_c maps calculated with the accepted water molecules located additional water molecules. Refinement in CNS of the entire structure with cross-validation and maximum likelihood refinement targets was completed after each round of interactive fitting using Powell and simulated annealing energy minimization protocols. Statistics for the model are given in Table 3.6.

3.8 Hydration and ion binding

3.8.1 Ions

A total of 15 Mn^{2+} ions were identified in the structure (see Table 3.7). Most were bound directly to the N7 atom of purine bases at an average distance of 2.45±0.19 Å. Ten Mn^{2+} ions were bound to the N7 atom of guanine and two were bound to the N7 atom of adenines at the 5' termini. Mn^{2+} 461 appears to be bound octahedrally through water bridges in the minor groove of the DNA with one of the water bridges clearly visible. Of the identified waters, HOH 256, may be a similarly coordinated Mn^{2+}. Two Mn^{2+} ions, MN 466 and 473, appear to be shared between the N7 atoms of adjacent guanine bases. Potassium ions (see Table 3.8) were similarly bound to the N7 atoms of purines with adenines and guanines equally likely. The average distance of the potassium ions from an N7 atom was 3.36±0.19 Å for the seven potassiums.

Chloride ions were identified as being bound in two positions which are equivalent by dyad symmetry. The chloride is tightly bound in a pocket formed by main chain amide nitrogens of Ala 45, Gly 46, Ala 47 of H2A and by the amide nitrogen of Ser 91. The amide nitrogen of Ala 45 of H2A is in direct contact with a DNA phosphate. The proximity of the negatively charged phosphate is problematic but it may be supposed that the chloride bound to the octamer prior to DNA binding in the reconstitution reaction. A sulfate ion in the low-temperature octamer structure binds strongly in the same location providing an ion bridge stabilizing the two histone chains in the same

Table 3.7: Tabulation of Mn^{2+} ions in the NCP structure. MN 461 is octahedrally bound in the minor groove. Water, HOH 256, was identified as a water but may also be a Mn^{2+} octahedrally bound in the minor groove.

MN	bound to	distance (Å)	temperature factor
459	N7 Gua 280	2.30	61.2
460	N7 Gua 134	2.53	57.4
461	minor groove		69.3
462	N7 Gua 246	2.45	62.6
463	N7 Gua 121	2.38	73.3
464	N7 Gua 70	2.07	72.9
465	N7 Gua 39	2.54	97.7
466	N7 Gua 138	3.48	108.5
	N7 Gua 137	3.19	
467	N7 Gua 185	2.45	92.9
468	N7 Gua 267	2.33	82.9
469	N7 Ade 147	2.80	85.2
470	N7 Gua 100	2.39	97.2
471	N7 Ade 1	2.72	105.8
472	N7 Gua 18	2.47	124.2
473	N7 Gua 125	3.21	105.0
	N7 Gua 124	3.00	

Table 3.8: Tabulation of potassium ions in the structure

K	bound to	distance (Å)	temperature factor
452	N7 Ade 56	3.53	102.4
453	N7 Gua 216	3.42	83.9
454	N7 Gua 217	3.18	78.2
455	N7 Gua 227	3.04	95.1
456	N7 Ade 228	3.32	83.8
457	N7 Ade 175	3.48	119.9
458	N7 Gua 98	3.54	108.2
	N7 Gua 97	3.56	

manner that the chloride ion in the NCP stabilizes the C-terminus of the β-bridge as discussed above.

The single most intense peak in the $2mF_o$-DF_c electron density map was identified as a cacodylate ion. The identity of the ion was confirmed using omit maps and by replacing the cacodylate with a Mn^{2+} ion. In a $2mF_o$-DF_c map calculated with the cacodylate replaced by Mn^{2+}, the electron density displayed distinct lobes in which the cacodylate was fitted (see Figure 3.10). The cacodylate is placed within a pocket formed by side chains of Gln 76 and Asp 77 of 2H3 and the main chain carbonyl of Leu 22 of 2H4 on the NCP and by side chains of Glu 64 of 1H2A and His 49 of 2H2B as well as the main chain carbonyl of Val 48 of 1H2B on the symmetry neighbor.

3.8.2 Structural water

A total of 349 waters were identified in the structure. Forty-two, or 10%, are hydrogen bonded to main chain amide nitrogens. Of those, a total of 24 are hydrogen bonded only to amide nitrogens, 4 are involved in water bridges between adjacent main chain amide nitrogens on the same chain and one makes a water bridge between chains 2H3 and 2H4, and 13 are involved in forming water bridges between main chain amide nitrogens and DNA phosphates. There are 43 identified waters which are hydrogen bonded only to DNA phosphates while 24 are bound to DNA bases. Many of the remaining waters are involved in forming water networks with the histone core of the

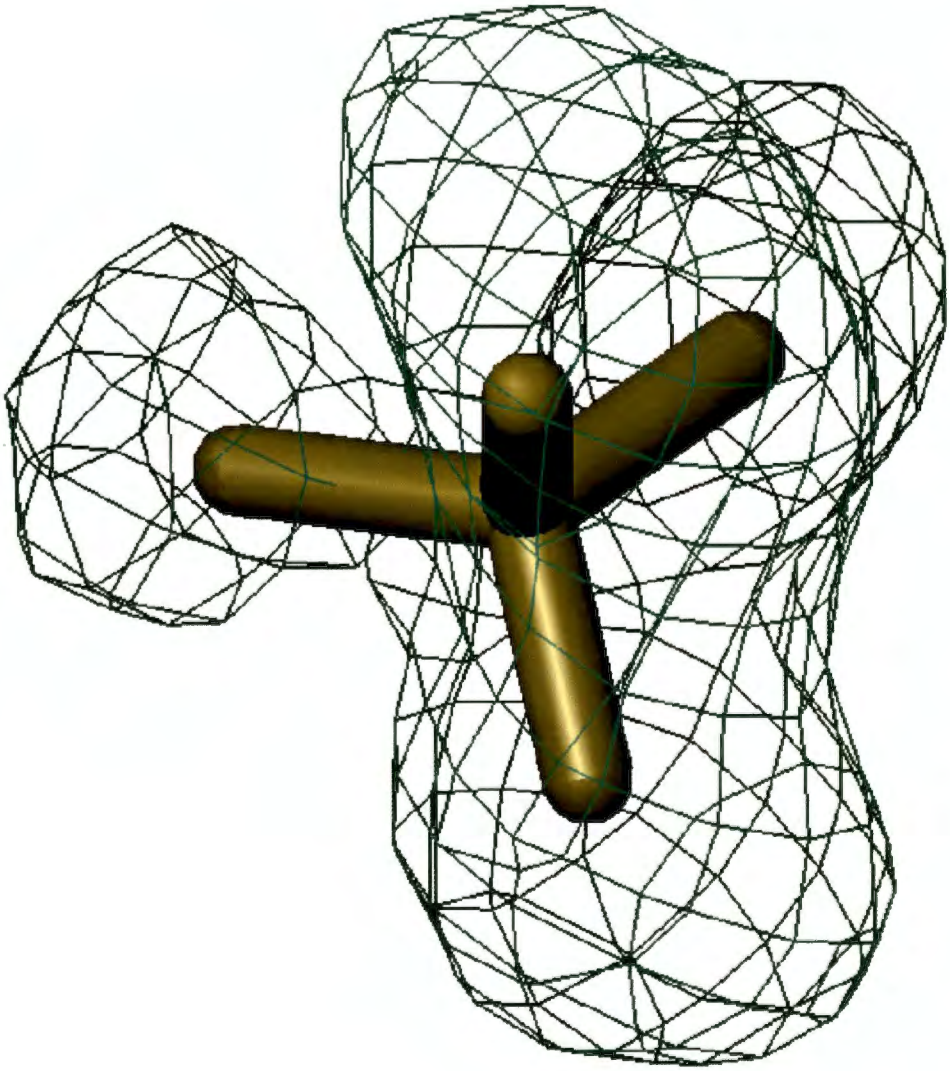

Figure 3.10: The electron density displayed at 3 times the σ level is shown for the $2mF_O$-DF_C electron density map calculated with the cacodylate ion replaced with a Mn^{2+} ion at the position of the arsenic atom. A model of the cacodylate has been placed within the resulting density to demonstrate the correspondence with the structure of the cacodylate ion.

NCP.

3.9 Structure of the nucleosome core particle at 2.5 Å

3.9.1 General features of the molecular structure

The model presented here represents a significant extension of the atomic position data for the NCP. We are able to model 152 more amino acids than were seen in the previously published 2.8 Å model (PDB ID code 1AOI) and the overall B-factors, especially for the DNA, are significantly lower. All DNA was fitted into observed density and no modeling of ideal structure was done to complete the structure. The DNA fragment used in this study includes terminal phosphates making this the longest fragment of DNA for which atomic position data exists. Ions and water molecules are seen in this structure and their modeled positions provide insight into the ordering of larger chromatin arrays.

The engineering of the NCP crystal used all of the constituent features to optimize symmetry within the NCP. Gentle salt extraction of native histone octamers from the mostly inactive nuclei of avian mature erythrocytes has provided a homogeneous population of unmodified histones with unaltered and thus physiologically authentic folding. The use of an engineered DNA palindrome based on a sequence designed by

nature to possess strong phasing signals, α-satellite DNA, was intended as a way to extend the two-fold symmetry of the histone octamer and result in NCPs with no dorsal-ventral asymmetry. Possibly due to the necessity of mutation at the dyad to introduce a restriction site, this prediction was off by 0.5 base pairs.

Arbitrary anatomical designations (see Figure 3.11) have been made to facilitate discussion of the structure of the NCP. The portion of the molecule where the DNA enters and exits the NCP is referred to as the top of the structure. The dyad axis of the histone core of the NCP divides the octamer into halves with each half, designated as side 1 and side 2 possessing one each of the histone proteins. Thus, the 2 subunits of histone H2A are designated as 1H2A or 2H2A depending on whether the subunit is placed on side 1 or side 2, respectively. The molecule is obviously asymmetric with one side having greater order in the tail regions. Asymmetry is also seen as a consequence of the dyad intersecting DNA base 73. This results in positioning 72 base pairs to the side of the dyad on one face, the ventral face, and 73 base pairs to the other face, the dorsal face. Remarkably, the asymmetries did not result in two-fold packing disorder. The fortunate alteration of dorsal and ventral orientations of NCPs in the lattice followed the continuous path of DNA base stacking between NCPs. The asymmetry of the DNA binding to the octamer core in this structure has provided insights into the plasticity of the protein-DNA interactions in the nucleosome, allowing formation of nucleosomes on random sequences of DNA throughout the nucleus. Binding of DNA to the superhelical ramp of the histone octamer is primarily mediated through

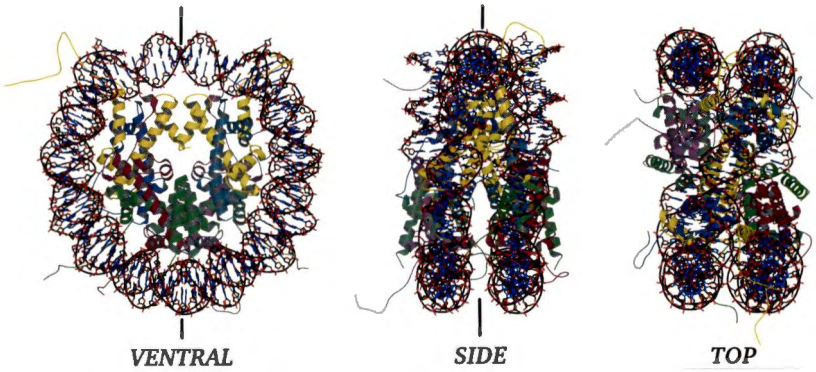

Figure 3.11: Views of the nucleosome core particle with the protein represented by ribbon models of the secondary structure elements, and DNA indicating base pairing between complementary strands. The pseudo-dyad axis is represented by vertical bars for both the ventral and side view. The pseudo-dyad axis passes through the center of the dyad view orthogonal to the plane of the page.

a series of binding sites constructed around an arginine residue whose side chain inserts into the minor groove of the DNA. These positions mark the closest approach of the minor groove to the protein surface and are here defined as minor groove-in positions or mGIs.

Although the histone octamer structure displays a two-fold symmetry about a central axis, the structures of the individual histones in the NCP deviate from symmetry in a number of positions. Least-squares fitting of the octamer moiety of the NCP onto the histone octamer structure indicates clearly that the disposition of the protein subunits deviates from symmetry upon binding with DNA. Figure 3.12 shows the displacement of the helices of individual histone subunits from the symmetrical histone octamer. This figure demonstrates some remodeling of the histone structure with DNA bound, particularly with respect to histone H2A and H4. Figure 3.13 similarly represents a

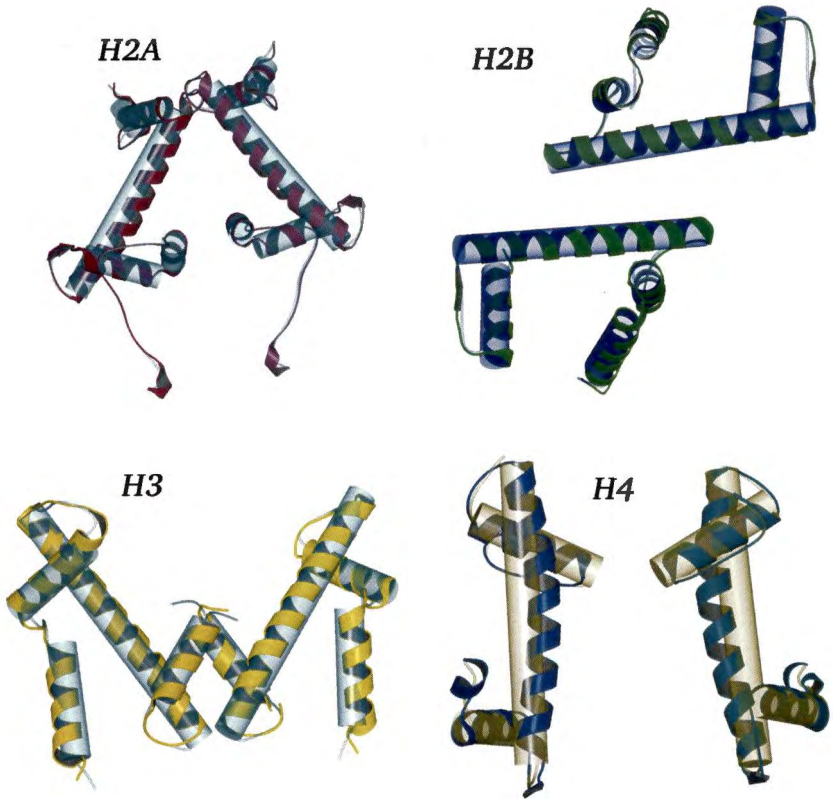

Figure 3.12: Ribbon models of histone α-carbons showing least-squares fitting of individual histone subunits from the octamer of the NCP compared with the same subunits in the histone octamer without DNA bound. Colored models represent the histone structure within the NCP while the transparent cylinders represent the boundaries of the low temperature histone octamer structure. Deviations from symmetry induced by binding of DNA are apparent from the differences in the disposition of helices.

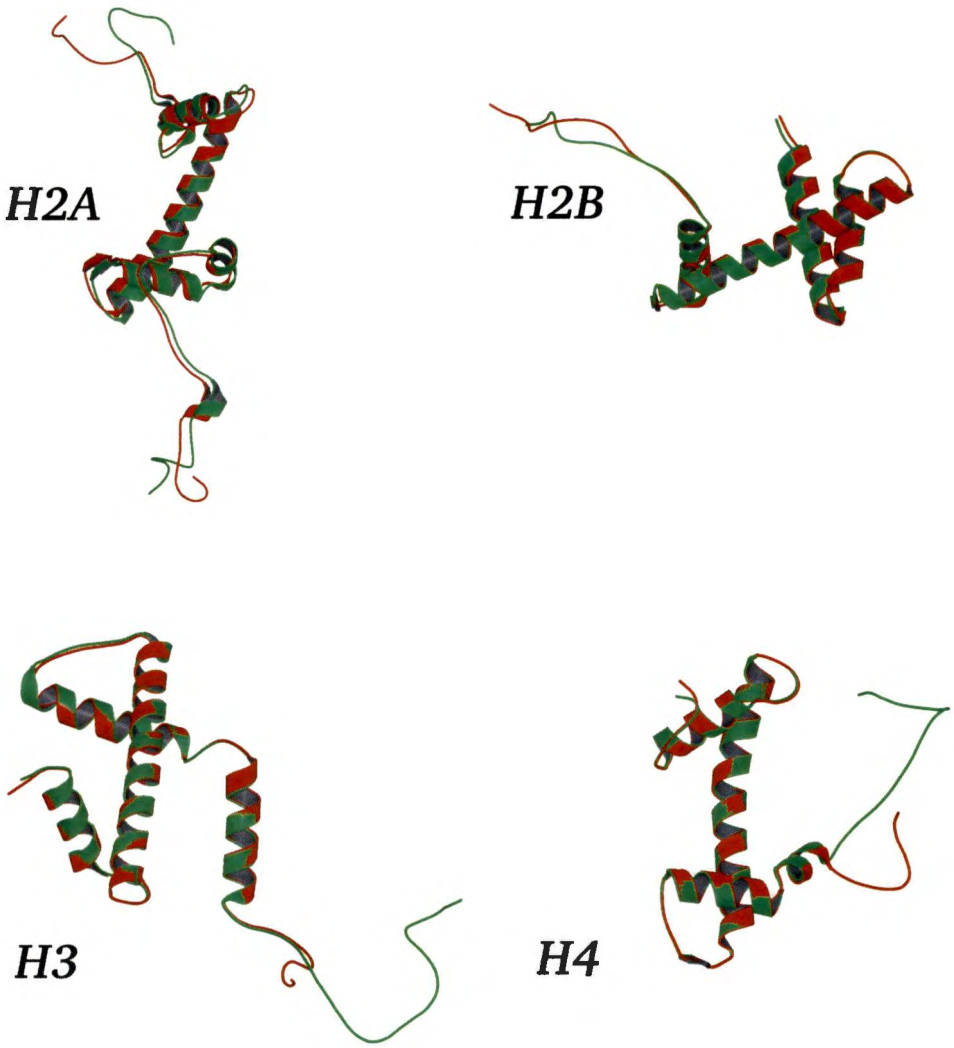

Figure 3.13: Ribbon models representing a least-squares fitting of homologous histone subunits in the NCP to reveal locations of asymmetry between the two halves of the histone octamer core of the NCP structure.

least squares fitting of the coordinates of homologous subunits except that only the histones with the NCP structure are used. This more detailed rendering of the structures shows that H3 and H4 are least impacted by DNA binding and that the dimer histones show the largest degree of remodeling. However, as will be discussed later, DNA binding is not the only factor influencing the remodeling seen here. Stacking interactions between the histone faces of adjacent NCPs are largely responsible for the asymmetry in the histone subunit structures.

The structure of the NCP is complex with numerous protein-protein and protein-DNA contacts. Discussion of the structure will involve separate discussions of the structure of the individual histone proteins, how the histones combine to construct the histone octamer core, the binding of DNA to the histones, interaction of the histone tails with DNA on the same and neighboring particles with the crystal lattice, and details of the structure of DNA as it exists on the nucleosome.

3.9.2 The histone fold and the handshake motif

The four core histones each possess a similar central structural design, the histone fold (Arents *et al.*, 1991), possibly produced by the symmetrical duplication of a helix-turn-helix motif (Arents & Moudrianakis, 1995; Ramakrishnan, 1995). The strongly conserved architecture of the histone fold motif consists of a long central helix with shorter helices attached by loop structures at the N-terminal and C-terminal

ends. The histone fold is illustrated in Figure 3.14. The elements of the histone fold structure therefore consist of three helices: a short helix on the N-terminal side of the symmetry center of the fold, NH, the long median helix, mH, and a short helix on the C-terminal side, CH. The helices are joined by loop NL between the NH helix and the mH helix, and loop CL between the median helix and the CH helix. The short terminal helices are folded back and rotated over the median helix. This structure is seen in a number of DNA binding proteins belonging to the histone fold family (Baxevanis *et al.*, 1995), which were identified after the first description of the histone fold motif (Arents *et al.*, 1991). The core histones assemble as heterodimers, (H2A·H2B) and (H3·H4), by interdigitation of the terminal helices and overlapping of the central helices as shown in Figure 3.14b. The tight association of histones in the handshake motif is characterized by extensive hydrophobic contacts between the helices and by a short β-bridge structure formed between the CL loop of H2A and the NL loop of H2B in the (H2A·H2B) heterodimer and between the CL loop of H3 and the NL loop of H4 in the (H3·H4) heterodimer. The β-bridges form part of the primary DNA docking sites on the histone surface, that are described later. The close contact of the histones in the heterodimers was described as a handshake motif by Moudrianakis and coworkers (Arents *et al.*, 1991).

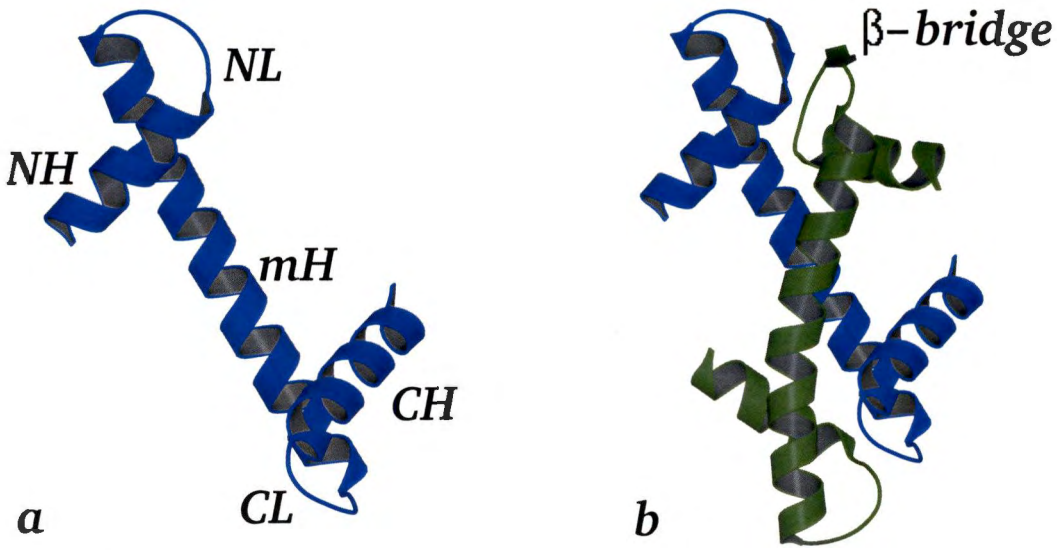

Figure 3.14: Structure of the histone fold motif and handshake interaction of histones. a. Structure of the histone fold motif using a ribbon α-carbon model of H4 as exemplar. The histone fold is constructed from 3 helices; the N-terminal helix, NH, the central or median helix, mH, and the C-terminal helix, CH, joined by loops; the N-terminal loop, NL, and the C-terminal loop, CL. b. Ribbon α-carbon model of H3·H4 heterodimer showing the 2 histones interdigitated in a handshake motif. The CL loop of one subunit is involved with the NL loop of the other in a short β-bridge.

3.9.3 Structure of the $(H3{\cdot}H4)_2$ tetramer

The heterotetramer $(H3{\cdot}H4)_2$ is referred to simply as the tetramer. The tetramer alone is capable of organizing DNA into a nucleosome-like structure (van Holde, 1988) which is termed the "tetrasome" by Prunell and colleagues (Hamiche *et al.*, 1996). The tetramer is assembled from (H3·H4) heterodimers by formation of a four-helix bundle composed of the mH and CH helices of the two H3s (see Figure 3.15). The four-helix bundle represents the only location within the core structure of the histone octamer at which an extensive interaction is made between homologous subunits. The location of the four-helix bundle immediately below the center of the DNA bound in the nucleosome suggests a pivotal role for this structure, and it will be occasionally referred to a the "hinge" region of the nucleosome.

3.9.4 Structure of the individual histone proteins

In addition to the basic architecture of the core histones, individual histones may possess accessory helices N-terminal or C-terminal to the histone fold motif. Each of the four core histones possesses an unstructured N-terminal tail which extends beyond the boundaries of the histone octamer structure to make contacts with intraparticle DNA as well as numerous interparticle contacts. An unusual feature of the NCP crystals is the extent and variety of interparticle contacts made by the N-terminal unstructured tails. Histone H2A also displays an extensive C-terminal tail which,

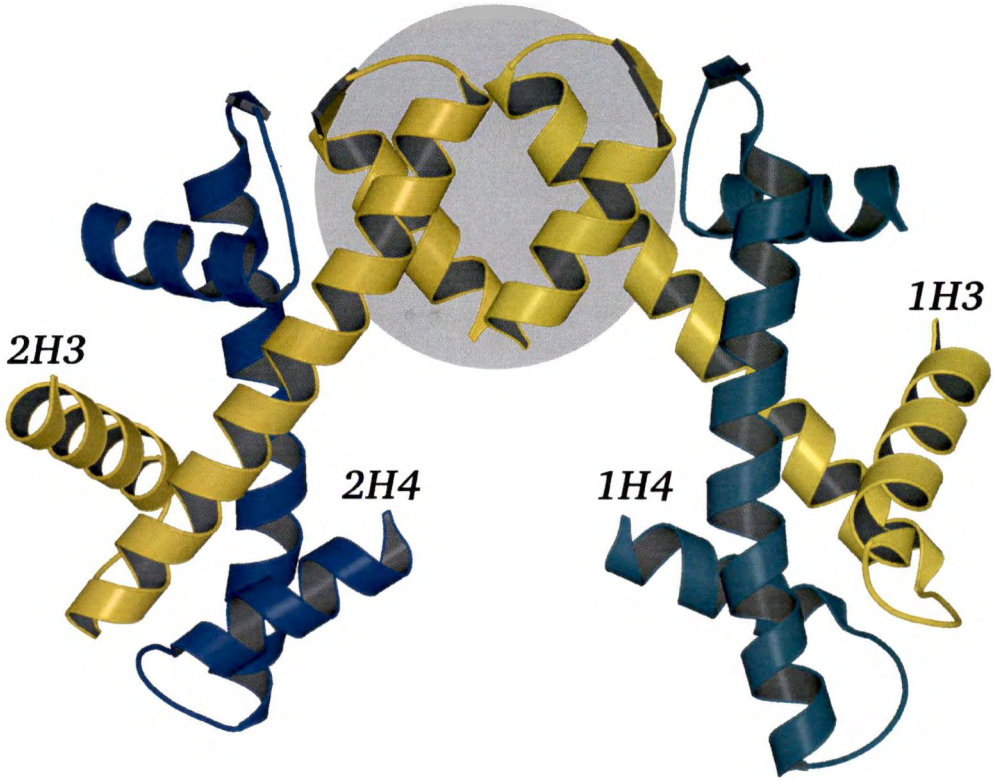

Figure 3.15: The histone fold domains of histones H3 and H4 associate in the handshake binding motif to form heterodimers which then form the $(H3{\cdot}H4)_2$ heterotetramer through the formation of a four-helix bundle at the C-termini of the 2 H3 subunits. The four-helix bundle is indicated with a grey background.

although random coil, is reasonably well ordered within the overall structure of the histone octamer and does not appear to be involved in interparticle interactions.

Alignments of the secondary structures of the individual histone subunits with their respective amino acid sequences are presented in Figure 3.16, Figure 3.17, Figure 3.18, and Figure 3.19. The secondary structure of each of the histones was examined by first passing the atomic coordinates of each histone chain through a secondary structure prediction program, DSSP (Kabsch & Sander, 1983). The results were checked by direct examination of the structure. In particular, the assignment of helices was checked by verifying main chain hydrogen bonding patterns. These figures show the extent of the chains observed in the octamer structure (PDB code 2HIO) (Arents *et al.*, 1991), that defines the boundaries of the unstructured tail regions of the protein in the absence of stabilizing interactions with bound DNA. The extent of the chains observed in the NCP structure are also indicated. Residues forming the histone fold of each subunit are shown within solid lines and residues within accessory helices are shown inside dashed lines. The examination of the structure of the histones in the NCP was facilitated by comparisons with the 2HIO histone octamer structure and this laboratory's low-temperature histone octamer structure, that will be presented in a separate publication (Harp *et al.*, 2000b). The low-temperature structure is characterized by higher resolution and details of ion binding. References to the histone structure refer to our 2.7 Å low-temperature structure.

Histone H2A (see Figure 3.16) possesses a 14-residue unstructured N-terminal tail

1H2A

```
  ν                    8°
N-SGRGKQGGKARAKAK -SRSSRAG- LQF -
                    16    22

PVGRVHRLLRKG - NYAERVG -AGAPVYLAAVLEYLTAEILELAGNAARDN- KKTRI -
26          37          45                          73
                                                 8°          ν
IPRHLQLAIRN D EELNKLL- GKVTIAQGGVLPNI -QAVLL- PKKTDSHKAKAK-C
79          90     97                 112  116
```

2H2A

```
  ν                    8°
N-SGRGKQGGKARAKAK -SRSSRAG- LQF -
                    16    22

PVGRVHRLLRKG - NYAERVG -AGAPVYLAAVLEYLTAEILELAGNAARDN- KKTRI -
26          37          45                          73
                                                 8°
IPRHLQLAIRN - DEELNKLL-GKVTIAQGGVLPNI -QAVLL- PKKTDSHKAKAK-C
79          89 90     97                 112  116
```

Figure 3.16: Secondary structure alignment with amino acid sequence of histone H2A. Helices belonging to the histone fold are in solid lines and accessory helices are within dashed lines. The extent of the chain visible in the NCP structure is denoted by ν over the final residue and the extent of the chain visible in the octamer structure is denoted by 8° over the terminal residue.

H2B

$\nu_1\nu_2$ 8°

N-PEPAKSAPAPKKGSKKAVTKTQKKGDKKRKKSRKES -

YSIYVYKVLKQVH - PDTGI - SSKAMGIMNSFVNDIFERIAGEASRLAHYN -
37 49 55 84

KRSTI - TSREIQTAVRLLL - PGELAKHAVSEGTKAVTKYTS - SK-C
90 102 103 123

Figure 3.17: Secondary structure alignment with amino acid sequence of histone H2B. Helices belonging to the histone fold are in solid lines and accessory helices are within dashed lines. The N-terminal extent of the chain visible in the NCP structure is denoted by ν_1 for 1H2B and ν_2 for 2H2B. The N-terminal extent of the chain visible in the octamer structure is denoted by 8° over the terminal residue. The C-terminus of the chain is visible in both structures.

H3

ν_2 ν_1 8°

N-ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRP -

GTVALREIRRYQKS - TELLI - RKLPFQRLVREIAQDFK - TDLRF -
44 57 63 79

QSSAVMALQEASEAYLVGLFEDTNLCAIHA - KRVTI - MPKDIQLARRIR - GERA-C
85 114 120 131

Figure 3.18: Secondary structure alignment with amino acid sequence of H3. Helices belonging to the histone fold are in solid lines and accessory helices are within dashed lines. The N-terminal extent of the chain visible in the NCP structure is denoted by ν_1 for 1H3 and by ν_2 for 2H3. The N-terminal extent of the chain visible in the octamer structure is denoted by 8°. The C-terminus is visible in both structures.

H4

ν_2 ν_1 8°

N-SGRGKGGKGLGKGGAKRHRKVLR -DNIQGI-TKPAIRRLARRG-GVKRISGL -

24 29 30 41

IYEETRGVLKVFLENVIRDAVTYTEHA- KRKTV -TAMDVVYALKRQ- GRTLYGFGG-C

49 76 82 93

Figure 3.19: Secondary structure alignment with amino acid sequence of H4. Helices belonging to the histone fold are in solid lines and accessory helices are within dashed lines. The N-terminal extent of the chain visible in the NCP structure is denoted by ν_1 for 1H4 and by ν_2 for 2H4. The N-terminal extent of the chain visible in the octamer structure is denoted by 8°. The C-terminus is visible in both structures.

Histone H2A (see Figure 3.16) possesses a 14-residue unstructured N-terminal tail beyond Lys 15 extending around the DNA gyres between mGI 4 and mGI 5 and interacting with both intraparticle DNA and with the N-terminal tail of H2A from the opposite side of a neighboring NCP. That is, 1H2A interacts with the N-terminal tail of 2H2A of a symmetry neighbor within the crystal lattice (see Figure 3.20 and Figure 3.21). The entire tail is visible in the NCP electron density map for 1H2A and to Gly 2 in 2H2A. At the base of the tail an N-terminal accessory helix, NaH, interacts directly with the DNA phosphate backbone.

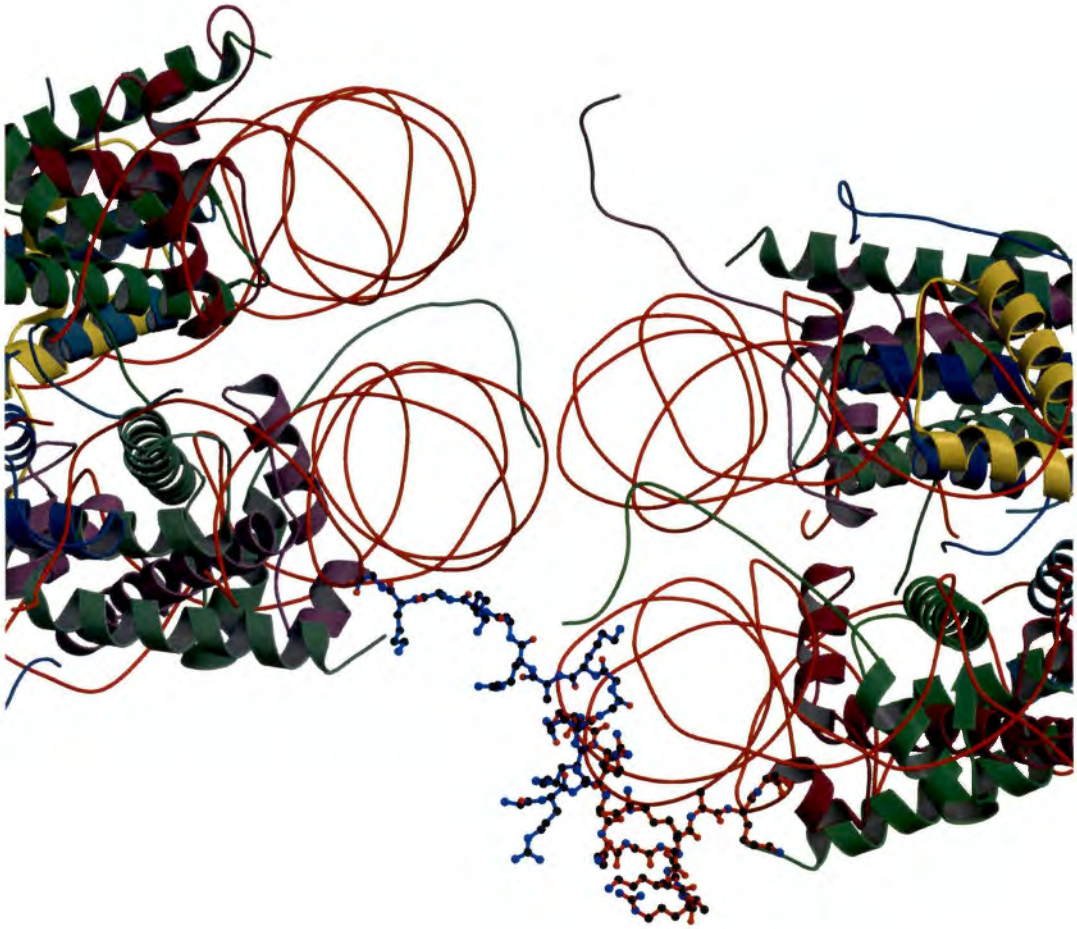

Figure 3.20: N-terminal tails of H2A are shown as ball and stick models to show the interaction of the H2A tails between neighboring NCPs. The carbon atoms of the N-terminal tail of 2H2A are shown in blue and those of 1H2A are in red. The tail of 2H2A extends across and interacts with the N-terminal tail of 1H2A on the neighboring NCP. At the base of the tail, a short accessory helix interacts directly with the DNA.

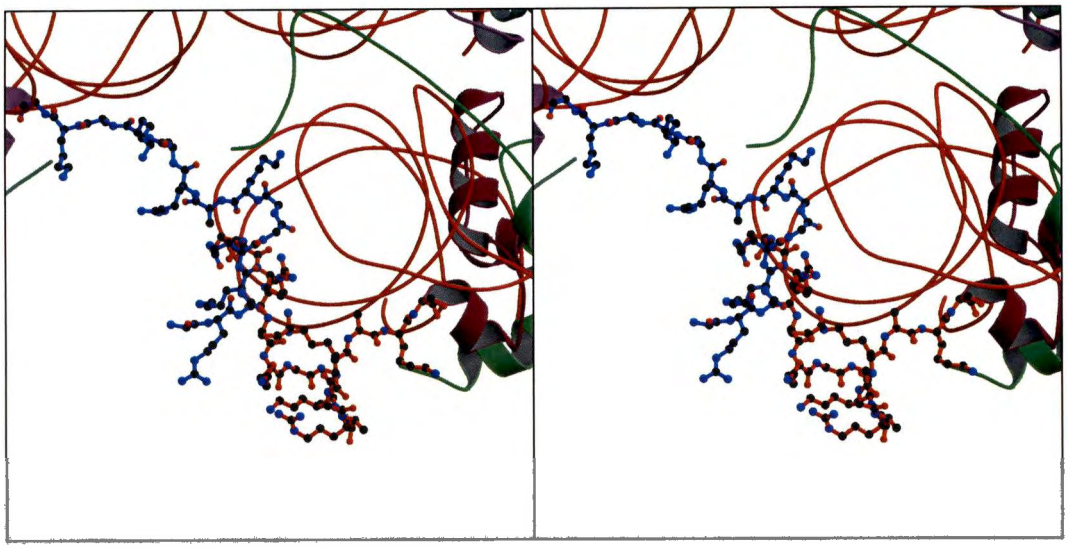

Figure 3.21: Stereo pair representing the interaction between H2A N-terminal tails of neighboring NCPs. The region shown in this figure is a magnification of the region of interaction between N-terminal tails of H2As in the previous figure.

H2A extends well beyond the histone fold as seen in Figure 3.22 and is the only one of the four core histones with a significant C-terminal tail. This tail is visible only to Leu 115 in the 2HIO histone structure leaving the 13 C-terminal residues disordered. Histone 1H2A is the only one of the eight protein subunits for which all of the C-terminal residues are not apparent in the NCP electron density map with the 3 C-terminal residues not visible. The H2A chain turns immediately at the C-terminus of the CH helix into a 3_{10} accessory helix, CaH, which then extends across the face of the octamer and into a random coil structure which serves to bind the (H2A·H2B) dimer to the $(H3{\cdot}H4)_2$ tetramer. The significance of a 3_{10} accessory helix is unclear but may reflect the fact that the C-terminal tail of H2A is not a disordered region in the same sense as the N-terminal tail domain.

The N-terminal disordered tail of histone H2B extends for 38 residues with only a few residues N-terminal to helix NH visible in the octamer structure. The N-terminal tail of H2B extends between DNA gyres with 17 residues of 1H2B and 18 of 2H2B visible in the NCP electron density map. The H2B chain bends abruptly past the CH helix and forms a well ordered accessory helix, CaH, which forms a large portion of the face of the histone core (see Figure 3.22). The histone fold of H2B interacts with the histone fold of H4 in the primary interaction binding the (H2A·H2B) dimer to the $(H3{\cdot}H4)_2$ tetramer.

Histone H3 N-terminal tails extend well beyond the DNA gyres near the entry and exit point of the DNA in the NCP. None of the N-terminal tail is visible in the histone

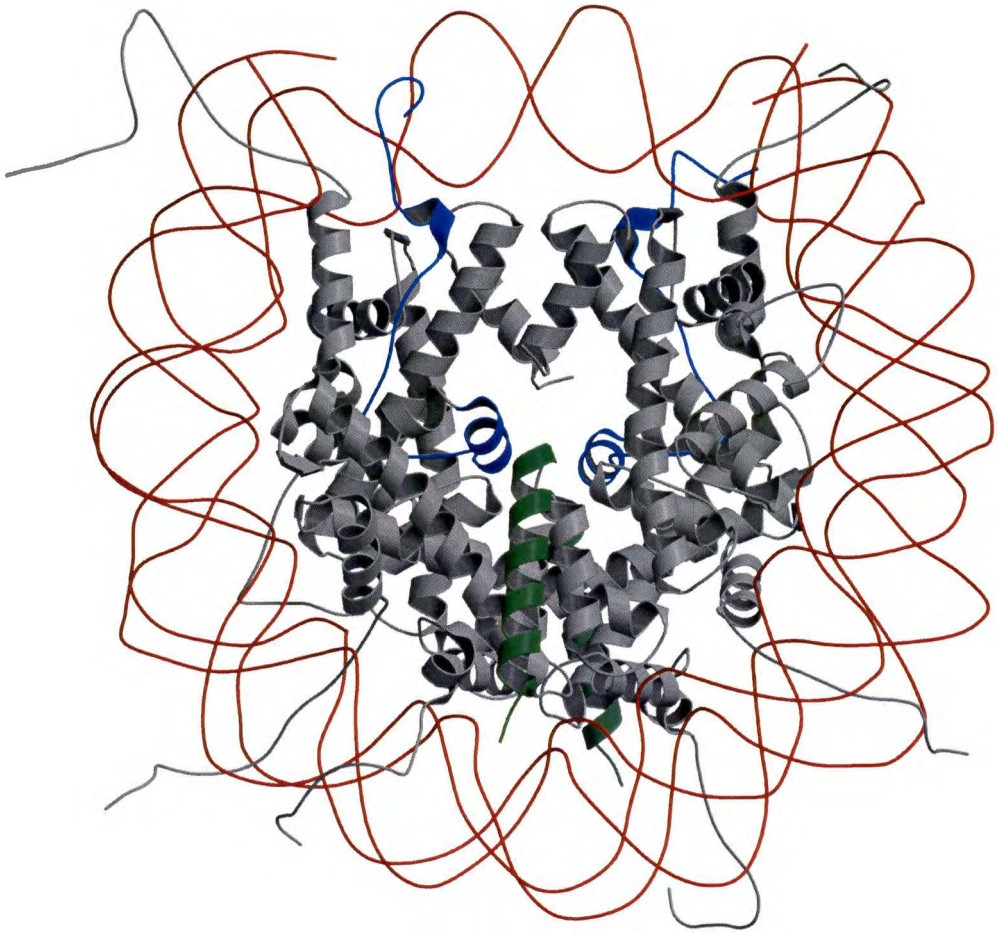

Figure 3.22: The C-terminal domains of the dimer histone subunits are shown in color with the remainder of the protein shown as grey. Histone H2A C-terminal domains are presented in blue for clarity. The C-terminal accessory helix of H2B is shown in green.

octamer structure while the tail is visible in electron density maps of the NCP for 14 residues in 1H3 and for 28 residues in 2H3. An N-terminal accessory helix displays significant interactions in stabilizing the DNA termini. Perhaps the most significant interaction of the histone fold of H3 is the binding of 1H3 to 2H3 which forms the foundation of the nucleosome structure.

The N-terminal tails of H4 extend from a 3_{10} N-terminal accessory helix which carries the tail out from the protein face of the NCP. The two H4 chains display a striking asymmetry in the orientation of the N-terminal tail domains with respect to symmetry neighbors. The tail of 1H4 extends to contact the DNA between ventral mGI 2 and 3 (see later discussion of DNA binding for definition of mGI positions) with Arg 23 making direct contact with a DNA phosphate. The tail then proceeds outwards to make indifferent contacts with 1H2B of a symmetry neighbor. The N-terminal tail domain of 2H4 is involved in an important packing interaction with a different symmetry neighbor by binding to an acidic region on the dimer surface of the neighboring NCP in the crystal lattice. This region is of particular interest as the interaction between the NCP and its neighbor is mediated by a very well ordered cacodylate ion located near the N-terminus of the 3_{10} accessory helix and the C-terminus of the NH helix of H3.

3.9.5 DNA binding to the octamer

For the purposes of discussion of the structure of the α-satellite DNA palindrome in the NCP, nucleotides were numbered 1 to 146, 5' to 3', with nucleotide 1 being the 5' end of the strand on the side of the NCP here defined as the ventral surface. The 3' end at nucleotide 146 lies on the dorsal surface of the NCP. The complementary strand is numbered, 5' to 3', as nucleotides 147 to 292 with nucleotide 292 base paired to nucleotide 1. The DNA strand numbered 1 to 146 is also referred to here as the Watson strand and its complement as the Crick strand.

Binding of DNA to the histone surface is primarily mediated by the insertion of arginine side chains at the minor groove-in position, mGI. The construction of the histone octamer serves to place the mGIs in such a way as to form a left-handed superhelical ramp with the mGIs spaced more or less evenly along the ramp. The positions were determined by locating all arginines inserted into the minor groove and determining which phosphates on either side of the minor groove were bound by putative salt bridges to the guanidino group of the arginine. The base position of individual mGIs was determined by the identity of the base along the Watson strand of the DNA positioned over the guanidino group of the arginine defining the mGI. These results are tabulated in Table 3.9. Figure 3.23 presents the mapping and numbering of the mGIs along the length of the DNA in the NCP. Numbering of mGI positions begins on either side flanking the pseudodyad of the particle at base 73 and proceeds outward to the DNA termini, with the 5' end of the Watson strand

Table 3.9: Tabulation of minor groove-in, mGI, positions. Each mGI is identified by a number from 1 to 6 counting outward from the pseudo-dyad and with a letter, v for ventral or d for dorsal. Each mGI is also identified by the arginine residue inserted into the minor groove of the DNA and by the type of binding motif, α or β as discussed in the text. The DNA phosphates listed are those positioned closest to the guanidino group of the inserted arginine residue and the base position is the base located above the inserted guanidino group away from the protein surface.

mGI	Arginine	type	DNA phosphates	base position
v 6	R77 1H2A	β	278/19	Ade 17
v 5	R33 1H2B	α	268/29	Ade 27
v 4	R42 1H2A	β	258/39	Thy 37
v 3	R83 1H3	β	246/51	Cyt 49
v 2	R63 1H3	α	237/60	Gua 58
v 1	R45 1H4	β	228/70	Ade 67/Gua 68
d 1	R45 2H4	β	217/81	Gua 78/Cyt 79y
d 2	R63 2H3	α	207/90	Cyt 88
d 3	R83 2H3	β	198-197/200	Gua 97/Gua 98
d 4	R42 2H2A	β	186/112	Ade 109/Ade 110
d 5	R33 2H2B	α	175/122	Thy 120
d 6	R77 2H2A	β	165/132	Thy 130

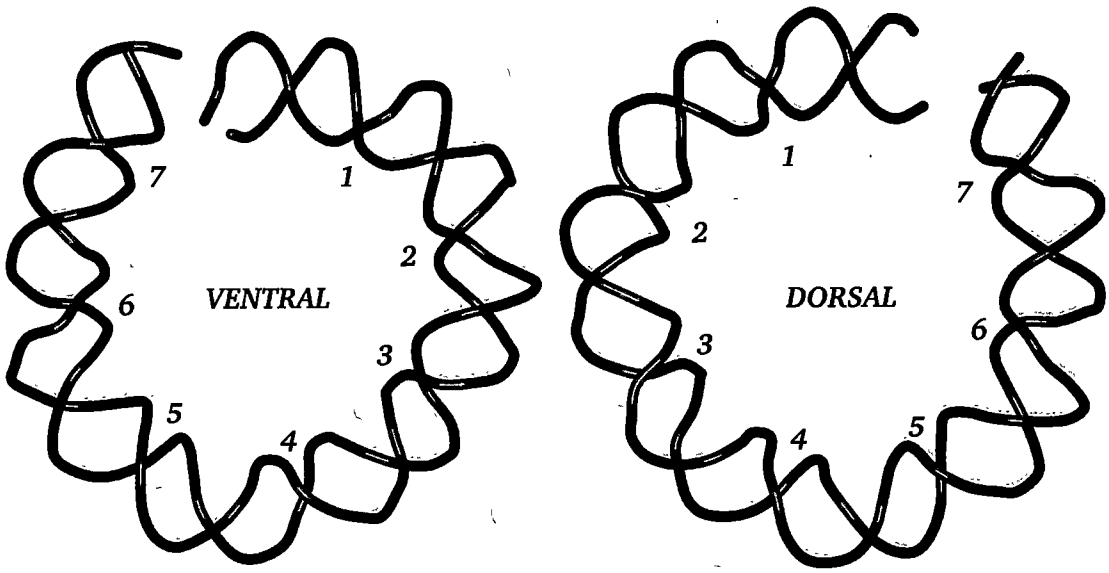

Figure 3.23: Mapping of minor groove-in, mGI, positions along the superhelical path of DNA in the nucleosome core particle for both ventral and dorsal halves with the zero position at the pseudodyad and numbering outward toward the DNA termini. Positions 1-6 correspond to mGI sites with an arginine inserted into the DNA minor groove.

defined as the plus side and the 3' end of the Watson strand defined as the minus side. Temperature factors for the phosphate atoms of the DNA were plotted as shown in Figure 3.24. The plot clearly shows the relationship between the temperature factors of DNA phosphorus atoms and the position of the phosphate moiety in relation to the surface of the histone core. The lowest temperature factors correspond to mGIs. The relationship of the temperature factor for DNA atoms to position is shown through color coding of atoms by temperature factor in Figure 3.25.

The mGIs represent binding motifs which are repeated regularly around the superhelical ramp of the octamer core. An arginine residue inserted into the minor groove appears as the central feature of the structure, that includes residues from 2 protein

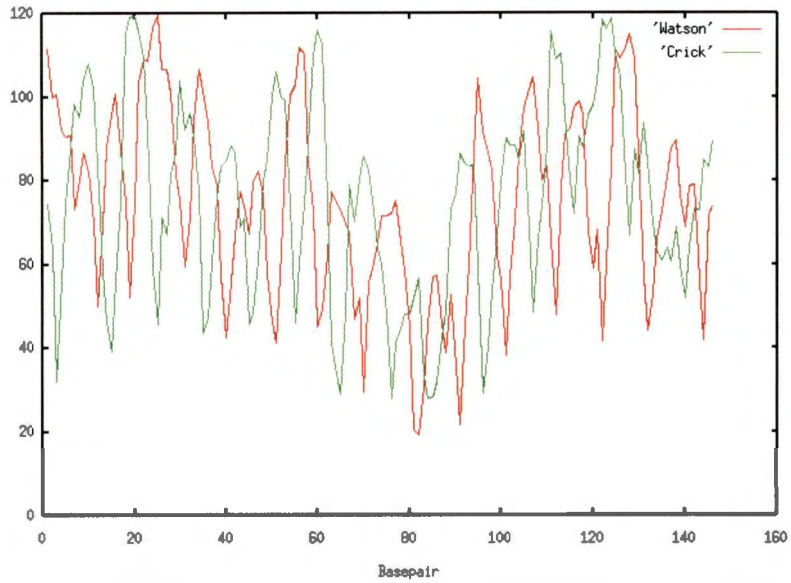

Figure 3.24: Plot of temperature-factors refined for phosphorus atoms along the backbone of the DNA in the nucleosome core particle structure. The identity of the strands is as discussed in the text.

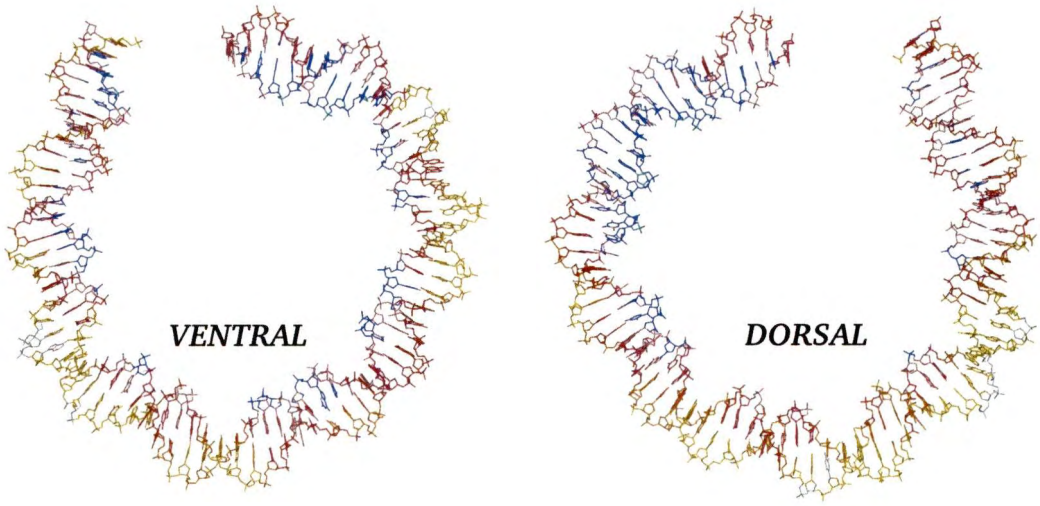

Figure 3.25: Temperature factors for DNA atoms are represented by color, from blue for lowest temperature factor to white for highest temperature factor. The DNA is broken into ventral and dorsal halves at the dyad.

subunits. Two basic designs are found and are here designated as α and β forms of the mGI.

αmGI DNA binding motif

The α form is constructed from a pairing of the N-termini of 2 central histone fold α helices. This binding motif was termed the paired-ends-of-helices, PEH, motif (Arents *et al.*, 1991) which interacts with the phosphate backbone through the dipoles produced in the helices. The αmGI binding is found at 2 and 5. The binding at 5 is accompanied by the insertion of Arg 33 of H2B into the minor groove. This position differs dramatically from all other mGIs in that the inserted arginine occurs on an N-terminal tail at a position not visible in the octamer structure (see Figure 3.26). Possible motion of the tail of H2B is restrained by the interaction of Arg 32 from the NH helix of H2A hydrogen bonding to Glu 35 of H2B. The NH helix of H2B is directed toward the phosphate group of Gua 268 at ventral mGI 5 and Gua 122 at dorsal mGI 5. The main chain amides of Ile 39 and Ser 38 of H2B contact the phosphate through water bridges, with additional contacts from the amides of Ser 36. The phosphate also is directly hydrogen bonded to Arg 29 from the NH helix of H2A. The NH helix of H2A is not directed toward the DNA phosphates but is accompanied by an N-terminal accessory helix, which interacts with the phosphate of Gua 31 at ventral mGI 5 and with Gua 177 at dorsal mGI 5. The interactions include direct hydrogen bonding to the main chain amide nitrogen of Arg 17 and a

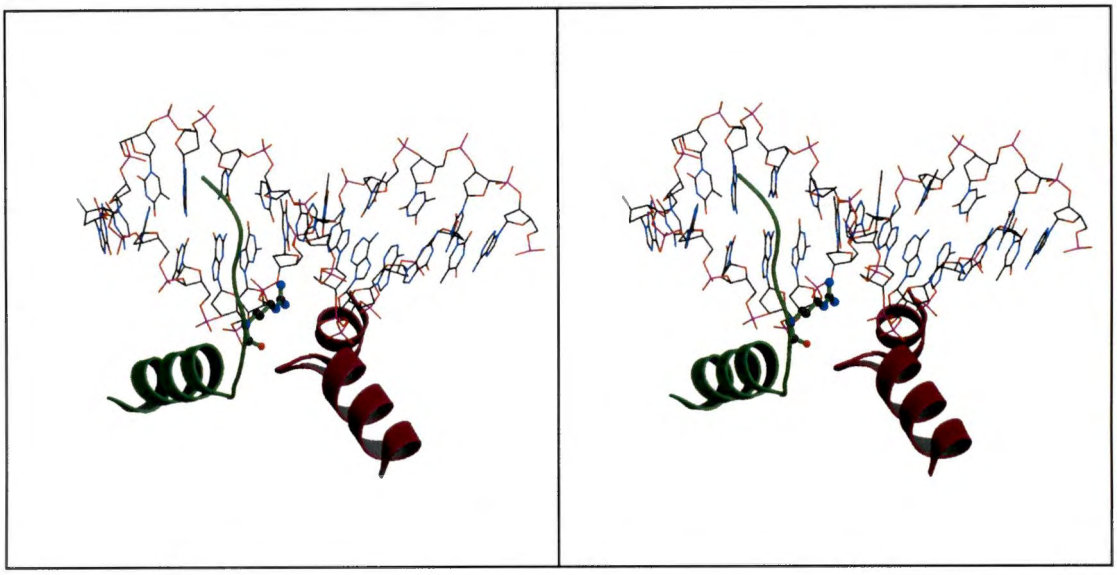

Figure 3.26: Stereo view of the structure of the αmGI binding site at dorsal mGI 5. The inserted arginine, Arg 33 of H2B, is shown in ball-and-stick rendering. The NH helix of H2A is not directed toward a DNA phosphate but is accompanied by an N-terminal accessory helix which interacts directly with a DNA phosphate outside of the minor groove. H2B is shown in green and H2A in lavender.

water bridge from the amide nitrogen of Ser 18 of the H2A N-terminal accessory helix. The N-terminal tail of 1H2B also passes between the DNA gyres and inserts Arg 29 into the minor groove between the phosphates of Thy 104 and Gua 192. Interestingly, the 2H2A tail also passes between the DNA gyres but does not insert into the minor groove between phosphates of Gua 249 and Thy 48 as might have been predicted from symmetry. Instead, it is deflected to place the guanidino group between the DNA gyres in proximity to phosphates of Thy 48 and Gua 125.

The αmGI binding sites at 2 are defined by the positions of Arg 63 of H3. A generalized view of the αmGI binding motif at this position is shown in Figure 3.27. The guanidino group is restrained from deeply inserting into the minor groove by Thr 30

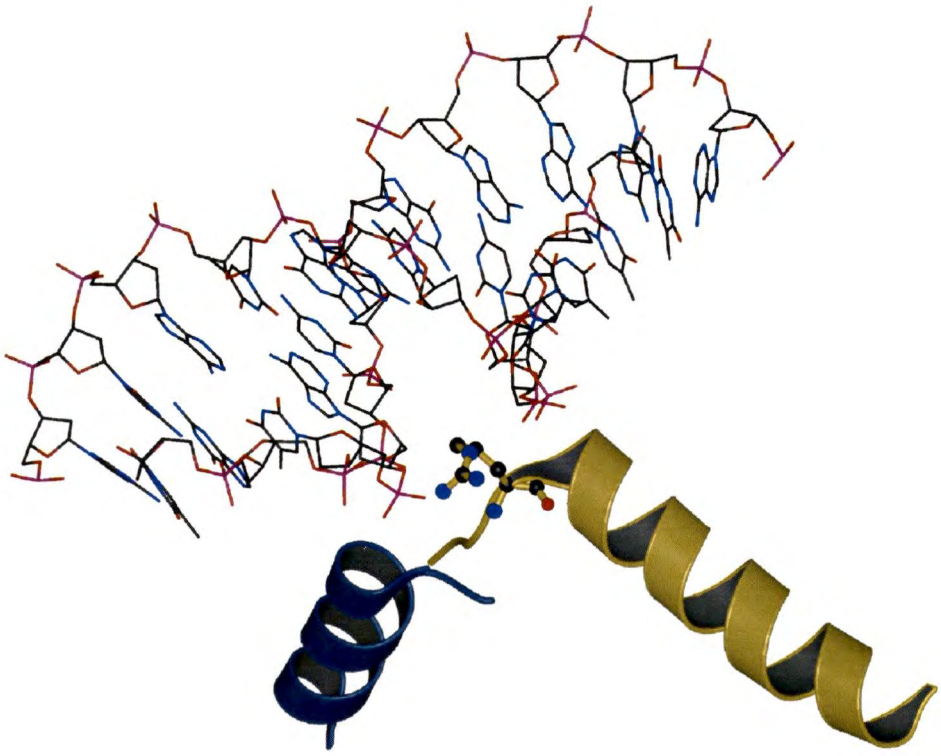

Figure 3.27: The αmGI motif found at mGI 2 is formed from the NH helices of H3 (yellow) and H4 (blue). The inserted arginine, Arg 63 of H3, is shown in ball and stick rendering. The N-termini of the paired helices are directed at phosphate groups on the DNA backbone.

of H4. The inserted arginine is positioned on a loop over a hydrophobic pocket. The NH helices of H3 and H4 are both involved in the formation of the binding site. At dorsal mGI 2 the N-terminus of the NH helix of H3 is directed toward the phosphate group of Thy 238 at dorsal mGI 2 and Thy 91 at ventral mGI 2 and main chain amide nitrogens of Leu 65 and Lys 64 are hydrogen bonded to the phosphate. The NH helix of H4 is directed toward the phosphate of Ade 61 at dorsal mGI 2 and Thy 208 at ventral mGI 2 but is displaced from the DNA backbone and does not make direct contacts with it although a water bridge is seen from Lys 31 of 2H4 to the phosphate at dorsal mGI 2. An unusual feature is the hydrophobic interaction in the major groove between Leu 65 of H3 and a thymine. In the octamer structure, Leu 65 of H3 is exposed to solvent and reinforced by the adjacent Pro 66. The position of this distinctive hydrophobic interaction in the major groove places it in the center of a run of four thymine residues which is intriguing in light of the importance of polyA tracts in the DNA sequence dependent phasing of nucleosomes.

βmGI DNA binding motif

The βmGI binding motif is found at mGI 6, mGI 4, mGI 3, and mGI 1. In general terms, the βmGI binding site is constructed from the CL loop of H3 and the NL loop of H4 or the CL loop of H2A and the NL loop of H2B. The loops are joined in a short, parallel β-bridge structure that forms a platform over which the phosphate backbones along the DNA minor groove are positioned. In most cases, the guanidino group of

the inserted arginine is restrained by hydrogen bonding to a threonine on the opposite loop. It may be that this interaction is designed to position the guanidino group in the DNA minor groove. However, the interaction of the guanidino group with a threonine is not invariant. The hydroxyl-carrying or acidic groups flanking the arginine may also serve to position the arginine side chain on the surface of the histone octamer in the absence of DNA. An examination of the binding pattern of the DNA at the mGIs (see Table 3.9) revealed that the phosphates directly bound to the arginine were separated by 5 bp or about one half of a helical turn. When one phosphate contacts an arginine at a mGI site, the next phosphate bound will be the one closest to the site on the other strand which is one half of a helical turn away. This suggests that the binding of the DNA takes place by a pinching of the minor groove with concomitant opening of the major groove. This may provide an alternative explanation for the significance of polyA tracts adjacent to an mGI as the polyA tract serves to narrow the minor groove in a 5' to 3' direction. Thus DNA binding to the octamer occurs at the mGI by the interaction of the arginine with phosphates on opposite strands and separated by one half of a helical turn. It would appear that the function of the remainder of the octamer is to place the arginines at appropriate positions. The propensity for the interaction of the arginine with tetrahedral ions such as phosphate or sulfate is shown in the octamer structure, where two sulfate ions are bound at Arg 45 of H4 in approximately the same positions as the phosphates at the same arginine at dorsal mGI 1 (see Figure 3.28). This is also seen in the structure of the

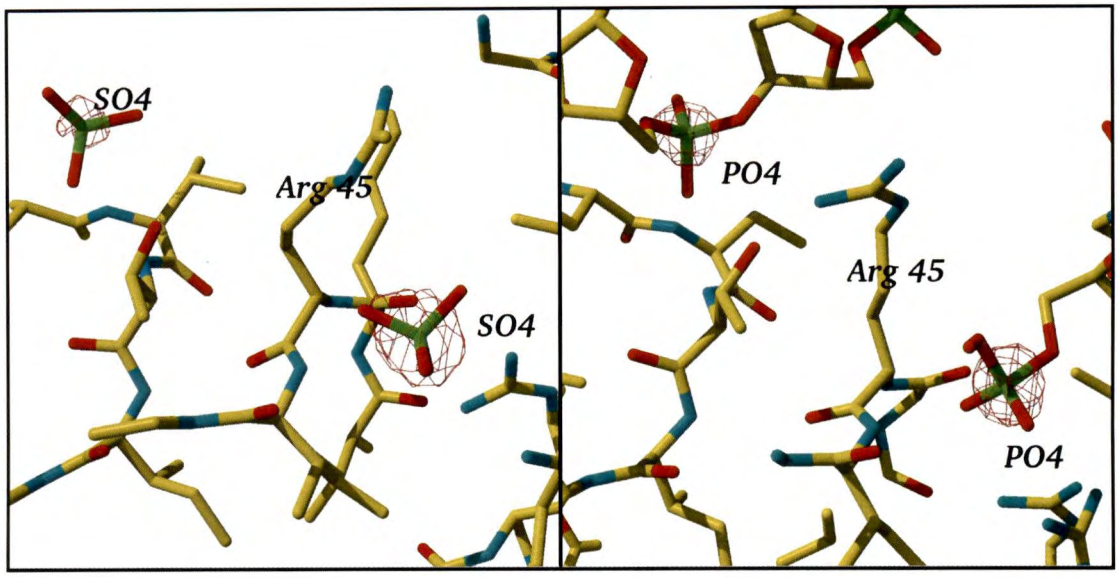

Figure 3.28: A comparison of the environment of Arg 45 of histone H4 in the octamer structure and at the dorsal mGI 1 of the nucleosome core particle. In the octamer structure, on the left hand panel, sulfate ions are bound in approximately the same positions as the DNA phosphates as shown in the right hand panel. Electron density is shown at 5 times the σ level of the $2F_o$-F_c map.

archael histones HMfA and HMfB (PDB ID code 1B67), where a sulfate ion is bound in approximately the same position.

The lowest temperature factors seen on the DNA backbone in Figure 3.24 occur adjacent to the nucleosomal DNA dyad center at a point at which the DNA minor groove contacts dorsal mGI 1. Examination of the structural details of the binding site reveal that the guanidino group of Arg 45 on the N-terminal loop of the histone fold of 2H4 is inserted into the minor groove and approximately equidistant from the phosphates of Gua 217 and Gua 81. Figure 3.28 presents a comparison of the environment of Arg 45 of H4 in both the NCP and the octamer structures. Interestingly, the phosphate positions seen in the NCP are nearly the same as the positions of sulfate ions bound

in the octamer structure. The binding site is further stabilized through arginines and lysines from 2H4 (Arg 35, Lys 31, Lys 44, Arg 39), Arg 116 from 2H3, and Lys 115 from 2H2A contributing to ionic stabilization of the DNA backbone. In addition, in the NCP, Arg 36 of H4 hydrogen bonds directly with the phosphate of Ade 207. The guanidino group of Arg 45 is restrained in both the octamer and NCP structures by hydrogen bonding with Thr 118 of 2H3. In the NCP structure, this interaction constrains the insertion of the guanidino group into the minor groove. In the octamer, the same interaction of Arg 45 with Thr 118 appears to restrain the motion of the guanidino group, which extends above the protein surface much like a grappling hook to bind DNA to the octamer.

At mGI 3, the interaction of the N-terminus of the mH helix of H3 with DNA phosphates is carried through by direct hydrogen bonding of the main chain amide nitrogen of Ser 86 to the phosphate and by water bridges between the phosphate oxygen and the amide nitrogen of Ser 87 as well as the hydroxyl group of Ser 87 (see Figure 3.29). At ventral mGI 3, the side chain of Gln 85 of H3 is also water bridged to the same phosphate. At dorsal mGI 3, Gln 85 reaches across the minor groove to form a water bridge with the phosphate of Ade 102 while the side chain is restrained by interactions with the C-terminus of the CH helix of 2H4. These interactions are formed through hydrogen bonding to the amide nitrogen of Ala 83 and by a water bridge to the amide nitrogen of Met 84 of 2H4. The phosphate group bound at the N-terminus of the mH helix of H3 at ventral mGI 3 is Cyt 50 and at dorsal mGI 3 is Cyt 196; these are

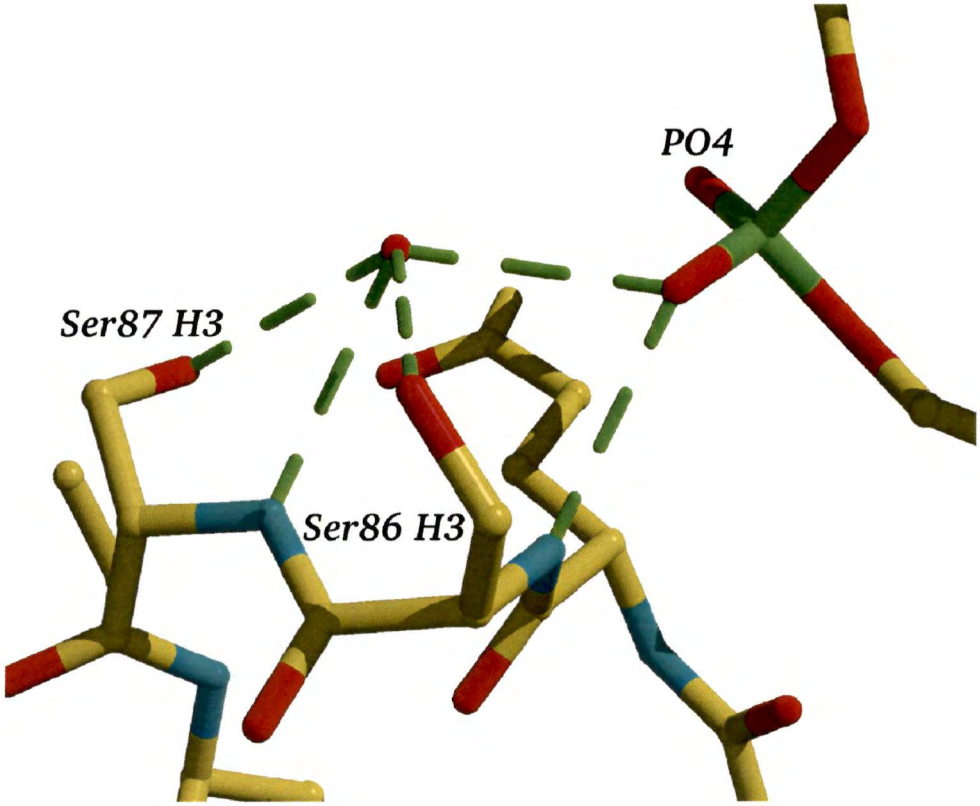

Figure 3.29: Direct binding at the N-terminus of the mH helix of H3 to DNA at mGI 3 is made through induced dipole of the α-helix and by direct hydrogen bonding of main chain amides as well as a water bridge to DNA phosphate.

identical positions by palindrome symmetry of the DNA. The inserted arginine, Arg 83 of H3, does not make any apparent bonds at either ventral mGI 3 or dorsal mGI 3 and does not interact with Thr 80 of H4 on the opposite side of the β-bridge. Thr 80 of H4 is hydrogen bonded to a DNA phosphate through the amide nitrogen and the hydroxyl group. A hydrogen bond to the same phosphate is also formed by the adjacent amide nitrogen of Lys 79 of H4. This interaction binds the phosphate of Cyt 247 at ventral mGI 3 and Cyt 101 at dorsal mGI 3 which are also identical by palindrome symmetry of the DNA sequence. Stabilizing interactions on both sides of the minor groove are provided by hydrogen bonding of side chains of Arg 72 of H3 and Lys 79 of H4 to DNA phosphates. At ventral mGI 3, Arg 78 of H4 can be seen to have two conformations, one of which involves hydrogen bonding to the phosphate group of Ade 248 and a water bridge to the side chain of Asn 84 of 1H2B. The second conformation shows the guanidino group of Arg 78 directly hydrogen bonded to Asp 85 of 1H4. This second conformation is the only one seen at dorsal mGI 3.

A chloride ion bound in a pocket formed by main chain amide nitrogens stabilizes the C-terminus of the β-bridge formed between the NL loop of H2A and the CL loop of H2B at mGI 4 (see Figure 3.30). The inserted arginine, Arg 42 of 1H2A, makes a hydrogen bond contact with the O4' of Ade 257 at ventral mGI 4, while at dorsal mGI 4 the guanidino group of Arg 42 is bound to a water. Thr 88 of H2B does not interact with the guanidino group of the inserted arginine. The main chain amides of Ser 87 and Thr 88 of H2B hydrogen bond with the phosphate of Gua 40 at ventral

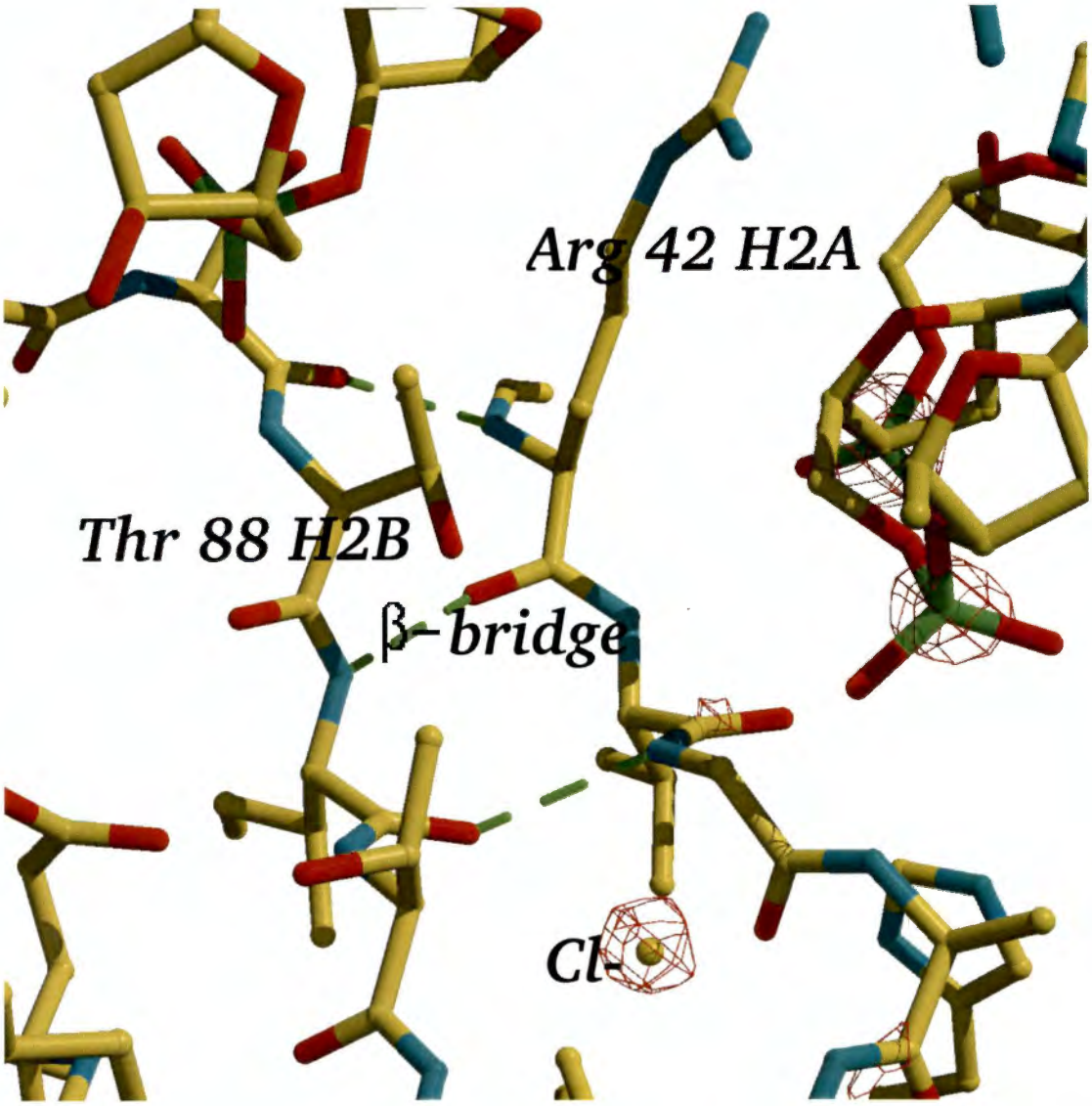

Figure 3.30: The structure of the β-bridge at mGI 4 is stabilized by the binding of a chloride ion in a pocket formed by main chain amide nitrogens. The guanidino group of Arg 42 of H2A is inserted into the DNA minor groove. Electron density is shown at 5 times the σ level of the $2F_o$-F_c map.

mGI 4 and of Gua 185 at dorsal mGI 4. The hydroxyl of Ser 87 is also positioned to make hydrogen bond contact with a phosphate oxygen but only clearly does so at dorsal mGI 4. Further stabilizing interactions are provided through direct hydrogen bonding of Arg 35 of H2A and Arg 86 of H2B to phosphates on opposite sides of the minor groove.

The βmGI binding site at 6 differs somewhat from the other βmGIs. The inserted arginine, Arg 77, is placed on the CL loop of H2A and the threonine which might be expected to hydrogen bond to it, Thr 52 of H2B, points inward toward the histone core. The inserted arginine is bounded by Thr 72 of H2A on one side and Asp 51 of H2B on the other. Main chain amide nitrogens of Thr 76 and Arg 77 of H2A hydrogen bond directly to a DNA phosphate, Cyt 132 at dorsal mGI 6 and Ade 19 at ventral mGI 6. Arg 77 of H2A is inserted deeply into the minor groove and hydrogen bonds to the O4' on the ribose moiety of Gua 131 (see Figure 3.31).

An interesting feature of the DNA in the NCP structure is the path taken by the double helix between neighboring NCPs. As seen in Figure 3.32, the 5' ends are abutted in a quite unnatural manner while the hydrophobic stacking of base rings is easily maintained moving from one NCP to another. The DNA used in this study was not treated to remove the 5' phosphates.

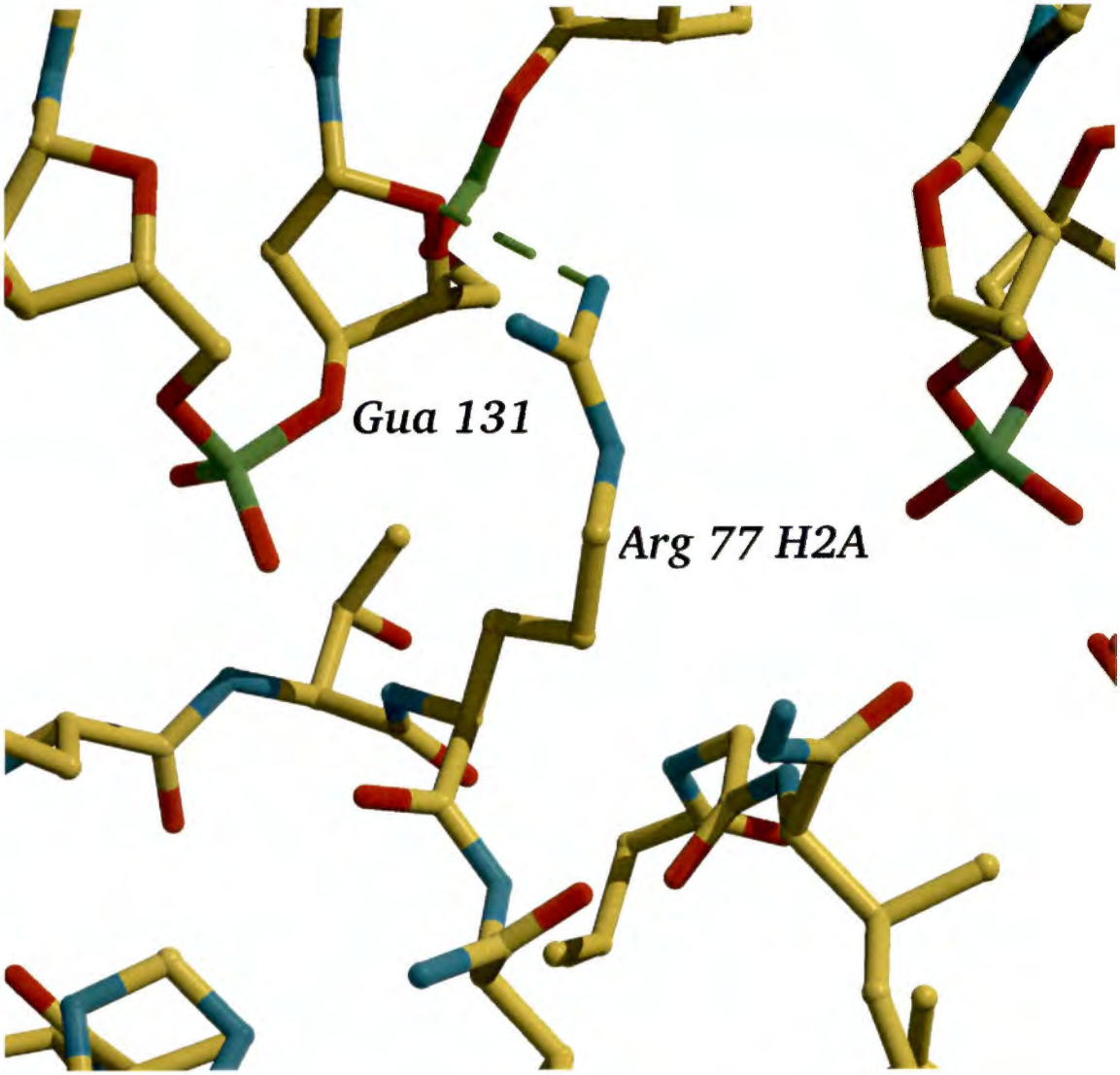

Figure 3.31: The structure of dorsal mGI 6 has the inserted arginine, Arg 77 of H2A, hydrogen bonding to the O4' atom of Gua 131.

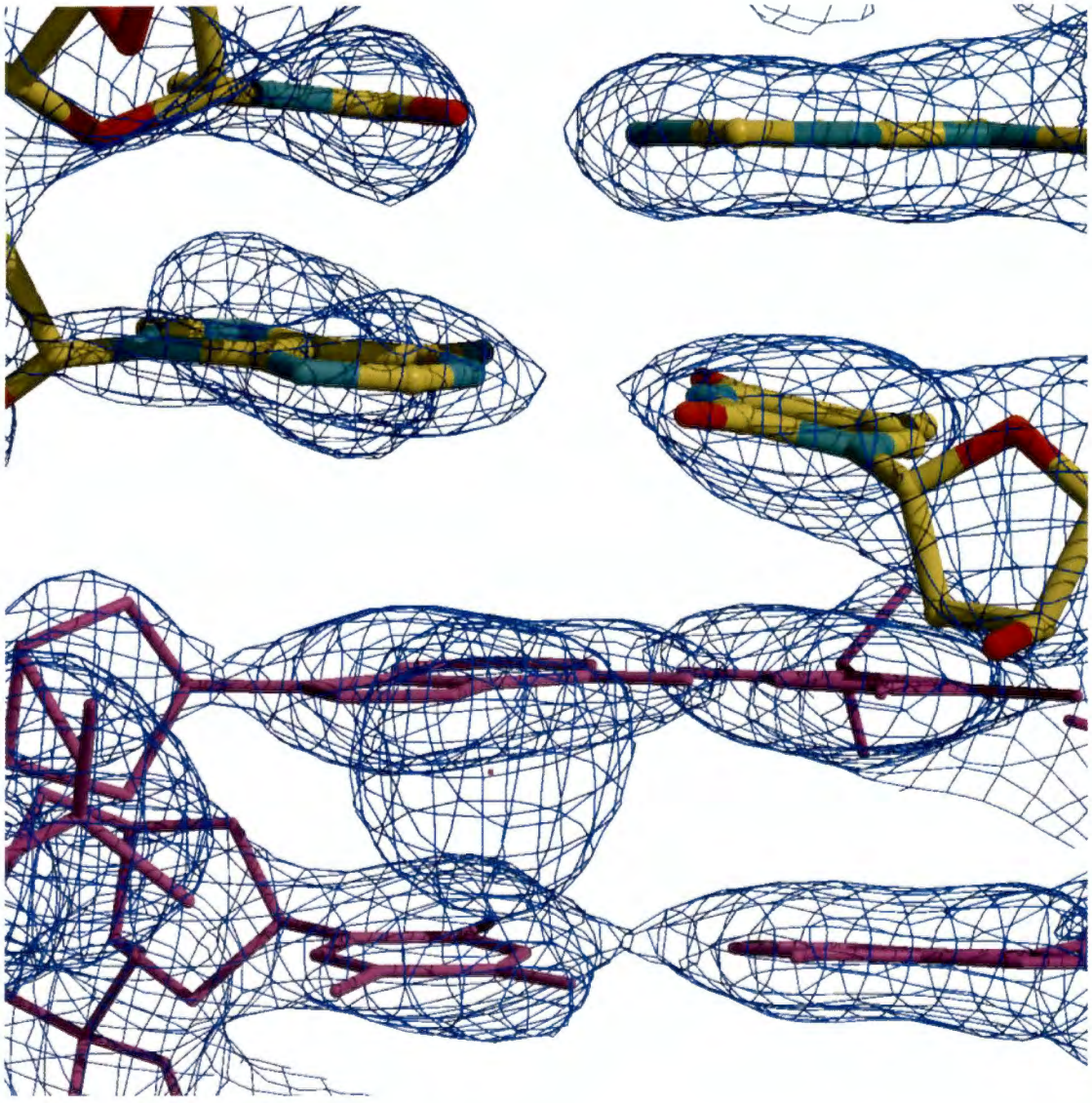

Figure 3.32: Base stacking is continuous between neighboring NCPs in the crystal lattice as the DNA exits one NCP and enters the next. Electron density is displayed at one times the σ level of the $2F_o$-F_c map.

3.9.6 Folding of the DNA into the nucleosome core particle

The role of arginines in the binding and folding of DNA into the NCP has long been known. Using NCPs prepared by micrococcal nuclease digestion of chicken erythrocyte chromatin, it was determined (Ichimura *et al.*, 1982) that 14 arginine residues were involved in binding to DNA and were released for chemical modification only above 0.6 M NaCl. Although the precise identity of the 14 arginine residues was not established, it is likely that they were the 14 arginines seen to insert into the minor groove of the DNA in the NCP structure, 12 of which have been used here to define the mGI positions. The guanidino group of arginine binds strongly through salt bridges to tetrahedral-type anions such as the phosphate group on the DNA backbone. In the octamer structure, sulfate ions are bound to the exposed side chains of several of the arginines at latent mGI sites in the approximate positions that would be occupied by phosphate groups when DNA is bound as seen in Figure 3.28. The arginine side chains at the latent mGI sites extend above the protein surface to provide anchor points for attachment of DNA to the histone octamer. It has long been known that the central approximate 100 bp of DNA in the NCP are more resistant to micrococcal nuclease than the termini (van Holde *et al.*, 1980). The strength of the binding is further suggested by hydroxyl radical footprinting of nucleosome cores at high salt and high temperature (Bashkin *et al.*, 1993), which demonstrated that translational and rotational positioning is maintained even at 0.8 M NaCl or at 75°C. Mechanisms for displacement of DNA from the histone core *in vitro* involve melting of the DNA

duplex or disruption of the binding by very high salt concentrations or by cationic molecules such as the protamines, which possess polyarginine segments.

The necessity of mutating the native α-satellite sequence to provide a synthetic *Eco*R I site may have contributed to the failure of the palindrome to phase with perfect symmetry by altering the helical parameters of the native sequence at the dyad. Energy minimizations of the central 22 base pairs using the DUPLEX program (Hingerty *et al.*, 1989) clearly demonstrated the tendency of the *Eco*R I recognition sequence to bend the fragment away from the histone surface with the bend centered at the dyad (unpublished results). We have earlier postulated (Uberbacher *et al.*, 1988) that the histones bind DNA cooperatively and that upon initial binding, the histones "communicate" to their symmetry-related counterparts which properties of the DNA to select. From an examination of the temperature factors of the phosphates along both strands of DNA in Figure 3.24, it is clear that the phosphodiester backbone is more stably positioned on the side of the dyad possessing the longer DNA strand. The lowest temperature factors are those for the phosphates bound at dorsal mGI 1. The DNA must bind first on one side of the histone core in a conformation which provides the lowest free energy of binding and then be accommodated on the other side by binding in a slightly less favorable mode. However, since the histone binding sites are not strictly uniform in nature, the binding is not uniformly misdirected and the base positions at ventral mGI 5 and 6 are as expected by palindrome sequence symmetry. The DNA between ventral mGI 2 and 3 is underwound as compared to

its symmetry-related counterpart (see Figure 3.33), as is the DNA between ventral mGI 4 and 5. This is in partial disagreement with the results of Luger, *et al.* (Luger *et al.*, 1997) who found only a difference between the DNA segments between mGI 2 and 3. Interestingly, these 2 regions are flanked by α mGIs on one side (mGI 2 and mGI 5). The mixture of α and βmGI motifs may account for some of the plasticity in DNA binding seen here while still providing for precise phasing.

A further very important consequence of the asymmetry in the DNA is that the non-equivalence of the two sides of the NCP as it packs into the crystal lattice provides an explanation to the question of how the NCP packed. The original intention of crystallizing NCPs containing a DNA palindrome (Harp *et al.*, 1996) was to produce a two-fold symmetric particle for which the dorsal-ventral asymmetry did not exist. Because of the shape of the NCP, it can, like a coin, land in one of two orientations resulting in two-fold disorder in the crystal if the two orientations are not equivalent. The single orientation was intriguing in that it meant that some other factor was involved which allowed the NCPs to pack in only one orientation. If we refer to the short side of the DNA as the ventral side and the long side of the DNA as the dorsal side, an analysis of crystal packing shows that the DNA exiting the NCP is abutted to DNA exiting a symmetry neighbor. The ventral side of the DNA of the nucleosome core particle abuts the dorsal side of the DNA of a symmetry neighbor and the dorsal side of the DNA abuts the ventral side of the DNA of another symmetry neighbor in the same plane. This can be clearly seen in Figure 3.34. If the NCP is shown

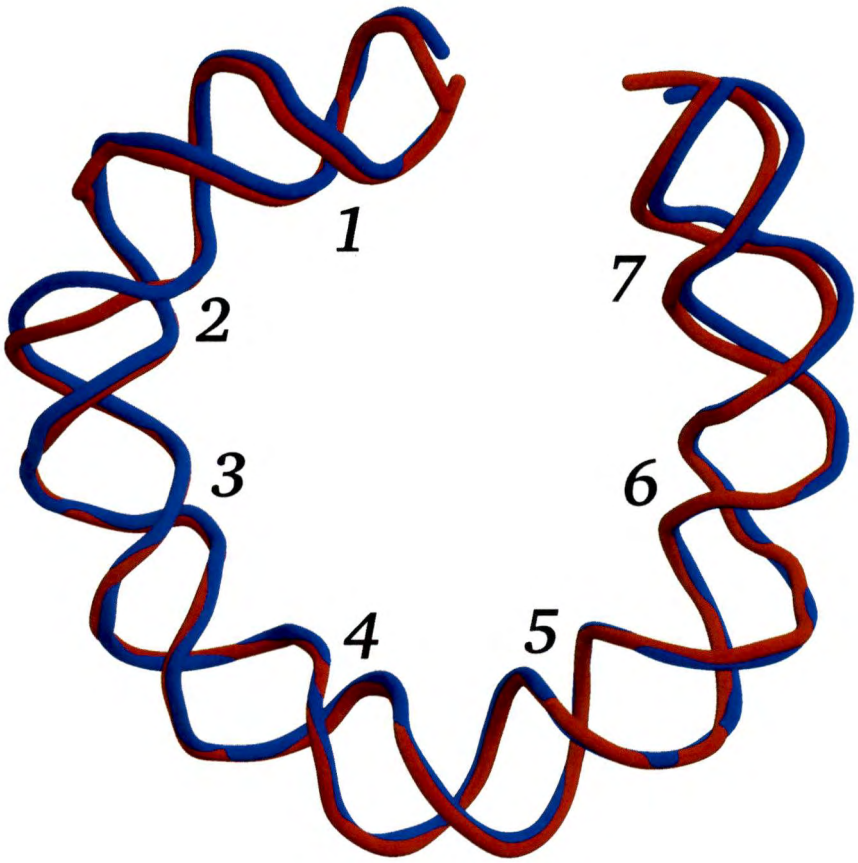

Figure 3.33: Overlapping of the phosphodiester backbone trace of the 2 halves of the DNA demonstrates binding asymmetry. The minor groove-in sites are numbered as in the previous figure. The 73 base pair dorsal half is shown in blue and the 72 base pair ventral half in red. Asymmetry in the path of the DNA backbone are revealed between positions 2 and 3 and, to a lesser degree, between positions 5 and 6.

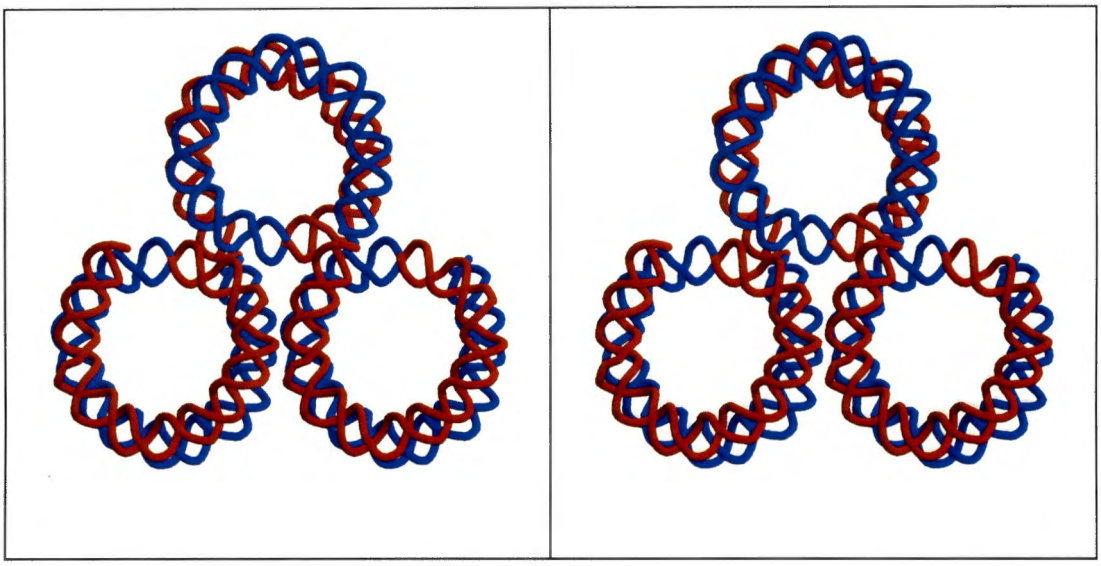

Figure 3.34: Stereo packing diagram showing the continuous path of DNA from the NCP to symmetry neighbors. DNA from three NCPs are shown with the 72 bp, ventral, half colored red and the 73 bp, dorsal, half in blue. This shows that the NCPs pack in only one orientation depending on the orientation of NCPs already situated in the crystal.

in a dorsal view, the two symmetry neighbors sharing the same continuous path of DNA will present their ventral side in Figure 3.34. Thus, between dyad centers of two such symmetry related NCPs there are 145 base pairs of DNA excluding the base pairs at either dyad. This fact means that the NCP does not have two equally likely choices of packing with either the ventral or dorsal surface down but can only pack in one orientation as determined by the orientation of the NCP to which it joins DNA termini. Unfortunately, the question of how the packing of NCPs in the crystal relates to the packaging of nucleosomes in higher order structure is unclear. Certainly, the incorrect bond polarity at the interface of the DNA termini between symmetry neighbors is an artifact. Also, the stacking of NCPs in the direction of the superhelical axis places the symmetry neighbors with alternating polarity with

respect to the dyad axis, that is, in a head-tail-head fashion which would not be expected from any model of the condensed (30 nm) fiber with the linkers disposed toward the center.

DNA helical parameters

An examination of the helical parameters between adjacent base pairs was conducted using the program rna (Babcock *et al.*, 1993). Folding of DNA onto the superhelical ramp provided by the histone octamer core results in overwinding of the duplex and an increase in the helical twist angle between neighboring base-pairs. The helical repeat of DNA in solution has been measured as 10.5 base pairs per turn of the double helix. The helical repeat of the DNA bound to the NCP reflects overwinding resulting from bending and is measured as approximately 9.9±1.4 base pairs per turn. The twist angle data were further broken down into segments delimited by mGI positions as shown in Table 3.9. To examine the pattern of winding and the symmetry relation of the winding parameter, a plot was made of base pairs per turn, calculated by dividing 360° by the twist angle in degrees (see Figure 3.35). Examination of the plot revealed that the primary asymmetry between mGIs 2 and 3 is reflected in the twist at these positions. The ventral segment was significantly overwound with respect to the symmetry related dorsal segment. The overall symmetry of the structure is reflected in the data but the resolution of the diffraction data and higher temperature factors for the DNA versus protein residues did not allow for precise placement of bases.

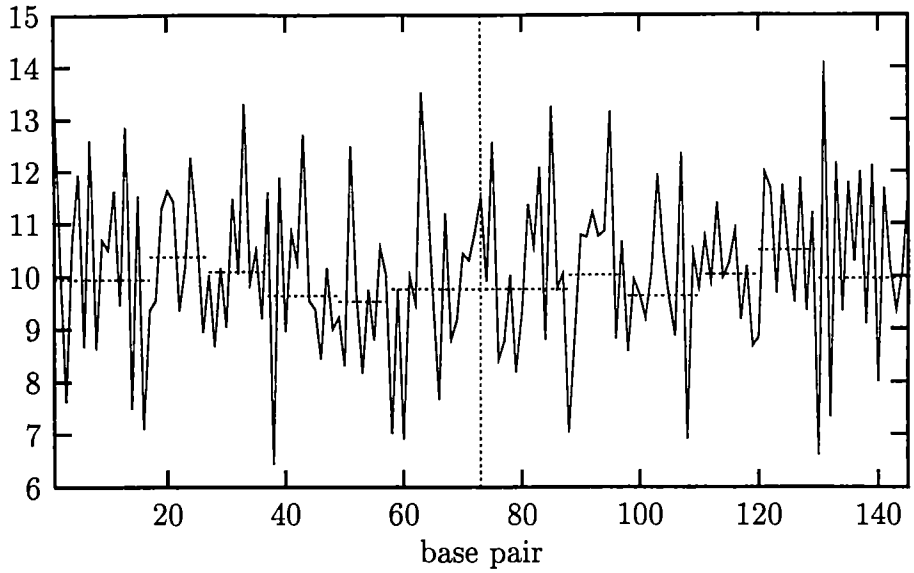

Figure 3.35: Plot of helical twist expressed as base pairs per turn over the length of the DNA. The average repeat length in base pairs per turn is shown as a dashed line covering the extent of the DNA between mGI positions. Data for the center of the DNA is combined between ventral mGI 2 and dorsal mGI 2.

An examination of the structure of the DNA at the dyad was undertaken with the collaboration of Dr. Brian Hingerty to attempt to reproduce the structure of the DNA on the nucleosome using an energy minimization from the sequence data alone (Hingerty *et al.*, 1989). Unconstrained runs of the sequence of the central 22 base pairs confirmed that the *Eco*R I recognition sequence placed at the center actually tended to bend away from the histone octamer surface (data not shown) however, the precise parameters of the structure did not match the experimentally determined model. This was especially evident in the width of the minor groove, which was significantly narrowed in the theoretical structure. Runs of the sequence were also performed using distance constraints measured from the experimental structure. These results suffered from the same narrowing of the minor groove. Additional modifications of

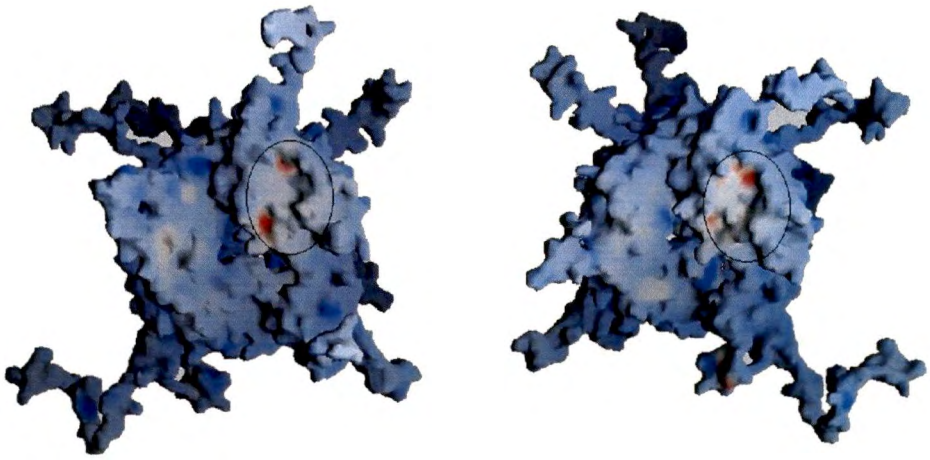

Figure 3.36: A GRASP electrostatic surface potential map of the histone octamer core in both ventral, on the left, and dorsal, on the right, views. An acidic patch on the dimer face in both views is the most striking feature of the map. The acidic patch is indicated by an ellipse in both views.

the theoretical parameters affecting the electrostatic force field were made to bring the theoretical results in line with the experimental parameters using simple models derived from structures in the PDB database and presented at the conference Computational Biology 2000 in Nice, France (Hingerty *et al.*, 2000). Work on modeling of curved DNA in the nucleosome is continuing.

3.9.7 Stacking interactions between histone faces

A striking feature of the surface of the histone face of the NCP is the presence of a number of acidic residues on the dimer face (see Figure 3.36). These residues include Glu 56, Glu 61, Glu 64, Asp 90, Glu 91, and Glu 92 of H2A and Glu 110 of H2B

on the ventral face of the NCP. The N-terminal tail of 2H4 from one NCP lies along the acidic patch forming numerous hydrogen bonds and water bridges between the basic residues on the H4 tail from the dorsal surface of the neighboring NCP and the acidic residues on the opposing dimer surface. The H4 tail is well visualized to Lys 8. Significantly, that there is no evidence of acetylation of lysines on the H4 tail although the histones are native chicken erythrocyte octamers. The cacodylate ion lies in a pocket formed on one side by Asp 77, Gln 76, and Glu 73 on the NH helix and NL loop of 2H3 as well as the main chain carbonyl of Leu 22 of 2H4 (see Figure 3.37 and Figure 3.38). The opposing NCP forms the other side of the pocket with the cacodylate ion interacting with main chain carbonyls of Gln 47 and Val 48 and the side chain of His 49 of 1H2B. This interaction appears to be responsible for the asymmetry seen in the dimer histones in Figure 3.13. The asymmetry is related to the interaction between neighboring NCPs along the acidic patch on the dimer surface (see Figure 3.39) and is due to the fact that only the ventral dimer face of one NCP and the dorsal tetramer face of the adjacent NCP are involved. This stacking interaction may be physiologically relevant, however, it seems unlikely due to the alternating orientation of the NCPs in the stack. This arrangement does not correspond to current models of higher order chromatin structure.

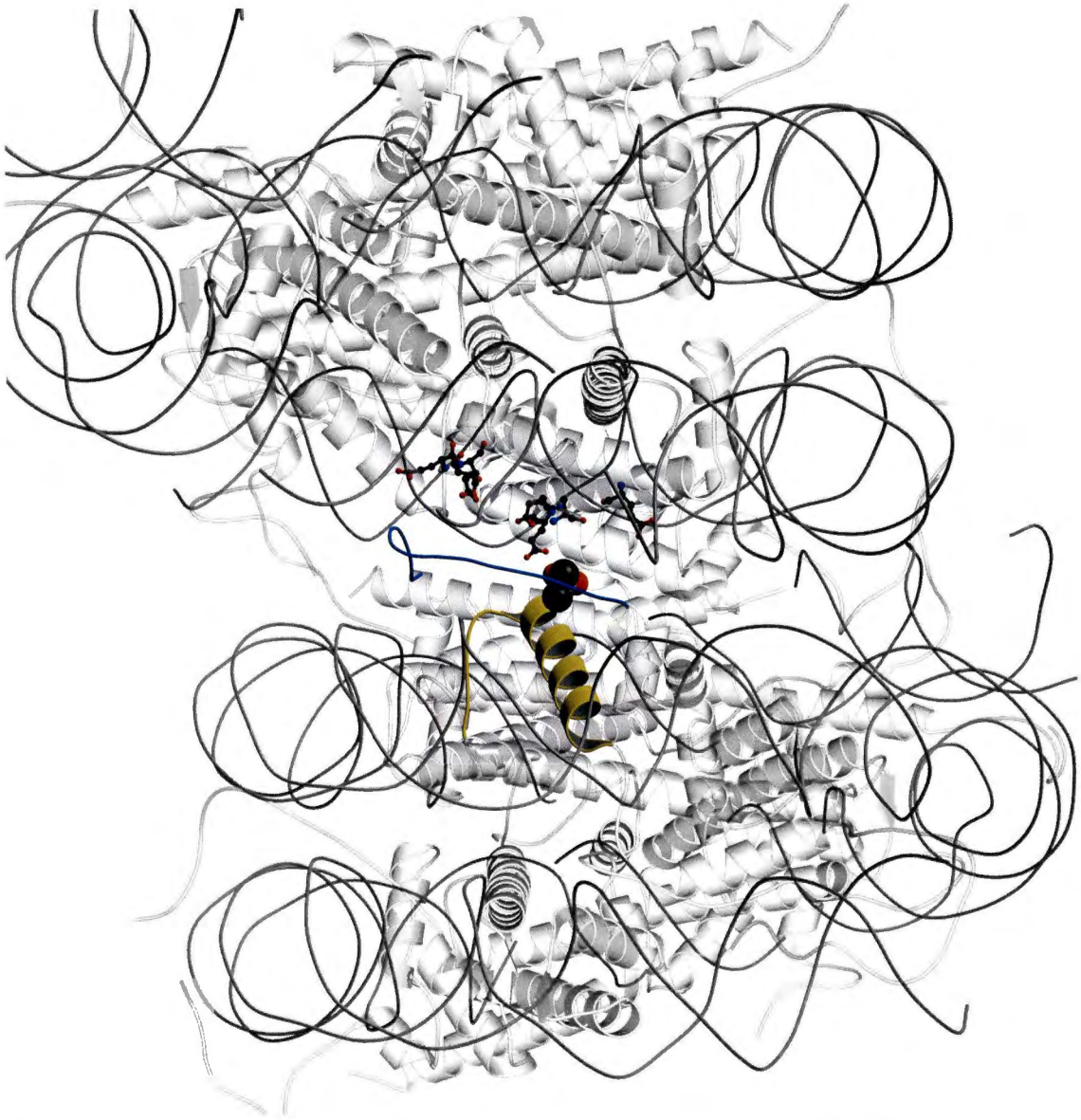

Figure 3.37: Stacking interactions between histone faces of neighboring NCPs in the crystal lattice. A central feature of the stacking interactions is a strongly bound cacodylate ion. The N-terminal tail of H4, shown in blue, and the NH helix and NL loop of H3 form a pocket containing the cacodylate ion on one particle. The neighboring particle interacts with the cacodylate through the NH helix and NL loop of H2B. The N-terminal tail of H4 is also involved in significant interactions with a patch of acidic residues, rendered as ball-and-stick, on the dimer face of the neighbor NCP. Note that the orientation of the dyad alternates between the two NCPs. Features of the interaction are identified and shown in detail in the next figure.

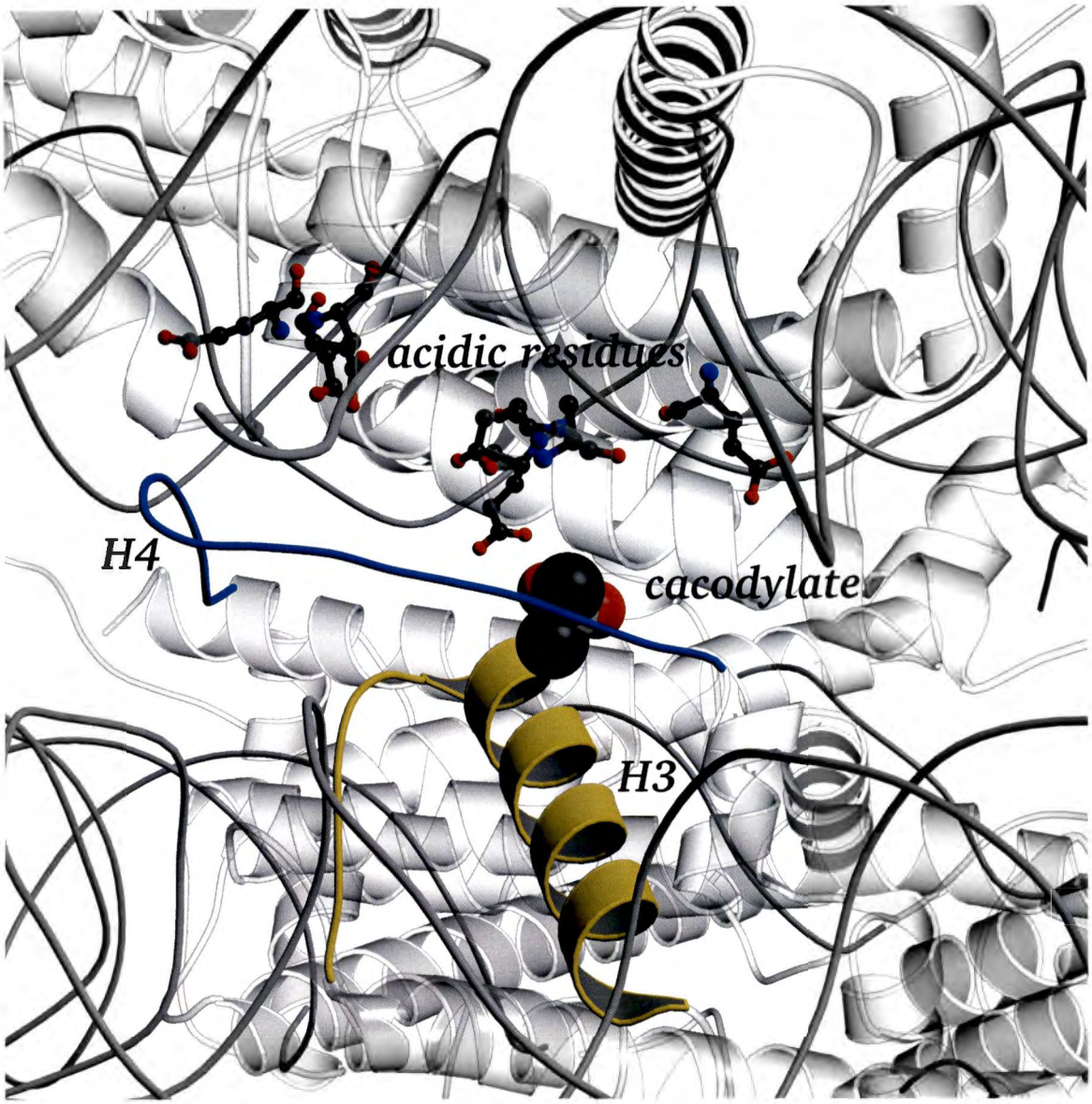

Figure 3.38: Enlarged view of the previous figure showing details of the stacking interactions between two neighboring NCPs involving the dorsal $(H3{\cdot}H4)_2$ tetramer face of one NCP and the ventral (H2A·H2B) dimer face of the adjacent NCP. The acidic residues forming the surface acidic patch on the dimer face are rendered as ball-and-stick. The H4 tail is in blue and the region of H3 involved in the interaction is in yellow. The cacodylate is shown in CPK rendering.

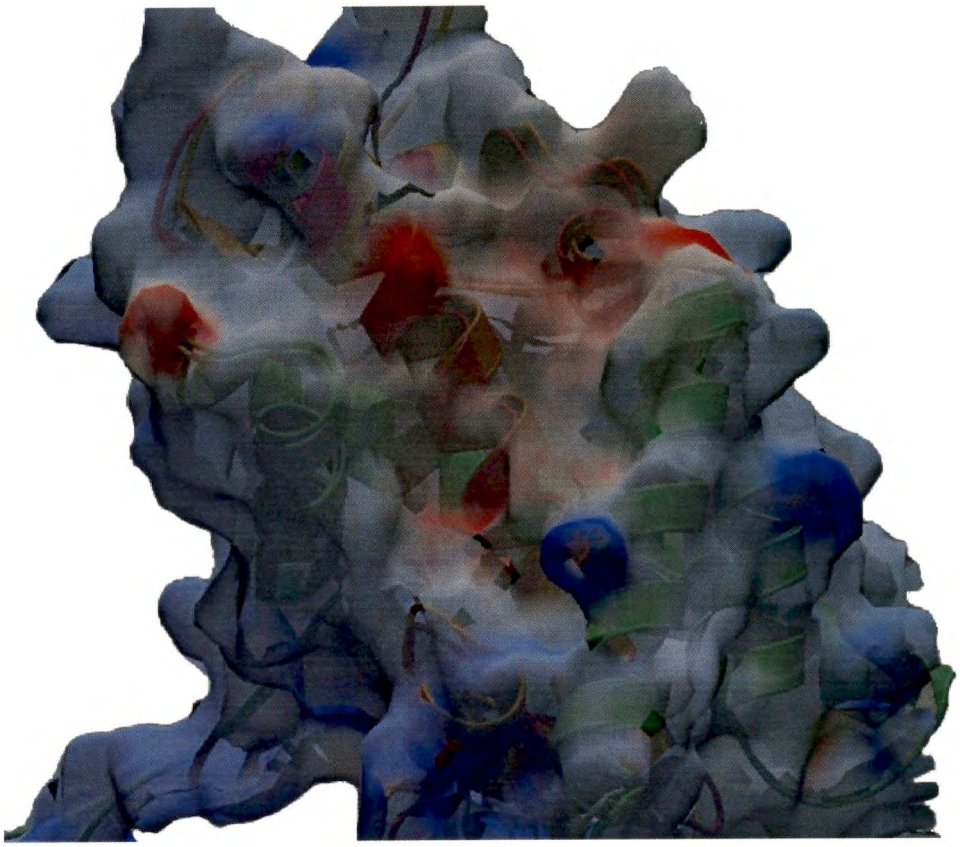

Figure 3.39: In this figure, a GRASP electrostatic potential surface has been been laid over a model of the dimers, which have been superimposed by least-squares fitting of the dorsal dimer onto the ventral dimer to demonstrate that the asymmetry of the histones is related to the stacking interaction involving the acidic region on the dimer face.

Chapter 4

Discussion

4.1 Microgravity crystallization: Was it worth it?

Microgravity crystallization was used to improve the quality of NCP crystals. Ground-based crystallization suffers from two factors, convection currents and sedimentation of growing crystals. Convection currents in ground-based crystallization work against the need for carefully controlled diffusion rates and the sedimentation of crystals eliminates the addition of molecules to crystal surfaces in contact with the bottom of the crystallization chamber. In the microgravity environment, nascent crystals do not sediment but remain suspended in the mother liquor. Addition of molecules to the growing crystal surface results in a depletion zone around the crystal. In the absence of convection currents, the depletion zone is maintained and movement of particles

across the depletion zone to add to the crystal is entirely diffusion driven. This results in a preference for monomers with higher diffusion coefficients and reduces the likelihood of adding aggregates and larger impurities to the crystal.

The primary objective for microgravity crystallization of the NCP derived from the growth habit of NCP crystals which resulted in roughly hexagonal rods. Addition of NCP to the crystal occurred more readily along the long axis of the crystal. The differential growth rate tended to produce hollow rods with a severely reduced cross-section for diffraction. The relatively high average temperature factor characteristic of NCP crystals produced low diffraction intensities for reflections, making crystal volume an important factor in the data collection strategy. Microgravity was imminently successful in reducing the occurence of hollow crystals. Experiments performed during the flight of the USML-2 mission were the first attempted for the NCP. Extensive launch delays resulted in poor results. However, one crystal was observed to be a completely solid and nearly perfect hexagonal rod. That crystal, seen in Figure 3.6, was the first of its kind to be observed. It was concluded that the crystal must have grown in microgravity. This provided the impetus to continue experiments on the space station Mir. Thereafter, microgravity experiments routinely returned solid crystals.

There has been much controversy concerning the value of microgravity for crystallization of biological macromolecules. Opponents have often cited the lack of novel results coming from the space shuttle and Mir space station experiment programs.

The American Society for Cell Biology published a position paper in 1998 (American Society for Cell Biology, 1998) calling for an end to funding of microgravity crystallization. The position statement claimed that purity and stability of macromolecules is of primary importance to success in crystallization experiments and that microgravity had produced no significant results. A recent acrimonious exchange in the pages of *Science* (Malakoff, 1999; Delucas, 1999; Laver, 1999) typified the differences of opinion. An unfortunate overrepresentation of the value of microgravity crystallization as the solution of the structure of an important flu drug led to destructive criticism which has severely tarnished the image of NASA's microgravity crystallization projects.

The results reported here provide a real example of the utility of microgravity for improvement of crystal perfection for the NCP. The 2.5 Å structure is based entirely on data collected from a relatively small crystal grown on the Space Shuttle as part of the USML-2 mission. The data, which were collected at beamline X-12C of the National Synchrotron Light Source at Brookhaven National Laboratory, have not been matched in quality and completeness by any subsequent ground-based crystal. Ground-based crystals have also been given the advantage of macromolecular crystal annealing, which was discovered later in the research program, as well as the advantage of the higher intensity of a third-generation synchrotron source. High quality ground-based crystals were taken to the IMCA beamline at the Advanced Photon Source at Argonne National Laboratory. Those data (not shown) were, again, similar

in anisotropy and resolution, to the ground-based data illustrated in Figure 3.8.

4.2 Macromolecular crystal annealing.

A study of the nature of MCA and its applicability to macromolecular crystals other than the NCP crystals was undertaken in the belief that the effect itself was of marked significance to both the current research project and to macromolecular crystallography in general. The ability of a macromolecular crystal to undergo a rewarming to ambient temperature from cryogenic conditions without loss of diffraction was not believed possible at the time and there were no published reports of such phenomena. At about the time that the first phase of the study of MCA was complete, a study was published (Sauer & Ceska, 1997) in which a simple comment was made that the crystals used could be taken out of the cold nitrogen gas stream, warmed to ambient temperature, and then flash-cooled a second time without loss of diffraction quality. No mention was made of improvement in diffraction quality. Two publications resulted from the work done on MCA. The first (Harp *et al.*, 1998) reported the first studies of the phenomenon. The second (Harp *et al.*, 1999) reported studies undertaken in response to the publication of a paper (Yeh & Hol, 1999) that described a variation of the technique based on one set of experiments using a single crystal system. The decision to divert time and resources from the studies of the NCP to the investigation of the phenomenon of MCA was supported by receipt of the Oxford Cryosystems

Cryocrystallography Award for the report of the work at the 1997 American Crystallographic Association Annual Meeting in St. Louis and a Technical Achievement Award from Lockheed Martin Energy Research Corporation in the following year.

4.2.1 The definition of success

> Definition of successful annealing: *The application of an annealing protocol without a detectable loss of diffraction quality.*

For the purposes of the study of MCA, the definition of a successful annealing is given as "the application of an annealing protocol without a detectable loss of diffraction quality." This definition became important as more crystals were examined and it became apparent that the effect of the annealing was proportional to the increase in mosaicity caused by the original flash-cooling. If the crystal did not exhibit an increase in mosaicity after the initial flash-cooling, macromolecular crystal annealing could still be applied to the crystal without damage to the diffraction quality of the crystal. It was important to take this fact into account when analyzing the results of these studies. The phenomenon also had some important benefits of a purely technical nature. The ability to anneal a crystal without damage to the diffraction quality provided a new tool to simplify handling of macromolecular crystals and provided new and important applications of the protocol in addition to overcoming an increase in mosaicity caused by flash-cooling. Such applications included rescue

of mishandled crystals and the ability to collect cryogenic data on the same crystal before and after treatments such as heavy metal derivative soaks. It has become routine in this laboratory to screen crystals by collecting diffraction images on the flash-cooled crystal and using the macromolecular annealing protocol to return the crystal to ambient temperatures, at which time the crystal can be subjected to a heavy metal derivative soak. Cryogenic data collected on the soaked crystal can then be compared to the original diffraction data to determine unequivocally if the soak had an effect on diffraction.

The usefulness of MCA for the recovery of mishandled crystals can be illustrated by two examples. If a crystal is mounted on a loop and flash-cooled in an undesirable orientation, annealing can be applied to correct the orientation of the crystal to improve the data collection strategy. An application which has been termed the "Lazarus effect" involves the accidental icing of a flash-cooled crystal by transient exposure to warm air. This effect is illustrated in Figure 4.1. The crystal had become cloudy after the accidental icing, and MCA was applied to bring diffraction back to normal. Before the discovery of macromolecular crystal annealing, such crystals were considered to have been destroyed and were always discarded.

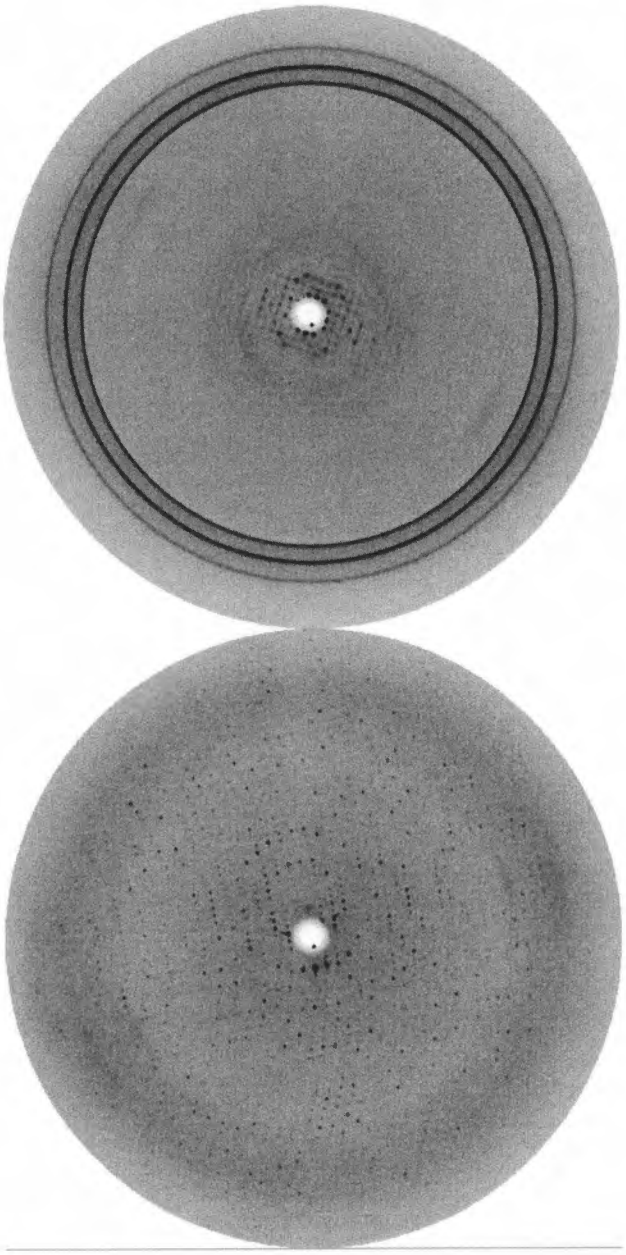

Figure 4.1: An example of the use of macromolecular crystal annealing to recover a mishandled crystal. An NCP crystal had been flash-cooled and used to collect X-ray diffraction data for about 2 days. Instrument maintenance then required that the crystal be stored in liquid nitrogen. After remounting the crystal to continue data collection, the first image (top) displayed signs of damage due to transient warming during handling. The crystal was annealed and diffraction returned to normal (bottom).

4.3 Structure of the nucleosome core particle as seen in the 2.5 Å model

4.3.1 Influence of the solution environment on the crystal structure

NCP crystals taken directly from the mother liquor do not diffract well and the unit cell dimensions are highly variable. The primary source of variability appears to be stacking along the long axis of the crystal, the c-axis. For this reason, the crystals have been stabilized by adding an organic cosolvent such as MPD. With increasing MPD concentration, the length of the c-axis is reduced and the diffraction limit extended to an optimum level of MPD, beyond which diffraction begins to suffer. It was once thought that MPD and similar organic cosolvents served to lower the activity of the bulk water in the crystal by organizing water structure and thus remove water available for hydration of protein structures. However, MPD acts conversely to increase the sphere of hydration on proteins (Timasheff, 1993; Pittz & Timasheff, 1978) and it now appears that the dielectric constant is the operative parameter (Flock *et al.*, 1996; Arscott *et al.*, 1995). Of particular concern to the structure of the NCP is the effect of MPD on the structure of DNA. It is not clear what, if any, impact MPD may have on bending due to phased A-tracts (Sprous *et al.*, 1995; Dickerson *et al.*, 1996). The effect of an organic cosolvent on the mobility of NCPs

in gel electrophoresis (Pennings *et al.*, 1992) may explain the response of the c-axis dimension to varying levels of MPD. A decrease of the dielectric constant increases the charge repulsion effect of phosphates and decreases bending at the DNA termini. This allows the establishment of appropriate base stacking between termini of neighboring particles and decreases the variability of the orientation of the dyad axis with respect to the c-axis.

DNA superhelical structure and nucleosome design

The design of the structure of the histone octamer core of the nucleosome appears to have been engineered for two intimately related functions; packaging, and command-and-control. The command-and-control function is mediated primarily through the N-terminal tail domains. The basis of the command-and-control functions is not directly answered by the NCP structure, however, very significant insights into the packaging function are provided. In order to package the entire eukaryotic genome into chromatin and fit the entire assemblage into the cell nucleus, evolution has had to engineer the structure of the nucleosome to optimize the degree of compaction. The high degree of conservation of the sequences of core histones between remotely related taxa (van Holde, 1988) attests to the significance of this engineering. The histone fold and handshake motifs found repeating through the octamer provide a flexible framework to position the DNA binding sites, mGIs, at appropriate positions to allow for bending of DNA into a nucleosome structure. It seems to be no accident that the

bend-persistence length of average B-form DNA is nearly the same as the length of DNA in the nucleosome core particle (Bustamante *et al.*, 1994; Smith *et al.*, 1996). The measured bend-persistence of average B-form DNA is 500 Å. The length of DNA in the NCP, 146 bp, is 496.4 Å assuming a canonical 3.4 Å /bp rise. DNA structure can be viewed as a rod-like polymer which on short length scales is rigid and on long length scales is quite flexible. The helical axis at a given base pair points in a particular direction and the direction of the helical axis at neighboring base pairs is highly correlated. The helical axis for base pairs further away will be less well correlated to a degree related to the so-called persistence of the direction of the helical axis, which relates to the flexibility of the rod. Bend-persistence length of an elastic rod is the length scale beyond which the elastic cost of bending is negligible. Thus, evolution has optimized the structure of the histone octamer to form nucleosomes on average or, equivalently, random sequence DNA in the genome. However, this is the length of DNA in the NCP which lacks linker DNA. DNA is not wrapped in two full superhelical turns around the histone core and, indeed, it can be seen that in the NCP structure (see Figure 3.33), the DNA termini beyond mGI 6 are nearly straight. It is now thought that the linker DNA enters and exits the nucleosome in a stem-like structure (Bednar *et al.*, 1998) rather than wrapping smoothly twice around each nucleosome. Most of the energy of bending is spent on the DNA in the NCP between ventral and dorsal mGI 6. Reference to the previously published 8 Å structure (Uberbacher & Bunick, 1989) of bulk chicken erythrocyte NCPs containing random sequence DNA

shows that this is likely to be a feature of the average nucleosome. As shown in Figure 4.2, the DNA termini appear to be nearly straight. Also evident in Figure 4.2 is some distortion in the helical structure at mGI 2, which may correspond to the distortion seen in the NCP structure here (see Figure 3.33), and suggests that the distortion seen in the NCP is not strictly a feature of the particular DNA sequence used here but may be a more general feature of DNA binding in the nucleosome.

DNA binding to the histone octamer

The general features of the binding of DNA to the left-handed superhelical ramp provided by the histone octamer were essentially correctly identified by Moudrianakis and his colleagues (Arents & Moudrianakis, 1993) from the structure of the histone octamer and modeling of the DNA. However, the central role of the arginine residues inserting into the minor groove at the mGI sites are here seen in detail along with such features as the hydrophobic interaction of H3 Leu 65 in the major groove. The structure is, of course, similar to nucleosome structures seen at lower resolution. However, as a result of the concatenation of crystal growth techniques and crystal engineering technology, the structure here represents a significant improvement in the resolution of the structure and the determination of the atomic position for both the protein and DNA components of the NCP.

The well established importance of polyA tract DNA, distributed approximately in phase with the helical repeat of DNA, to the ability of a DNA fragment to bind to

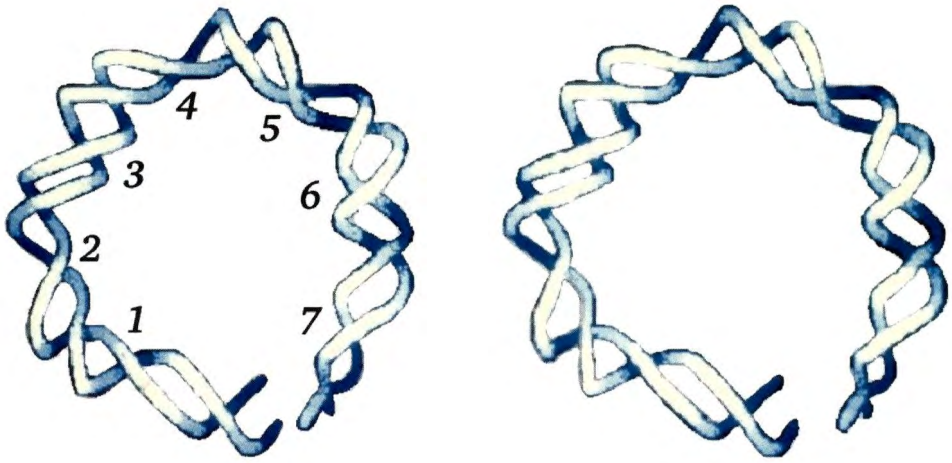

Figure 4.2: Stereo view of the DNA phosphate backbone trace from the 8 Å structure of the NCP. The trace represents an interpretation of the model for bulk chicken erythrocyte NCPs containing random sequence DNA trimmed by nuclease digestion. The model provides a view of the average structure of DNA in the nucleosome. The mGIs are numbered as in the 2.5 Å NCP structure. Only one half of the DNA is presented as asymmetries are not resolved in the 8 Å model. A distortion of the helical structure of the DNA is evident at mGI 2 in accordance with the structure at 2.5 Å, and the DNA termini beyond mGI 6 are essentially straight.

the histone octamer is here given a novel interpretation. Apart from the controversy concerning curvature in polyA tract DNA (Sprous *et al.*, 1995; Sprous *et al.*, 1999; Haran *et al.*, 1994; Dickerson, 1998; Olson & Zhurkin, 2000), it is generally agreed that polyA tracts serve to narrow the minor groove. In nucleosomal DNA sequences, AA dinucleotide steps flanking the minor groove-in position bring DNA phosphates closer together and provide an improved substrate for insertion of an arginine side chain at the mGI binding site. The path of the phosphate backbone also more closely approximates the separation of the histone fold loops forming the foundation of a βmGI binding motif. Thus, the width of the minor groove at the point at which DNA contacts the mGI binding motif may be the important factor determining rotational phasing of nucleosomes.

Packing interactions

The packing of the NCP in the crystal lattice is unlikely to reflect real packing of nucleosomes connected by linker DNA. The strong interaction of the protein faces of the NCP in the crystal structure involves a distinctive acidic surface on the dimer which binds to the tetramer surface of a neighboring NCP, particularly through the N-terminal tail of H4. This interaction is certainly suggestive of strong binding between nucleosomes in the condensed or 30 nm fiber even though the head-to-tail orientations of the two particles in the lattice is not appropriate to any current model of higher-order structures. The extremely tight binding of a cacodylate ion with-

in the dimer-tetramer interface connecting 2 NCPs provides a putative mechanism explaining the mutagenic properties of dimethyl arsenate. In addition to its use in crystallization buffers, cacodylate has been used as an herbicide and is the major metabolite of inorganic arsenic compounds, and is known to cause chromosomal aberrations (Kashiwada *et al.*, 1998).

4.3.2 Applications of the NCP structure to questions of function

The structure reported here represents more than simply a picture of the NCP. It represents a great wealth of data and a tool which can be applied to the analysis of functional studies. At the time of this writing, the structure has been accessed in efforts to answer questions related functional issues. Two examples will be discussed. The first involves the question of the function of an H3-like protein, D6H3, in the nematode *Caenorhabditis elegans* which is known to be localized to the centromere. And, the second involves an issue of possible functions for selective methylation of four arginine residues at the C-terminus of H3 in the so-called hinge region of the nucleosome.

Homology modeling of an H3-like protein

As part of a collaboration with Dr. Steven Henikoff of the Fred Hutchinson Cancer Research Center, Seattle WA, we received clones and sequences for a series of H3-like proteins which had been shown to replace H3 in nucleosomes localized to centromeres. One of these, D6H3 from *C. elegans*, was expressed, purified, and used to reconstitute nucleosomes using recombinant H2A, H2B, and H4 based on chicken protein sequences. The D6H3 formed authentic nucleosomes (data not shown). An examination of the sequence of D6H3 revealed the presence of 3 cysteine residues; one near the N-terminus and two near the C-terminus. H3 contains only a single cysteine residue, C110, whose function is not clear although it occurs at the hinge region formed by the four-helix bundle at the H3-H3 interface. The C-α atoms of the two residues are 7.8 Å apart so that the sulfhydryl groups are not in a position to form a disulfide bridge between the two halves of the tetramer. The amino acid sequence of D6H3 was subjected to homology modeling using the Swiss-Model server (Guex & Peitsch, 1997; Peitsch, 1996; Peitsch, 1995), with the coordinates of H3 in the NCP as the model. The resulting coordinates were then used to replace H3 in the NCP structure. Examination of the structure immediately revealed the strong structural homology of D6H3 and H3 and the position of Cys 79 of D6H3 placed in precisely the same position as Cys 110 in H3. Further, Cys 83 of D6H3 was positioned at the C-terminus of the mH helix of H3 (see Figure 4.3). These cysteines may well be involved in preventing mobility in the four-helix bundle and prevent the

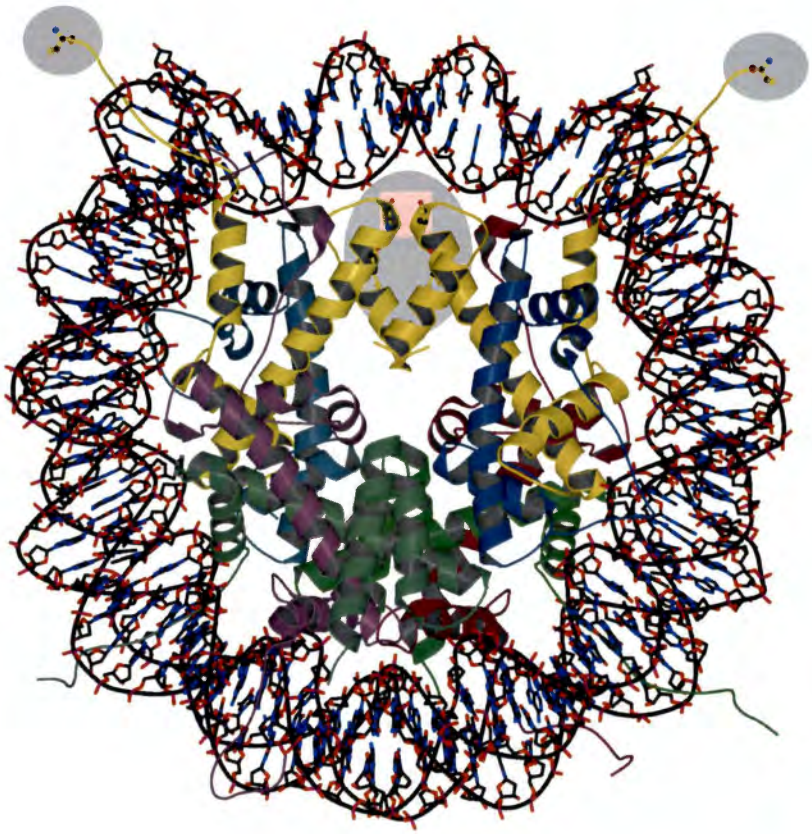

Figure 4.3: Homology models of *C. elegans* H3-like protein D6H3 inserted into the NCP structure. Grey oval indicate regions of the structure containing cysteine residues in D6H3. Cys 83 at the C-terminus of the mH helix of D6H3 is further indicated with a pink square.

chiral transition proposed by Prunell and colleagues (Hamiche *et al.*, 1996). They may also function in making the DNA in the nucleosome accessible for transcription. The cysteine near the N-terminus of D6H3 intriguingly occurs near the base of the N-terminal tail. Even providing for the *caveat* that the nucleus provides a reducing environment, the dorsal and ventral positions of the N-terminal cysteine in D6H3 is highly suggestive of internucleosomal interactions in chromatin structure at the centromere.

Methylation of arginines at the C-terminus of H3

In the course of a collaboration with Dr. Dana Aswad of the University of California, Irvine, the question was raised concerning the possible function of methylation of arginines at the C-terminus of H3 by coactivator-associated protein arginine methyltransferase, CARM1. In addition to methylation of arginines on the N-terminal tail of H3, CARM1 was shown to selectively methylate arginines 128, 129, 131, and 134 (Schurter *et al.*, 2000). Investigation of the possible functional role of methylation of these arginines was undertaken by examination of the positions of the arginine side chains and determination of the possible structural consequences of mono- and dimethyl arginines at these positions. The disposition of the side chains of the four arginine residues at the C-terminus of H3 can be seen in Figure 4.4. The arginines function to stabilize the four-helix bundle in the hinge region of the NCP. In particular, Arg 131 is buried with hydrogen bonding to Arg 134 is solvent exposed and

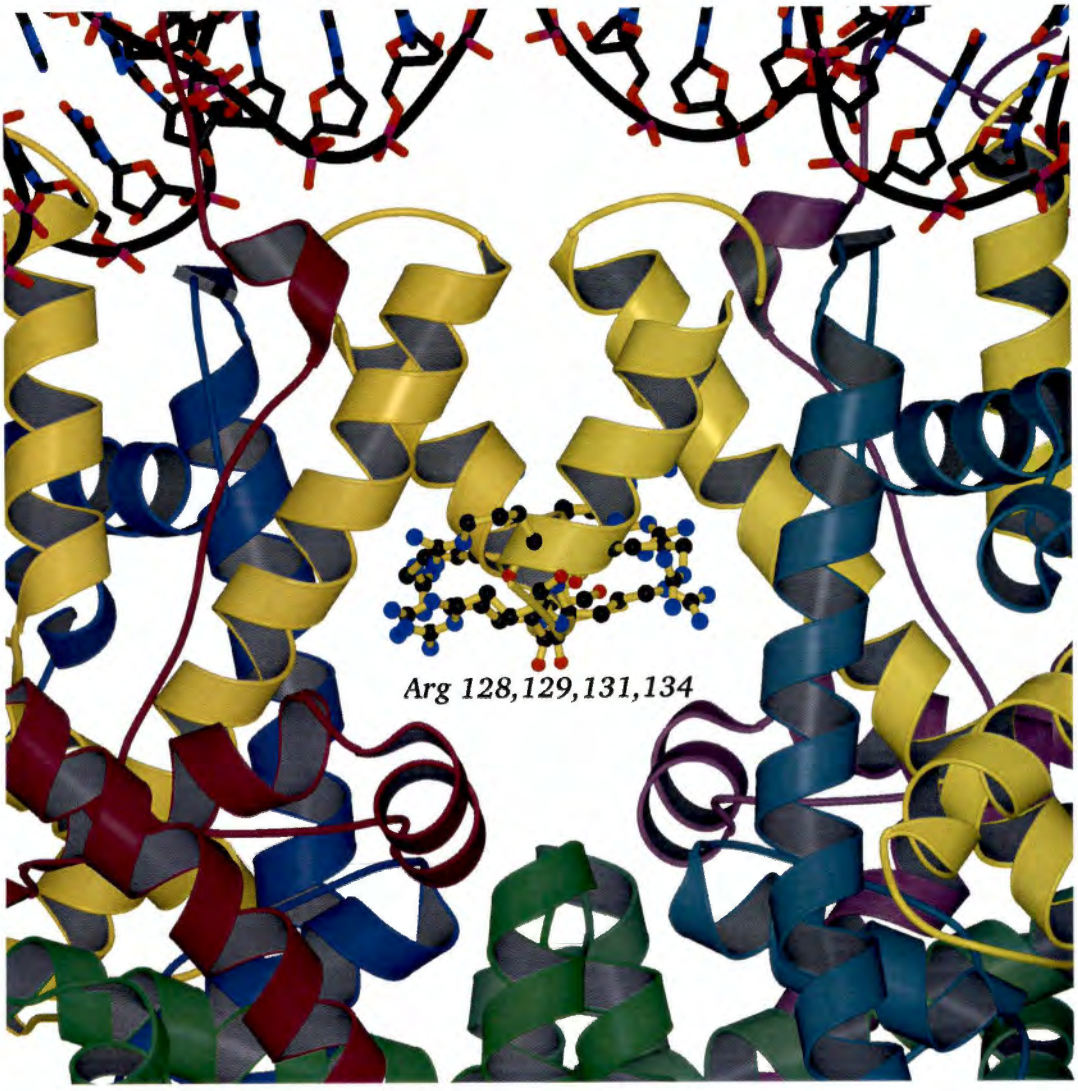

Figure 4.4: Positions of arginines 128, 129, 131, and 134 on the C-terminus of H3

could conceivably be a substrate for a protein methyltransferase. However, Arg 131 is buried and hydrogen bonded to Asp 106 and Glu 133 of the same H3 chain as well as a number of solvent molecules. Arg 128 and Arg 129 are also involved in hydrogen bond contacts to stabilize the four-helix bundle. Arg 128 is hydrogen bonded to Glu 133, and Arg 128 of 2H3 is additionally water bridged to Tyr 99. Interestingly, Arg 129 on 2H3 is found in 2 conformations. One conformation is identical to that seen for Arg 129 on 1H3, with hydrogen bonding to Asp 106 on 2H3 and a water bridge to Glu 105 on 2H3. In the second conformation, the side chain is positioned to water bridge with Asn 110 on 2H2A. Both Arg 128 and Arg 129 appear capable of interacting in hydrophobic pockets within the four-helix bundle if methylated. Arg 128 could be involved in a hydrophobic pocket formed by Val 60 and Val 57 of H4 and Leu 103 and Ile 124 of the same H3 chain. Arg 129 could be involved in a hydrophobic pocket formed by Leu 109 and Ile 112 on the same H3 chain and Ile 130 on the same H3 chain. Methylation of the 4 arginines could destabilize the four-helix bundle at the hinge region and allow greater rotational mobility of the (H3·H4) heterodimers around the hinge region in nucleosomes in transcriptionally active chromatin.

4.3.3 Future directions

To better understand the functions of different covalent modifications that occur to the histones of the NCP, atomic positions for the histone tails must be extended and more precisely defined. Although these regions are sites of inherent disorder within

the molecule, crystal packing contacts have provided some clues about possible chromatin interactions. Sites of biological significance for acetylation, ADP-ribosylation, phosphorylation, methylation, and ubiquitination all occur at positions on the histone tails. The structural basis of the function of these post-translational modifications is now an even more important goal. The physiological role of these modifications is clear and may represent a kind of epigenetic code regulating precise functions of genetic elements (Strahl & Allis, 2000). Improved structural characterization of histone tails using native variant histones and with those engineered to contain specific modifications, may provide additional insight into these regulatory processes. However, the lack of clear evidence for the role of the histone tails in this structure suggests that a radically more complex solution is required. The α-satellite DNA sequence used in the NCP structure determination was originally subcloned from a *Bam*H I repeat of α-satellite DNA from the human X chromosome and is known to contain 12 evenly spaced and strongly phasing nucleosomal sequences. Reconstitution of nucleosomes onto the full length *Bam*H I fragment could result in a synthetic chromatin preparation which may be sufficiently homogeneous for crystallization. Such experiments are very likely to require the microgravity environment aboard the International Space Station. The need for an authentic chromatin structure in elucidating the role of the histone tails is clear and immediate.

Recent studies assessing the potential for cryogenic helium in macromolecular cryocrystallography (Hanson *et al.*, 1999; Hanson *et al.*, 2000) bode well for further

elucidating the structure of the histone tails. An open-flow helium cryostat capable of holding a crystal at a temperature below 20 K at a third-generation synchrotron source has been shown to increase diffraction resolution, lower temperature factors, and improve electron density for a number of crystal systems including the NCP. Although first experiments with NCP crystals showed some evidence of structure modulation due to the extreme temperature, we are confident of acquiring greater detail about histone tail atomic positions, modifications, and interactions, as well as providing more details of the structure of DNA in the nucleosome.

With the solution of the NCP structure, opportunities arise to use this structure as the anchor for structural studies of other chromatin assemblages. Modified nucleosomes occur at the centromere, with histone H3 replaced by H3-like proteins. These modified nucleosomes serve as sites of attachment for fibrils during sister chromatid disjunction at metaphase. We have produced nucleosome core particles containing the H3-like protein D6H3 from *C. elegans*. This protein contains 3 cysteine residues rather than the single residue in H3. A structural model (Harp, *et al.* in preparation) shows that one of the cysteines is structurally homologous to that in H3 and that one of the cysteines occurs at the C-terminus of the mH helix of D6 in a position to form a disulfide bond across this hinge region of the tetramer. This is in contrast to H3, for which cysteines are not within disulfide bond-forming distance. A third cysteine residue is located on the N-terminal tail domain of D6H3.

Other studies need to be directed towards increasing the database of DNA sequences

known to atomic resolution. One simple set of experiments includes changing the length of the DNA sequence to look at the effects of 145 and 147 bp DNA fragments on the NCP structure. We have also engineered a palindrome from the other half of the full length α-satellite DNA sequence used in this structure. Our identification of correctly phasing α-satellite DNA sequences (Uberbacher *et al.*, 1988) and the development of the specific palindromic DNAs for crystallization have made high-resolution NCP structures attainable (Harp *et al.*, 1996). However, NCPs with non-palindromic DNA that diffract to lower resolution have been made and could more accurately reflect the DNA sequences found in nature. A 3.1 Å NCP structure using native α-satellite Human X chromosome DNA will be described in a separate publication.

Binding of the NCP with other DNA interacting proteins offers an opportunity to provide atomic resolution models of DNA modulating complexes. One of those complexes is that formed between the NCP and HMG 14/17 (Mardian *et al.*, 1980; Uberbacher *et al.*, 1982; Trieschmann *et al.*, 1998; Postnikov & Bustin, 1999). The regions of highest temperature factors on the DNA correlate to binding sites for HMG 14/17. Structures of complexes will help establish the role of HMG 14/17 and other DNA interacting proteins in gene regulation and chromatin assemblage.

The structure of the NCP leaves many important questions unanswered and it should be emphasized that the NCP is not a physiological unit but an artifact of nuclease digestion. The structure of the nucleosome has been 25 years in the making (Kornberg

& Lorch, 1999) and is not yet complete. In this study, the absence of linker DNA resulted in an asymmetry in the binding of the C-terminal domains of H2A and may have obscured the physiological role of the N-terminal tails of H3 which emerge between the gyres of DNA near the termini of the DNA in this structure. Future crystallographic studies must focus on higher-order structures of chromatin including the linker DNA and the linker histones before these structural issues can be resolved.

Bibliography

Bibliography

Adams, P. D., Pannu, N. S., Read, R. J., & Brunger, A. T. (1997). *Proceeding of the National Academy of Sciences USA,* **94**, 5018–5023.

American Society for Cell Biology (1998). Approved by the ASCB Council, Spring, 1998.

Arents, G., Burlingame, R., Wang, B.-C., Love, W. E., & Moudrianakis, E. N. (1991). *Proceedings of the National Academy of Science, USA,* **88**, 10148–10152.

Arents, G. & Moudrianakis, E. (1995). *Proceedings of the National Academy of Sciences, USA,* **92**, 11170–11174.

Arents, G. & Moudrianakis, E. N. (1993). *Proceedings of the National Academy of Science, USA,* **90**, 10489–10493.

Arscott, P. G., Ma, C., Wenner, J. R., & Bloomfield, V. A. (1995). *Biopolymers,* **36**, 345–364.

Babcock, M. S. & Olson, W. K. (1993). *Computation of Biomolecular struc-*

tures:Achievements, Problems, and Perspectives chapter A new program for the analysis of nucleic acid structure interpretation, pp. 65–85. Heidelberg: Springer-Verlag.

Babcock, M. S., Pednault, E. P. D., & Olson, W. K. (1993). *Journal of Biomolecular Structure and Dynamics,* **11**, 597–628.

Babcock, M. S., Pednault, E. P. D., & Olson, W. K. (1994). *Journal of Molecular Biology,* **237**, 125–156.

Bakayev, V. V., Melnikov, A. A., Osicka, V. D., & Varshavsky, A. J. (1975). *Nucleic Acids Research,* **2**, 1401.

Bashkin, J., Hayes, J. J., Tullius, T. D., & Wolffe, A. P. (1993). *Biochemistry,* **32**, 1895–1898.

Baxevanis, A. D., Arents, G., Moudrianakis, E. N., & Landsman, D. (1995). *Nucleic Acids Research,* **23**, 2685.

Bednar, J., Horowitz, R. A., Grigoryev, S. A., Carruthers, L. M., Hansen, J. C., Koster, A. J., & Woodcock, C. L. (1998). *Proceeding of the National Academy of Sciences, USA,* **95**, 14173–14178.

Bentley, G. A., Finch, J. T., & Lewit-Bentley, A. (1981). *Journal of Molecular Biology,* **145**, 771.

Bentley, G. A., Lewit-Bentley, A., Finch, J. T., Podjarny, A. J., & Roth, M. (1984). *Journal of Molecular Biology,* **176**, 55–75.

Berman, H. M., Westbrook, J., Feng, A., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). *Nucleic Acid Research,* **28**, 235–242.

Brunger, A. T. (1992). *Nature,* **355**, 472–474.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., & Warren, G. L. (1998). *Acta Crystallographica,* **D54**, 905–921.

Burlingame, R. W., Love, W. E., & Moudrianakis, E. N. (1984). *Science,* **228**, 546.

Burlingame, R. W., Love, W. E., Wang, B.-C., Hamlin, R., Xuong, N.-H., & Moudrianakis, E. N. (1985). *Science,* **228**, 546.

Bustamante, C., Marko, J. F., Siggia, E. D., & Smith, S. (1994). *Science,* **265**, 1599–1600.

Carter, D. C., Lim, K., Ho, J. X., Wright, B. S., Twigg, P. D., Miller, T. Y., Chapman, J., Keeling, K., Ruble, J., Vekilov, P. G., Thomas, B. R., Rosenberger, F., & Chernov, A. A. (1999a). *Journal of Crystal Growth,* **196**, 623–637.

Carter, D. C., Wright, B., Miller, T., Chapman, J., Twigg, P., Keeling, K., Moody,

K., White, M., Click, J., Ruble, J. R., Ho, J. X., Adcock-Downey, L., Bunick, G., & Harp, J. (1999b). *Journal of Crystal Growth,* **196**, 602–609.

Chambon, P. (1978). *Cold Spring Harbor Symposia on Quantitative Biology,* **42**, 1209–1234.

Clark, R. J. & Felsenfeld, G. (1971). *Nature New Biology,* **229**, 101–106.

Collaborative Computational Project No. 4 (1994). *Acta Crystallographica,* **D50**, 760–763.

Crowther, R. A. & Klug, A. (1975). *Annual Review of Biochemistry,* **44**, 161.

D'Anna, J. A. & Isenberg, I. (1973). *Biochemistry,* **12**, 1035–1043.

D'Anna, J. A. & Isenberg, I. (1974a). *Biochemistry,* **13**, 2098–2104.

D'Anna, J. A. & Isenberg, I. (1974b). *Biochemistry and Biophysics Research Communications,* **61**, 343–347.

D'Anna, J. A. & Isenberg, I. (1974c). *Biochemistry,* **13**, 4987–4992.

D'Anna, J. A. & Isenberg, I. (1974d). *Biochemistry,* **13**, 4992–4997.

Delucas, L. J. (1999). *Science,* **284**, 1106–1107. in Letters.

Dickerson, R. E. (1998). *Nucleic Acids Research,* **26**, 1906–1926.

Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., von Kitzing, E., Lavery, R., Nelson, H. C., Olson, W. K., Saenger, W.,

Shakked, Z., Soumpasis, D. M., Tung, C.-S., Wang, A. H.-J., & Zhurkin, V. B. (1989). *Nucleic Acids Research,* **17**, 1797–1803.

Dickerson, R. E., Goodsell, D., & Kopka, M. L. (1996). *Journal of Molecular Biology,* **256**, 108–125.

Drenth, J. (1994). *Principle of protein X-ray crystallography.* : Springer-Verlag New York, Inc.

Drew, H. R., Wing, R. M., Takano, T., Broka, C., Itakura, K., & Dickerson, R. E. (1981). *Proceeding of the National Academy of Science, USA,* **78**, 2179.

Finch, J. T., Brown, R. S., Rhodes, D., Richmond, T., Ruston, B., Lutter, L. C., & Klug, A. (1981). *Journal of Molecular Biology,* **145**, 757.

Finch, J. T. & Klug, A. (1978). *Cold Spring Harbor Symposia in Quantitative Biology,* **42**, 1–9.

Finch, J. T., Lewit-Bentley, A., Bentley, G. A., Roth, M., & Timmins, P. A. (1980). *Philosophical Transactions of the Royal Society of London B,* **290**, 635.

Finch, J. T., Lutter, L. C., Rhodes, D., Brown, R. S., Rushton, B., Levitt, M., & Klug, A. (1977). *Nature,* **269**, 29–36.

Flock, S., Labarbe, R., & Houssier, C. (1996). *Biophysical Journal,* **70**, 1456–1465.

Furey, W. (1997). *Methods in Enzymology,* **277**.

Garman, E. (1999). *Acta Crystallographica,* **D55**, 1641–1653.

Garman, E. F. & Schneider, T. R. (1997). *Journal of Applied Crystallography,* **30**, 211–237.

Guex, N. & Peitsch, M. C. (1997). *Electrophoresis,* **18**, 2714–2723.

Hahn, T., ed (1995). *International Tables for Crystallography,* volume A Space-Group Symmetry. : International Union for Crystallography, fourth edition.

Hamiche, A., Carot, M., Alilat, F., De Lucia, F., O'Donohue, M.-F., Revet, B., & Prunell, A. (1996). *Proceedings of the National Academy of Sciences USA,* **93**, 7588–7593.

Hanson, B. L., Harp, J. M., Kirschbaum, K., Parrish, D. A., Timm, D. E., Howard, A., Pinkerton, A. A., & Bunick, G. J. (2000). *Journal of Crystal Growth,* . submitted.

Hanson, B. L., Harp, J. M., Parrish, D. A., Bunick, C. G., Kirschbaum, K., Pinkerton, A. A., & Bunick, G. J. (1999). *Journal of Applied Crystallography,* **32**, 814–820.

Haran, T. E., Kahn, J. D., & Crothers, D. M. (1994). *Journal of Molecular Biology,* **244**, 135–143.

Harp, J. M., Hanson, B. L., Timm, D. E., & Bunick, G. J. (1999). *Acta Crystallographica,* **D55**, 1329–1334.

Harp, J. M., Hanson, B. L., Timm, D. E., & Bunick, G. J. (2000). *Acta Crystallographica,* **D56** In the press.

Harp, J. M., Hanson, B. L., Timm, D. E., & Bunick, G. J. (2000b). *Acta Crystallographica,* In preparation.

Harp, J. M., Palmer, E. L., York, M., Geweiss, A., Davis, M., & Bunick, G. J. (1995). *Electrophoresis,* **16**, 1861–1864.

Harp, J. M., Timm, D. E., & Bunick, G. J. (1998). *Acta Crystallographica,* **D54**, 622–628.

Harp, J. M., Uberbacher, E. C., Roberson, A., & Bunick, G. J. (1996). *Acta Crystallographica,* **D52**, 283–288.

Hauser, L. (1985). *Structural analysis of members of a repeated DNA family in primates.* PhD thesis University of California, Irvine.

Hewish, D. R. & Burgoyne, L. A. (1973). *Biochemical and Biophysical Research Communications,* **52**, 504–510.

Hingerty, B. E., Bunick, G., Harp, J. M., & Wang, L. (2000). Computational modeling of DNA: Treatment of dielectric, charges and van der Waals radii. In preparation.

Hingerty, B. E., Figueroa, S., Hayden, T. L., & Broyde, S. (1989). *Biopolymers,* **28**, 1195–1222.

Hjelm, R. P., Kneale, G. G., Suau, P., Baldwin, J. P., & Bradbury, E. M. (1977). *Cell,* **10**, 139–151.

Hooft, R. W. W., Vriend, G., Sander, C., & Abola, E. E. (1996). *Nature,* **381**, 272.

Hope, H. (1990). *Annual Review of Biophysics and Biophysical Chemistry,* **19**, 107–126.

Ichimura, S., Mita, K., & Zama, M. (1982). *Biochemistry,* **21**, 5329–5334.

Kabsch, W. & Sander, C. (1983). *Biopolymers,* **22**, 2577–2637.

Kashiwada, E., Kuroda, K., & Endo, G. (1998). *Mutation Research,* **413**, 33–38.

Klug, A., Finch, J. T., & Richmond, T. J. (1985). *Science,* **29**, 1109.

Klug, A., Rhodes, D., Smith, J., Finch, J. T., & Thomas, J. O. (1980). *Nature,* **287**, 509.

Kornberg, R. (1974). *Science,* **184**, 868–871.

Kornberg, R. D. & Lorch, Y. (1999). *Cell,* **98**, 285–294.

Kraulis, P. J. (1991). *Journal of Applied Crystallography,* **24**, 946–950.

Laver, W. G. (1999). *Science,* **284**, 2089. in Letters.

Luger, K., Mader, A., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). *Nature,* **389**, 251–260.

Lutter, L. C. (1978). *Journal of Molecular Biology,* **124**, 391–420.

Malakoff, D. (1999). *Science,* **284**, 1106–1107. in News Focus.

Mardian, J. K. W., Paton, A. E., Bunick, G. J., & Olins, D. E. (1980). *Science,* **209**, 1523–1536.

Matthews, B. W. & Czerwinski, E. W. (1975). *Acta Crystallographica,* **A31**, 480–497.

Maxam, A. M. & Gilbert, W. (1980). *Methods in Enzymology,* **65**, 499–560.

McRee, D. E. (1999). *Journal of Structural Biology,* **125**, 156–165.

Merritt, E. A. & Murphy, M. E. (1994). *Acta Crystallographica,* **D50**, 869–873.

Moudrianakis, E. N. & Arents, G. (1993). *Cold Spring Harbor Symposia on Quantitive Biology. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory,* **58**, 273–279.

Moudrianakis, E. N., Love, W. E., & Burlingame, R. W. (1985). *Science,* **229**, 1113.

Nicholls, A., K.Sharp, & Honig, B. (1991). *Proteins: Structure, Function, and Genetics,* **11**, 281–296.

Olins, A. L. & Olins, D. E. (1973). *Journal of Cell Biology,* **59** (252a).

Olins, A. L. & Olins, D. E. (1974). *Science,* **183**, 330–332.

Olins, A. L. & Olins, D. E. (1983). *Current Contents (Life Sciences),* **26** (10), 25.

Olson, W. K. & Zhurkin, V. B. (2000). *Current Opinion in Structural Biology,* **10**, 286–297.

Otwinowski, Z. & Minor, W. (1997). *Methods in Enzymology* volume 276 chapter Processing of X-ray data collected in oscillation mode, pp. 307–326. : Academic Press.

Palmer, E., Geweiss, A., Harp, J. M., York, M. H., & Bunick, G. J. (1996). *Analytical Biochemistry,* **231**, 109–114.

Pardon, J., Cotter, R. I., Lilley, D. M. J., Worcester, D. L., Campbell, A. M., Wooley, J. C., & Richards, B. M. (1978). *Cold Spring Harbor Symposia on Quantitative Biology,* **42**, 11–29.

Peitsch, M. C. (1995). *Bio/Technology,* **13**, 658–660.

Peitsch, M. C. (1996). *Biochemistry Society Transactions,* **24**, 274–279.

Pennings, S., Meerssemans, G., & Bradbury, E. M. (1992). *Nucleic Acids Research,* **20**, 6667–6672.

Pittz, E. P. & Timasheff, S. N. (1978). *Biochemistry,* **17**, 615–623..

Postnikov, Y. V. & Bustin, M. (1999). *Methods in Enzymology,* **304**, 133–155.

Ramakrishnan, V. (1995). *Proceedings of the National Academy of Sciences, USA,* **92**, 11328–11330.

Ravelli, R. B. G., Sweet, R. M., Skinner, J. M., Duisenberg, A. J. M., & Kroon, J. (1997). *Journal of Applied Crystallography,* **30**, 551–554.

Read, R. J. (1986). *Acta Crystallographica,* **A42**, 140–146.

Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D., & Klug, A. (1984). *Nature,* **311**, 532–537.

Richmond, T. J., Searles, M. A., & Simpson, R. T. (1988). *Journal of Molecular Biology,* **199**, 161–170.

Rodgers, D. W. (1994). *Structure,* **2**, 1135–1140.

Rodgers, D. W. (1997). *Methods in Enzymology,* **276**, 183–203.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory, New York: Cold Spring Harbor Laboratory Press, second edition.

Sauer, U. H. & Ceska, T. A. (1997). *Journal of Applied Crystallography,* **30**, 71–72.

Schurter, B. T., Koh, S. S., Chen, D., Bunick, G. J., Harp, J. M., Hanson, B. L., Henschen, A., Stallcup, M. R., & Aswad, D. W. (2000). . In preparation.

Simpson, R. T. & Stafford, D. W. (1983). *Proceeeding of the National Academy of Sciences, USA,* **80**, 51–55.

Smith, S. B., Finzi, L., & Bustamante, C. (1996). *Science,* **258**, 1122–1126.

Sprous, D., Young, M. A., & Beveridge, D. L. (1999). *Journal of Molecular Biology,* **285**, 1623–1632.

Sprous, D., Zacharias, W., Wood, Z. A., & Harvey, S. C. (1995). *Nucleic Acids Research,* **23**, 1816–1821.

Stein, A. & Page, D. (1980). *Journal of Biological Chemistry,* **255**, 3629–3637.

Stout, G. H. & Jensen, L. H. (1989). *X-ray structure determination: A practial guide.* : John Wiley and Sons.

Strahl, B. D. & Allis, C. D. (2000). *Nature,* **403**, 41–45.

Timasheff, S. N. (1993). *Annual Review of Biophysics and Biomolecular Structure,* **22**, 67–97.

Toldo, L. I. G. (1995). *Crystallographic Studies on Chromatin: The Nucleosome Core Particle and The Chicken Histone Octamer.* PhD thesis ETH, Zurich, Switzerland.

Trieschmann, L., Martin, B., & Bustin, M. (1998). *Proceedings of the National Academy of Sciences, USA,* **95**, 5468–5473.

Uberbacher, E. C. & Bunick, G. J. (1985a). *Journal of Biomolecular Structure and Dynamics,* **2**, 1033–1055.

Uberbacher, E. C. & Bunick, G. J. (1985b). *Science,* **229**, 1112.

Uberbacher, E. C. & Bunick, G. J. (1989). *Journal of Biomolecular Structure and Dynamics,* **7**, 1–18.

Uberbacher, E. C., Harp, J. M., Wilkinson-Singley, E., & Bunick, G. J. (1986). *Science,* **232**, 1247.

Uberbacher, E. C., Mardian, J. K. W., Rossi, R. M., Olins, D. E., & Bunick, G. J. (1982). *Proceedings of the National Academy of Sciences, USA,* **79**, 5258–5262.

Uberbacher, E. C., Wilkinson-Singley, E., Harp, J. M., & Bunick, G. J. (1988). In: *DNA Bending & Curvature*, (Olson, W. K., Sarma, M. H., Sarma, R. H., & Sundaralingam, M., eds) volume 3 of *Structure & Expression* pp. 139–158, Schenectady, NY: Adenine Press.

van den Akker, F. & Hol, W. (1999). *Acta Crystallographica,* **D55**, 206–218.

van Holde, K. E. (1988). *Chromatin.* New York: Springer-Verlag.

van Holde, K. E., Allen, J. R., Tatchell, K., Weischet, W. O., & Lohr, D. (1980). *Biophysical Journal,* **32**, 271–282.

Vriend, G. (1990). *Journal of Molecular Graphics,* **8**, 52–56.

Wang, B.-C. (1985). *Methods in Enzymology,* **115**, 90–112.

Wang, B.-C., Rose, J., Arents, G., & Moudrianakis, E. N. (1994). *Journal of Molecular Biology,* **236**, 179.

Widom, J. (1998). *Annual Reviews of Biophysics and Biomolecular Structure,* **27**, 285–327.

Williamson, R. (1970). *Journal of Molecular Biology,* **51**, 157–168.

Woodcock, C. F. L. (1973). *Journal of Cell Biology,* **59**, 368a.

Yang, T. P., Hansen, S. K., Oishi, K. K., Ryder, O. A., & Hamkalo, B. A. (1982). *Proceeding of the National Academy of Sciences USA,* **79**, 6593–6597.

Yeh, J. I. & Hol, W. G. J. (1999). *Acta Crystallographica,* **D54**, 479–480.

Vita

Joel Harp was born April 26, 1950 in Amarillo, Texas to A. M. (Buddy) and Theresa Harp. He graduated from Amarillo High School in 1968. Joel spent two years at Amarillo College and was then accepted in the Biology Department at West Texas State University (now West Texas A&M) in Canyon, Texas and received a Bachelor of Science in 1973. He remained at WTSU and entered the graduate program under Dr. Robert A. Wright and earned a Master of Science degree in Biology specializing in arid lands research and community ecology. He was then accepted into the doctoral program in the Department of Zoology at the University of Tennessee in Knoxville under Dr. Susan Riechert and Dr. M. L. Pan. His studies there included the taxonomy, behavioral ecology, and physiology and anatomy of spiders. His dissertation research involved the histology and ultrastructure of the spider digestive tissue. After having spent many years in that program, Joel underwent an early "mid-life crisis" and left the program after completing all requirements save writing the dissertation.

After ending his zoology career, Joel obtained a position, in 1983, as electron micro-

scope technician in the laboratory of Dr. Ada Olins in the Biology Division of Oak Ridge National Laboratory. It was here that he was introduced to chromatin and after 2 years in that position decided to accept a position with Drs. Gerard Bunick and Edward Uberbacher who gave him a remarkable second chance. Drs. Uberbacher and Bunick were working on the structure of the nucleosome core particle at 8 Å and needed a molecular biologist. At that time, Joel did not know a vector from an insert but was smart enough to not admit it. He was made a research associate, provided with a laboratory, a new copy of "Maniatis" and given the assignment of preparing nucleosome core particle crystals containing defined sequence DNA. During the years required to complete that task, he matured as a scientist and decided, under the generous auspices of Dr. Bunick, to enter the doctoral program in the Graduate School of Biomedical Sciences and continue working as an X-ray crystallographer. It has been a much better choice than working for a living.