INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®
NOTE TO USERS

This reproduction is the best copy available.

UMI
SITE-SPECIFIC DNA INTERACTIONS OF THE ARCHAEAAL HISTONE HMF FROM THE HYPERTHERMOPHILIC METHANOGEN METHANOTHERMUS FERVIDUS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Divya J. Soares, B.S., M.S.

*****

The Ohio State University

2001

Dissertation Committee:

John N. Reeve, Adviser

Joseph A. Krzycki

Charles J. Daniels

Pravin T. Kaumaya

Approved by

[Signature]

Department of Microbiology
ABSTRACT

HMf, a histone from the methanogenic archaeon, *Methanothermus fervidus*, is a member of the archaeal protein family that bears close homology to the eukaryal nucleosomal core histones H2A, H2B, H3 and H4 based on three-dimensional structures (Starich et al., 1996, Zhu et al., 1998, Decanniere et al., 2000). HMf contains homodimers and heterodimers of two small, basic polypeptides HMfA and HMfB that are ~85% identical and bind non-sequence specifically to DNA (Sandman et al., 1990, Bohrmann et al., 1994, Grayling et al., 1997). EM studies have shown that HMf forms nucleosomes *in vitro* and *in vivo* (Sandman et al., 1990, Pereira et al., 1997), and HMf nucleosomes have been found to be most similar to the (H3+H4) tetramer at the center of the eukaryal nucleosome (Bailey et al., 1999, Pereira and Reeve, 1999, Sandman and Reeve, 1999). The goal of this study was to identify amino acids that are responsible for contacting the DNA in archaeal nucleosomes. Site-specific mutagenesis of rHMfB was performed to generate ~50 variants. The abilities of these variants to form stable archaeal nucleosome-containing complexes with linear pBR322 DNA, with an 89 bp restriction fragment of this DNA and a 110 bp higher affinity Selex DNA were studied. rHMfB variants that did not generate such complexes in gel-shift assays had substitutions for residues at the N terminus or within α1, L1 and L2 regions which are homologous to the histone-fold residue locations in the eukaryal nucleosome. Variants that failed to produce
gel-shifts were further analyzed for their abilities to induce ligase-catalyzed circularization of an 88 bp DNA molecule, and to decrease the ellipticity of DNA at 275 nm (θ275) in CD assays. Experiments designed to test cooperativity of DNA binding showed that a combination of three residue substitutions in α1, L1 and L2 was needed to generate a rHMfB variant with no detectable DNA binding based on gel shift, circularization and θ275 reduction assays. Hence, even though DNA binding is independent at these three histone-fold locations, it is additive in that the loss of binding at one location does not prevent DNA binding at other locations. Higher affinity variants of rHMfB were also constructed and studied for their interactions with DNA using the aforementioned assays. These variants bound to DNA with a higher affinity and formed more stable nucleosomes due to additional lysine substitutions in the L1 and L2 regions that interact with the DNA minor groove. Finally, affinity constants of rHMfA, rHMfB and higher affinity rHMfB-G51K variant nucleosomes were measured using surface plasmon resonance and isothermal calorimetry assays which confirmed that these histones interact differently with DNA.
This work is dedicated to my family (Joe, Vilma, Ayres and Yvonne, Jitesh, Dinesh, Rohan, Rob, Melissa and Rupal) for their unwavering support and guidance through the years.
ACKNOWLEDGMENTS

I would like to thank my adviser Dr. John Reeve for giving me the opportunity to work in his lab. Dr. Reeve has always been very generous with his time and has been extremely supportive of my ideas. I would also like to thank him for giving me the freedom to take this project in any direction I chose. I would like to thank my committee members, Dr. Charles Daniels, Dr. Joseph Krzycki and Dr. Pravin Kaumaya for all their helpful suggestions in designing these experiments and for their time spent in serving on this committee. My special thanks goes out to Dr. Kathleen Sandman for all her creative ideas on this project as well as an overabundant supply of strains and proteins like topoisomerase I. I would also like to thank Dr. Suzette Pereira, Dr. Kathryn Bailey and Dr. WenTyng Li for their help in designing several DNA binding assays and Dr. Bailey for a generous supply of Selex DNA. I would like to thank Dr. Frederick Marc and Brent Christner for their help with computer-related issues and to all the members of the Reeve lab (both past and current) for making it such an enjoyable and stimulating experience. I would also like to thank Don Ordaz and John-David Sears for their help with fermentations.

I would like to acknowledge the contributions of several collaborators: Dr. Zvika Kelman and Dr. Frederick Schwarz at the University of Maryland (Center for Advanced Research in Biotechnology) for their time and help in developing the biacore and
microcalorimetry experiments. I would like to thank to Dr. Udo Heinemann for several helpful discussions.

I am most grateful to my dad, Ayres, for encouraging me to pursue this goal, for offering his constant support, and for being an inspiration to me every step of the way. I would also like to thank him for helping us with all our statistical analyses. I would also like to thank my brother, Jitesh, for his help in generating several of these figures and for all his support over the years. I am also grateful to my friends Todd Kavaleros, Cristy Utz, Leyla Gonzalez-Finzer and Dr. Mary Tichi for all their support and contribution of ideas to this project.
VITA


1995.................................................................B.S. in Microbiology,
                          The Ohio State University

2000.................................................................M.S. in Microbiology,
                          The Ohio State University

PUBLICATIONS


FIELDS OF STUDY

Major Field: Microbiology

Minor Field: Molecular Biology
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>iv</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>v</td>
</tr>
<tr>
<td>VITA</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xiv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xxi</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xxii</td>
</tr>
<tr>
<td>CHAPTER 1: GENERAL INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Genome compaction</td>
<td>1</td>
</tr>
<tr>
<td>Eukaryal nucleosome structure</td>
<td>4</td>
</tr>
<tr>
<td>Histone-like proteins in Bacteria</td>
<td>13</td>
</tr>
<tr>
<td>Nucleoid proteins in Archaea</td>
<td>18</td>
</tr>
<tr>
<td>HMF family of archaeal nucleosomes</td>
<td>22</td>
</tr>
<tr>
<td>The histone fold structure of eukaryal nucleosomes</td>
<td>28</td>
</tr>
<tr>
<td>The histone fold structure of rHMfB</td>
<td>35</td>
</tr>
<tr>
<td>Histone fold –DNA contacts in eukaryal nucleosomes</td>
<td>47</td>
</tr>
<tr>
<td>Interactions of histone extensions and tails</td>
<td>55</td>
</tr>
<tr>
<td>Eukaryal transcriptional regulation by histones</td>
<td>58</td>
</tr>
</tbody>
</table>
CHAPTER 2: IDENTIFICATION OF AMINO ACID RESIDUES IN THE
ARCHAEAL HISTONE, rHMfB, THAT CONTACT DNA

INTRODUCTION

MATERIALS AND METHODS

Chemicals, reagents and kits

Site-specific mutagenesis and purification of rHMfB variants

Folding studies of rHMfB variants using circular dichroism

Agarose EMSA with rHMfB variants

Polyacrylamide EMSA with rHMfB variants and 89 bp $[^{32}P]$-DNA

Polyacrylamide EMSA with rHMfB variants and $[^{32}P]$-Selex clone 1 DNA

Construction of (R19S+T54K) and (K13T+R19S+T54K) variants DNA circularization assays with rHMfB variants

Development of CD assays for archaeal histones

Copper-phenanthroline crosslinking of rHMfB cysteine variants

RESULTS

Construction and purification of rHMfB variants

Structural studies of archaeal histone folding

DNA binding assays of rHMfB variants using agarose
CHAPTER 3: DNA BINDING STUDIES OF THE HIGHER AFFINITY rHMfB VARIANTS

INTRODUCTION

MATERIALS AND METHODS

Chemicals, reagents and kits

Site-specific mutagenesis and purification of rHMfB variants

CD analyses of rHMfB variants

Agarose EMSA with the higher affinity rHMfB variants

Polyacrylamide EMSA with the higher affinity rHMfB
variants and 89 bp $^{32}$P-pBR322 DNA ........................................ 151
Polyacrylamide EMSA with the higher affinity rHMfB
variants and 110 bp $^{32}$P-Selex clone 1 DNA ............................. 151
CD assay for measurement of DNA binding of the higher
affinity rHMfB variants ............................................................... 152
DNA circularization assays with the higher affinity rHMfB
variants DNA topology studies of the higher affinity
rHMfB variants ........................................................................... 152
Preparation of DNA substrates for topology assays ................. 153
Topology assays of complexes with the higher affinity
rHMfB variants ........................................................................... 153
2-D agarose gel electrophoresis to distinguish between
negatively and positively supercoiled topoisomers .......... 154
RESULTS ...................................................................................... 155
Construction, purification and protein folding studies of
rHMfB higher affinity variants ................................................... 155
Agarose and polyacrylamide EMSAs with the higher
affinity variants ........................................................................... 156
Circular dichroism assays of rHMfB higher affinity variant
interactions with DNA ............................................................... 171
DNA circularization assays with the higher affinity rHMfB
variants ...................................................................................... 171
DNA topology studies with the higher affinity rHMfB variant

DISCUSSION

CHAPTER 4: AFFINITY STUDIES OF WILD-TYPE rHMfA AND rHMfB AND THE rHMfB VARIANTS FOR DNA

INTRODUCTION

MATERIALS AND METHODS

Reagents and kits

Construction and purification of DNA substrates for Biacore experiments

Purification of archaeal histones for Biacore experiments

Development of the surface plasmon resonance (Biacore) assay

Immobilization of the DNA on a sensor CM5 chip

Comparative binding of archaeal histones to the immobilized DNA

Equilibrium affinity assays of archaeal histone-DNA interactions using Biacore

Preparation of DNA for isothermal titration calorimetry (ITC) experiments

Purification of archaeal histones for ITC experiments

Development of ITC assays to evaluate archaeal
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Phylogenetic tree for members of the Euryarchaeota</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>The histone fold structure</td>
<td>6</td>
</tr>
<tr>
<td>1.3</td>
<td>Crystal structure of the nucleosome core particle</td>
<td>8</td>
</tr>
<tr>
<td>1.4</td>
<td>Alignment of the 4 eukaryal core histones based on primary sequence and structure</td>
<td>11</td>
</tr>
<tr>
<td>1.5</td>
<td>Primary sequence alignments of the archaeal histones with the eukaryal core histone consensus sequences</td>
<td>24</td>
</tr>
<tr>
<td>1.6</td>
<td>The (H3+H4) histone-fold DNA interactions</td>
<td>30</td>
</tr>
<tr>
<td>1.7</td>
<td>The (H2A+H2B) histone fold pair interactions with DNA</td>
<td>32</td>
</tr>
<tr>
<td>1.8</td>
<td>Schematic representation of the rHMfB crystal structure showing favorable salt bridge and/or hydrogen bond interactions</td>
<td>39</td>
</tr>
<tr>
<td>1.9</td>
<td>Schematic drawing of the (HMfA)\textsubscript{2} crystal structure showing ionic and/or hydrogen bond interactions</td>
<td>41</td>
</tr>
<tr>
<td>1.10</td>
<td>Schematic representation of the (HMfB)\textsubscript{2} crystal structure in the vicinity of V64</td>
<td>44</td>
</tr>
<tr>
<td>1.11</td>
<td>Schematic drawing of the (HMfA)\textsubscript{2} crystal structure in the</td>
<td></td>
</tr>
</tbody>
</table>
region surrounding R64..............................................................46

1.12 Schematic representation of the N terminal proline tetrad in
the (HMfA)_2 crystal structure......................................................49

1.13 Primary sequence alignments of the eukaryal histone fold
dimer pairs based on structure...................................................53

1.14 The eukaryal nucleosome crystal structure showing the N
and C terminal extensions and tails............................................57

1.15 Inter-nucleosomal contacts between the H4 tail and the
(H2A+H2B) dimer in the crystal structure.................................60

2.1 Tricine SDS-PAGE of fractions from a heparin column
used to purify rHMfB-(K13T+R19S+T54K)................................88

2.2 Comparison on CD spectra of rHMfB and rHMfB-(T54K)...........90

2.3 Agarose gel electrophoretic mobility shift assays of
rHMfB variants........................................................................93

2.4 Summary of agarose gel electrophoretic mobility
shift assays of the rHMfB variants.............................................95

2.5 Polyacrylamide gel-shift retardation assays of
complexes formed with wt rHMfB and rHMfB variants..............99

2.6 Summary of polyacrylamide gel retardation assay results
for the rHMfB variants with the 89 bp pBR322 DNA.................101

2.7 Summary of the polyacrylamide gel electrophoresis
retardation assay results for the rHMfB variants
2.8 Polyacrylamide gel electrophoresis retardation assays of complexes formed by wt rHMfB and rHMfB variants with Selex clone 1 DNA ................................................................. 106

2.9 Illustration of the R19S-T54a/R19Sa-T54 interactions ........................................... 108

2.10 Illustration of histone-fold residues in rHMfB predicted to participate in DNA binding ................................................................. 110

2.11 Comparison of the circular dichroism spectra of wt rHMfB and the rHMfB-(R19S+T54K) variant ........................................... 113

2.12 Circular dichroism spectral comparisons of wt rHMfB and the rHMfB-(K13T+R19S+T54K) variant ........................................... 115

2.13 Agarose electrophoretic mobility shift assay using the rHMfB-(R19S+T54K) variant ........................................... 117

2.14 Agarose electrophoretic mobility shift assay using the rHMfB-(K13T+R19S+T54K) variant ........................................... 119

2.15 Polyacrylamide gel electrophoretic retardation assay of complexes formed by the rHMfB-(R19S+T54K) variant ........................................... 121

2.16 Polyacrylamide gel electrophoretic retardation assay of complexes formed by the rHMfB-(K13T+R19S+T54K) variant ........................................... 123

2.17 Polyacrylamide gel electrophoretic retardation assay of
complexes formed by wt rHMfB, rHMfB-(R19S+T54K) and (K13T+R19S+T54K) variants with the Selex 110 bp cycle 8 clone 1 DNA.................................................................125

2.18 Ligase-catalyzed DNA circularization assays in the presence of wt rHMfB and rHMfB variants........................128

2.19 Circular dichroism assays of rHMfB and rHMfB variants binding to pUC19 DNA..............................................131

2.20 Copper phenanthroline crosslinking of rHMfB variants...............................................................................133

2.21 Illustration of rHMfB residues identified as being involved in DNA binding.................................................136

2.22 Illustration identifying the positions of residues that were changed to cysteines in rHMfB variants..................146

3.1 Tricine SDS-PAGE of fractions from a heparin column used to purify rHMfB-(E18K+G51K)..................................158

3.2 Circular dichroism spectra of wt rHMfB and the rHMfB-E18K variant.............................................................160

3.3 Circular dichroism spectra of wt rHMfB and the rHMfB-G51K variant.............................................................162

3.4 Circular dichroism spectra of wt rHMfB and the rHMfB-(E18K+G51K) variant.................................................164

3.5 Agarose gel EMSA of complexes formed by the rHMfB-E18K, G51K and (E18K+G51K) variants.................166
3.6 Polyacrylamide EMSA of complexes formed by the 
rHMfB-E18K, G51K and (E18K+G51K) variants ...............168

3.7 Polyacrylamide EMSA of complexes formed by 
the rHMfB-E18K, G51K and (E18K+G51K) variants 
with the 110 bp Selex cycle 8 clone 1 DNA ..................170

3.8 Circular dichroism assays of the rHMfB-G51K 
variant binding to pUC19 DNA ......................173

3.9 Circular dichroism assays of the rHMfB-(E18K+G51K) 
variant binding to pUC19 DNA ..............175

3.10 Decrease in the ellipticity of pUC19 DNA at 275 nm 
resulting from addition of wt rHMfB, G51K and 
(E18K+G51K) variants .............................................177

3.11 Ligase-catalyzed DNA circularization assays of DNA binding 
by wt rHMfB and the rHMfB-E18K, G51K and 
(E18K+G51K) variants .............................................179

3.12 Topology assays of DNA complexes formed by wt rHMfB and 
The rHMfB-E18K variant .............................................182

3.13 Topology assays of the DNA in complexes formed by the 
rHMfB-G51K and (E18K+G51K) variants .....................184

3.14 Illustration of the clusters of positively-charged residues in the 
L1 and L2 sites in the rHMfB-(E18K+G51K) variant 
tetramer ..........................................................188
3.15 Why does DNA binding by the rHMfB-(E18K+G51K) variant prevent ligase-catalyzed DNA circularization? ...............................192

4.1 Sensorgram of streptavidin crosslinking to the sensor surface............................................................................................204

4.2 Sensorgram showing the immobilization of the 5' biotinylated 93 bp Selex clone 20 DNA to a streptavidin-activated sensor surface.............................................................................................206

4.3 Superimposed sensorgrams of archaeal histone interactions with the immobilized, biotinylated 93 bp Selex clone 20 DNA..............213

4.4 Superimposed sensorgrams of DNA binding by wt rHMfB and the rHMfB-G51K and (E18K+G51K) variants.........................216

4.5 Affinity analysis of wt rHMfA binding to a 93 bp Selex cycle 8 clone 20 DNA.....................................................................218

4.6 Affinity analysis of wt rHMfB binding to a 93 bp Selex cycle 8 clone 20 DNA..............................................................220

4.7 Affinity analysis of wt rHMfA binding to a 93 bp fragment of pLITMUS 28 polylinker.........................................................222

4.8 Affinity analysis of wt rHMfB binding to a 93 bp fragment of pLITMUS 28 polylinker..............................................................224

4.9 Titration of the energy released by sequential injection of the 80 bp Selex clone 20 DNA into a solution of wt rHMfB ..............228

4.10 Isothermal calorimetric scan of the titration of the 80 bp Selex xix
clone 20 DNA into a solution containing the rHMfB-(E18K+G51K) variant..............230
<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Oligonucleotide sequences used to generate mutations in rHMfB</td>
<td>68</td>
</tr>
<tr>
<td>2.2</td>
<td>Salt conditions used to purify rHMfB variants by heparin affinity chromatography</td>
<td>76</td>
</tr>
<tr>
<td>2.3</td>
<td>Folded α-helical contents of rHMfB variants and list of variants that were not produced in <em>E. coli</em></td>
<td>85</td>
</tr>
<tr>
<td>4.1</td>
<td>DNA sequences used in the different assays with archaeal histones</td>
<td>201</td>
</tr>
<tr>
<td>4.2</td>
<td>Affinity dissociation constants of archaeal histones binding to different DNA molecules</td>
<td>225</td>
</tr>
</tbody>
</table>
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>amp</td>
<td>ampicillin</td>
</tr>
<tr>
<td>bp</td>
<td>base pair(s)</td>
</tr>
<tr>
<td>cbt</td>
<td>chicken blood topoisomerase I</td>
</tr>
<tr>
<td>CD</td>
<td>circular dichroism</td>
</tr>
<tr>
<td>Cpm</td>
<td>counts per minute</td>
</tr>
<tr>
<td>C-terminus</td>
<td>carboxyl terminus of the protein</td>
</tr>
<tr>
<td>Da</td>
<td>dalton</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNase</td>
<td>deoxyribonuclease</td>
</tr>
<tr>
<td>ds</td>
<td>double-stranded</td>
</tr>
<tr>
<td>dNTPs</td>
<td>equimolar mixture of dATP, dGTP, dCTP, dTTP</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>disodium ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EM</td>
<td>electron microscopy</td>
</tr>
<tr>
<td>EMSA</td>
<td>electrophoretic mobility shift assay</td>
</tr>
<tr>
<td>5X GLB</td>
<td>gel-loading buffer at 5-fold concentration</td>
</tr>
<tr>
<td>h</td>
<td>hour(s)</td>
</tr>
<tr>
<td>ITC</td>
<td>isothermal titration calorimetry</td>
</tr>
<tr>
<td>K-citrate</td>
<td>potassium citrate buffer</td>
</tr>
<tr>
<td>L</td>
<td>liter</td>
</tr>
<tr>
<td>min</td>
<td>minute(s)</td>
</tr>
<tr>
<td>MW</td>
<td>molecular weight</td>
</tr>
<tr>
<td>NLS</td>
<td>nucleosome-like structures</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>N-terminus</td>
<td>amino terminus of the protein</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>r</td>
<td>recombinant</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>rpm</td>
<td>rotations per minute</td>
</tr>
<tr>
<td>s</td>
<td>second(s)</td>
</tr>
<tr>
<td>ss</td>
<td>single-stranded</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>%T, %C</td>
<td>%T (total concentration of acrylamide and N,N'-methylenebisacrylamide)</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>Tm</td>
<td>melting temperature</td>
</tr>
<tr>
<td>tricine</td>
<td>N-[tris(hydroxymethyl)glycine]</td>
</tr>
<tr>
<td>Tris</td>
<td>tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>U</td>
<td>units of enzyme activity</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>Vol</td>
<td>volume</td>
</tr>
<tr>
<td>w/v</td>
<td>weight/volume</td>
</tr>
<tr>
<td>w/w</td>
<td>weight/weight</td>
</tr>
</tbody>
</table>

%C (percentage of total acrylamide concentration (T) that is \( N,N' \)-methylenebisacrylamide)
CHAPTER 1

GENERAL INTRODUCTION

Genome Compaction

Comparisons of rRNA sequences have shown that Archaea, Bacteria and Eukarya comprise the three domains of life (Woese et al., 1990). While the Eukarya have their genomes organized into a large defined nucleus that is separated from the cytoplasm by a nuclear membrane, the prokaryotic domains namely the Archaea and Bacteria which have relatively smaller genomes, lack a nuclear membrane and have their genomic DNA contained in an irregular-shaped nucleoid region (Sandman et al., 1998). Despite these differences in the degrees of nuclear organization, all cells face a common packaging problem of containing and organizing long DNA molecules within the physical limits of the cell. DNA organization helps prevent spontaneous DNA aggregation at high concentrations and also overcomes the electrostatic repulsions of the negatively charged DNA phosphate backbone. But DNA is also a dynamic structure and hence a balance must be maintained between confining the DNA within cellular limits and allowing its accessibility for DNA replication, repair and transcription.

Cells have evolved several solutions to this packaging problem including the use of architectural chromosomal proteins. Histones are one class of these proteins that exist in Archaeal and Eukaryal kingdoms but don’t have true bacterial homologs (Figure 1.1).
Figure 1.1. Phylogenetic tree for members of the Eukaryarchaeota. This tree is based on comparisons of small subunit rRNA sequences (Olsen et al., 1994). Organisms that are known to contain histones are marked with an asterisk. While the branching order is represented accurately, the branch lengths do not correspond to exact evolutionary distances. This figure has been adapted from Sandman et al., 1998.
Figure 1.1.
In Eukarya, histones package DNA into nucleosomes that are linearly arranged to form chromatin fibers and eventually condensed into chromosomes. Genome compaction in Archaea and Bacteria, on the other hand, is relatively unclear. Topoisomerases are thought to induce DNA supercoiling and positively charged ions and small compounds like polyamines might contribute to DNA compaction (Worcel and Burge, 1972, Pettijohn, 1988). Prokaryotes also contain an abundance of "histone-like" proteins that share size, charge and sequence similarities with histones but do not have the conserved tertiary structural motif called the "histone fold" (Figure 1.2) and, furthermore, form complexes with DNA that are different from nucleosomes (Schmid, 1990, Reeve et al, 1997). The only true prokaryotic homologs of the eukaryal histones are in the HMf family in Euryarchaeota (Figure 1.1).

Eukaryal nucleosome structure

Greater than 80% of the eukaryal genomic DNA is packaged into nucleosomes which are found every 200 +/- 40 bp in all eukaryotic genomes (Noll, 1974, McGhee and Felsenfield, 1980). Chromatin appears as uniformly structured "beads on a string" when viewed by electron microscopy (Olins and Olins, 1974). Nucleosomes are made up of nucleosome core particles, stabilized by linker histones like H1, H5 or variants. The crystal structure of the nucleosome core particle, resolved to 2.8 Å, shows ~146 bp of DNA wrapped in 1.65 negative toroidal supercoils around the outside of a protein core, containing 2 copies each of the 4 different histone molecules H3, H4, H2A and H2B and a total molecular mass of 206 kDa (Luger et al., 1997) (Figure 1.3). This confirmed the previous crystal structure of the histone octamer, resolved to 3.1 Å, that showed a
Figure 1.2. The histone fold structure. Eukaryal core and archaeal histones both contain the histone fold motif (Arends et al., 1991) that comprises 3 α helices (α1, α2 and α3), separated by β-sheet loop structures (L1 and L2). The two monomers interact hydrophobically in a head-to-tail ‘handshake’ motif, crossing between the α2 helices. The N termini contact each other on one end of the dimer but the C termini do not interact on the opposite end (Starich et al., 1996, Luger et al., 1997). This schematic represents the (HMfA)2 dimer crystal structure that does not contain the long N and C terminal extensions and tails found in the eukaryal histones. This figure has been adapted from Decanniere et al., 2000.
Figure 1.2.
Figure 1.3. Crystal structure of the nucleosome core particle. The image on the left shows the view down the superhelical axis while the figure on the right shows a perpendicular view to it. The 4 core histones H2A (yellow), H2B (red), H3 (blue) and H4 (green) are shown surrounded by 146 bp of DNA (turquoise and brown). A 4-helix bundle, formed by the interaction of the α2 and α3 helices of the two H3 molecules and aligned with the center of the DNA, is shown on the top. This figure has been adapted from Luger et al., 1997.
tripartite wedge-shaped arrangement (Arents et al., 1991). These 4 core histones are small (11-16 kDa), basic proteins that exhibit low sequence homology but share a highly conserved tertiary structure, a helix-loop-helix, called the "histone fold" (Figure 1.2). The histone fold comprises 3 α-helices separated by 2 loops that have β-sheet structures (Arents et al., 1991). N and C terminal extensions and tails that flank the histone folds of the eukaryal core histones make up 45% of the mass of the nucleosome core, and might be involved in higher order structure formation and histone post-translational modifications like phosphorylation, methylation, acetylation and ADP-ribosylation for gene regulation (Luger et al., 1997, Luger and Richmond, 1998, Wolffe, 1994a and 1994b, Wolffe, 1995) (Figure 1.4). However, studies with H3 and H4 have shown that the tails can be removed without loss of nucleosomal positioning or integrity (Pruss et al., 1995).

In solution, each histone monomer interacts with its heterologous partner to form the head-to-tail dimers like (H2A+H2B) and (H3+H4) through extensive hydrophobic interactions described as a "handshake motif". Two (H3+H4) dimers associate to form a tetramer. Core histones only form heterodimers in the specific pairings of (H3+H4) and (H2A+H2B). The major dimerization interface is the long central α2 helix. The (H3+H4) tetramer is stabilized by a 4-helix bundle, formed by interactions between the C-terminal regions of α2, and α3 helices of the two H3 molecules. In the presence of DNA or high salt concentration, these dimers assemble into an octamer composed of a central (H3+H4) tetramer flanked, on each side, by an (H2A+H2B) dimer. The (H2A+H2B) dimers bind to the (H3+H4) tetramer through a structurally homologous 4-
Figure 1.4. Alignment of the 4 eukaryal core histones based on primary sequence and structure. The α helices are underlined and the specific location of each α helix and loop region is indicated on the top. The sites at which arginine residues penetrate the DNA minor groove at the indicated SHL locations are marked by ^, while the lines above the residues show the hydrogen bonds between L1 and L2 in a β-sheet interaction. The overlying lines show buried arginine-aspartate salt bridges in the (H3 + H4) pairs and interrupted lines indicate the absence of this ionic interaction in the (H2A + H2B) pairs. Residues highlighted in blue represent the H3-H3' 4 helix bundle, while amino acids highlighted in brown show the H4-H2B 4 helix bundle. Contact sites between α2-α2 helices in (H3+H4) and (H2A+H2B) are shown in blue-green and orange, respectively. The N and C terminal extensions and tails are shown at the bottom of the figure and are italicized. The positions of residues that were not clearly defined in the nucleosome crystal structure are shown in gray. The DNA contact sites are labeled by superhelix location (SHL) above the sequences. Sites of ubiquitination are marked by a "U", acetylation sites in vivo are denoted by "a", and positions of trypsin cleavage that remove the tails of the histone core particles are marked by "T". The region of H4 that makes an inter-nucleosomal contact with the (H2A+H2B) dimer is underlined in orange. This figure was adapted from Luger et al., 1997.
Figure 1.4.
helix bundle, formed by interaction between the α2 and α3 helices of H2B and H4 (Arents et al., 1991, Arents and Moudrinakis, 1993, Luger et al., 1997) (Figure 1.3).

The histone tetramer plays a vital role in nucleosome organization and function. The (H3+H4) tetramer protects ~73 bp from micrococcal nuclease digestion and wraps DNA in negative or positive supercoils (Hayes et al., 1991, Hamiche et al., 1996). Orientation of the dimer-dimer interface appears to determine the direction of supercoiling (Hamiche and Richard-Foy, 1998). However, addition of (H2A+H2B) dimers in the eukaryal nucleosome restricts the tetramer to negative DNA supercoiling (Hamiche et al., 1996). Additionally, the (H3+H4) tetramer determines nucleosome positioning and initiates nucleosome assembly (Hayes et al., 1991, Dong and Van Holde, 1991).

The DNA helix is not uniformly bent around the histone core. DNA sequences have varying flexibilities in addition to the bending induced by histones. DNA in eukaryal nucleosomes is overwound (~10.2 bp/turn) compared to free DNA in solution (~10.5 bp/turn) (Luger et al., 1997, Rhodes and Klug, 1981). Intrinsically flexible or curved DNA fragments have been found to be preferentially associated with histones because these sequences are more energetically favored to accommodate to nucleosomal structure than other DNA molecules (Thåström et al., 1999). Eukaryal nucleosomes are found to be preferentially associated with the curved Crithidia fasciculata DNA which contain several repeats of (A)_{4-6} tracts, in phase with the DNA helix (Kitchin et al., 1986). Also, genomic SELEX experiments have shown inherently curved DNA molecules to have the highest affinity for histone octamers (Widlund et al., 1997). Rotational positioning of eukaryal nucleosomes has shown upto a 100-fold higher affinity
for synthetic DNA sequences containing alternating, anisotropically flexible G/C and A/T compressions at ~5 bp intervals corresponding to the major and minor grooves, respectively, than bulk genomic DNA. This is consistent with the face of the DNA helix that is in contact with the histone core being compressed, and creating alternating major and minor groove compressions (Shrader and Crothers, 1989). Tandem CTG repeats are also strong positioning elements for eukaryal nucleosomes (Wang et al., 1995, Godde and Wolfe, 1996). These repeats form unusual structures that have been associated with human neuromuscular diseases (Bacolla et al., 1997).

**Histone-like Proteins in Bacteria**

Prokaryotic cells do not contain histones but have several small, abundant, basic DNA binding proteins that bend and compact DNA but do not share structural similarity with the histones. Furthermore, EM studies of prokaryotic nucleoids do not show repeating arrays of nucleosomes (Pettijohn, 1988). Amongst the most well-characterized proteins in this class are HU, IHF, FIS and H-NS.

**HU**

HU is a prototypic DNA-binding protein in the *Escherichia coli* HU family of proteins and is found in all eubacteria examined to date as well as in plant chloroplasts and in the archaeon *Thermoplasma acidophilum* (Hayat and Mancarella, 1995). In *E. coli*, HU is a small (18 kDa), basic heterodimeric protein in the nucleoid region composed of HU-α and HU-β, each having a molecular mass of 9 kDa (Rouvière-Yaniv, 1978, Rouvière-Yaniv and Kjeldgaard, 1979). The two monomers appear to be ~70%
homologous and form heterodimers in the enteric bacteria like *E. coli* and *Salmonella typhimurium* but homodimers in all other bacteria (Oberto *et al*., 1994). HU is the most abundant DNA binding protein and *E. coli* has ~ 60,000 monomers per cell. Each HU dimer binds ~200 bp of DNA (Drlica and Rouvière-Yaniv, 1987). The crystal structure of the HU homodimer from *Bacillus stearothermophilus* (HUBst) shows a lobster-shaped structure containing a hydrophobic central core and two arms that reach into the sugar-phosphate backbone of the major and minor grooves of DNA and bend it. Each HU monomer contains 3 α-helices, involved in dimerization, and an anti-parallel β-sheet structure that binds DNA (Tanaka *et al*., 1984, White *et al*., 1989). Mutagenesis studies have also shown that ionic interactions in the flexible arms of HU are important in DNA binding (Goshima *et al*., 1990).

HU-DNA interactions have been widely studied *in vitro*. HU was identified in a search for transcription activating factors and binds DNA in a non-sequence specific manner. Electron microscopy studies of HU-DNA complexes, *in vitro*, reveal nucleosome-like structures containing 275 bp of DNA bent around 9 HU dimers, but these complexes were not observed *in vivo* (Rouvière-Yaniv *et al*., 1979, Broyles and Pettijohn, 1986, Pettijohn, 1988). Even though there is enough HU in *E. coli* to compact a large amount of genomic DNA, the digestion patterns of repeating nucleosomal units are absent (Pettijohn, 1988). HU introduces negative supercoils into relaxed plasmid DNA, it circularizes short (99-126 bp) DNA fragments, and preferentially interacts with negatively supercoiled DNA, cruciform or Holliday and stem-loop DNA-containing structures (Rouvière-Yaniv *et al*., 1979, Hodges-Garcia *et al*., 1989, Shindo *et al*., 1992, Pontiggia *et al*., 1993, Bonnefoy *et al*., 1994).
HU acts as a stimulatory accessory protein in the initiation of DNA replication in 
also enhances Lac repressor binding to operator sequences and CAP binding to upstream 
regulatory regions which involve DNA bending and looping (Pettijohn, 1988, Flashner 
and Gralla, 1988). HU can also act as both an activator or repressor of transcription, 
depending on the template used (Drlica and Rouvière-Yaniv, 1987). Mutations in either 
hupA (encodes HU-α) or hupb (encodes HU-β) in *E. coli* function as wild type strains 
presumably as they do in organisms that have only one gene (Storts and Markovitz, 1988, 
Pettijohn, 1988). However, mutations or deletions in both genes result in cells that are 
still viable but defective in DNA replication, cell division, Hin-mediated site-specific 
recombination and transposition of bacteriophage Mu (Husimann *et al.*, 1989, Jaffe *et al.*, 
1997, Paull *et al.*, 1993). Transposition of hupA and hupB are autoregulated by internal 
levels of HU-α and HU-β, and HU might also control synthesis of other proteins like 
IHF. Therefore it is thought that HU, as a sequence-independent DNA binding protein, is 
localized and acts cooperatively with other sequence-specific DNA binding proteins in 
gene regulation (Pettijohn, 1988).

**IHF**

The integration host factor, IHF, is a small (~20 kDa), basic protein that was 
originally identified as a necessary host factor for integrative recombination of the 
bacteriophage lambda and shares substantial primary sequence homology with the HU 
protein but is much less abundant (Miller and Friedman, 1980, Rice, 1997). In *E. coli*, 
IHF is a heterodimer composed of an 11 kDa subunit IHF-α (encoded by himA) and a 9
kDa IHF-β subunit (encoded by himD) that binds sequence-specifically to DNA in the minor groove and induces some of the sharpest bends seen with DNA binding proteins to date between 140-180° (Craig and Nash, 1984, Yang and Nash, 1989, Thompson and Landy, 1988, Sun et al., 1996). The co-crystal of the E. coli IHF-DNA complex shows structural homology with the B. stearothermophilus HU homodimer with the exception that the IHF-DNA contacting arms are more ordered (Rice et al., 1996, Tanaka et al., 1984, Ellenberger and Landy, 1997). Like HU, IHF also introduces negative supercoils into the DNA (Lopez-Garcia et al., 1998). However, unlike HU, IHF binds to specific sites in promoters and to phage attachment (att) sites, namely a 5 bp AT-rich element followed by two consensus binding sites, located one helical turn away, namely TATCAA and TTG. The protein core makes extensive DNA-phosphate contacts while conserved proline residues, in each arm of the protein, intercalate between DNA bases and cause sharp bends which reverse the direction of the DNA (Rice et al., 1996). This sequence dependency of IHF binding that is dictated by the DNA structural constraints of the minor groove, and not the specific DNA bases, is called “indirect readout” (Rice, 1997).

IHF is referred to as a “big bender” because it plays several roles in processes that involve shaping of DNA like lambda site-specific recombination. However, it is interesting that HU or DNA, bent at the IHF consensus site, can substitute for IHF in lambda recombination, demonstrating that these proteins might not act alone. Additionally, IHF serves as a transcriptional activator in the control of gene expression, it is involved in initiation of DNA replication and partitioning of replicated DNA into daughter cells, and also enhances transposition (Ellenberger and Landy, 1997).
FIS

The factor for inversion stimulation, FIS, is a small, abundant protein in the *E. coli* nucleoid that binds to bent DNA (Murphy and Zimmerman, 1997). Unlike IHF, however, FIS binds DNA as a homodimer (MW ~ 25 kDa) with some specificity for a weak consensus sequence containing a helix-turn-helix motif, namely (G/T)NN(T/C)(A/G)NN(A/T)NN(T/C)(A/G)NN(C/A), and bends the DNA ~ 90° (Thompson and Landy, 1988, Finkel and Johnson, 1992). FIS does not belong to the HU family of proteins but shares primary sequence homology with NtrC, a DNA-binding protein that activates transcription of σ54-dependent promoters (Johnson et al., 1988). FIS is also involved in its own transcriptional autoregulation as well as transcriptional activation of rRNA and tRNA genes (Ninnemann et al., 1992, Ross et al., 1990, Nilsson et al., 1990). In addition, FIS stimulates DNA inversion catalyzed by various recombinases and DNA replication (Sandmann et al., 1996).

H-NS

The histone-like nucleoid structuring protein, H-NS (also known as H1a), is a small, abundant DNA binding protein in the enterobacteria. In *E. coli*, it binds and compacts DNA as a homodimer of ~15.5 kDa monomer subunits in 3 isoforms of different isoelectric points (Spasksy et al., 1984). H-NS binds cooperatively and with specificity to bent, A/T rich fragment and non-specifically to DNA fragments that are completely saturated with H-NS (Tupper et al., 1994, Lucht et al., 1994). However, non-specific binding might not be very significant *in vivo* as the intracellular amount of H-NS
protein matches the number of specific binding sites. H-NS modulates transcription in the intestinal environment and is under autoregulatory control (Atlung and Ingmer, 1997).

Nucleoid proteins in Archaea

Primary sequence alignments of the archaeal nucleoid proteins divide them into 4 groups: HTa from *Thermoplasma acidophilum*, the MCI family from *Methanosarcinaceae*, Sac and Sso DNA binding proteins from *Sulfolobus* species and the HMf family of archaeal histones. Except for HTa which is a homolog of the bacterial HU protein, the archaeal nucleoid proteins are not related to each other or to other bacterial DNA binding proteins (Grayling et al., 1994, Sandman et al., 1998).

HTa

Primary sequence and structural comparisons suggest that HTa, a histone-like protein from *T. acidophilum* (optimum growth temperature 59°C and pH 1-2) is homologous to the HU family of bacterial nucleoid proteins (DeLange et al., 1981). HTa is an approximately 9 kDa abundant chromosomal protein that is composed of ~ 23% of basic amino acids, protects DNA from thermal denaturation and increases its melting temperature by ~ 40°C (Searcy, 1975, DeLange and Williams, 1981, DeLange et al., 1981). Five phenylalanine residues in the hydrophobic core of HTa might contribute to the thermostability of this protein (Searcy, 1986). HTa assembles as a tetrameric complex on a 40 bp DNA fragment and forms nucleosome-like structures *in vitro* and *in vivo*. 
although repeated arrays have not been observed in vivo (Searcy and Stein, 1980, Bohrmann et al., 1990).

MC1

MC1 is an abundant nucleoid proteins in the mesophilic *M. barkeri* (optimal growth temperature 35°C) and is found in many other thermophilic *Methanosarcinaceae* sp. (Chartier et al., 1985, Chartier et al., 1989, Derkacheva and Kagramanova. 1994). These ~ 11 kDa proteins comprise up to 90% of all DNA-binding proteins in these species but only account for 14% of DNA-associated proteins in *Methanosarcina* CHTI 55 (Chartier et al., 1988). MC1 contains 93 amino acids with a large number of basic and acidic residues but no hydrophobic domain (Chartier et al., 1989). One MC1 protein molecule binds ~11 bp of DNA and this interaction is not cooperative (Culard et al., 1993). MC1 preferentially interacts with curved A/T rich DNA as well as negatively supercoiled DNA and protects dsDNA from thermal denaturation (Teyssier et al., 1996, Chartier et al., 1988). DNA cyclization studies show that MC1 induces large kinks in DNA (Laine et al., 1991). EM studies also reveal sharp bends in the DNA complexed with MC1 but nucleosome-like structures are absent (Le Cam et al., 1999). Moreover, MC1 does not protect DNA from micrococcal nuclease digestion and does not induce DNA wrapping like histones (Laine et al., 1991, Toulmé et al., 1995).

**DNA binding proteins from Sulfolobus species**

Small, abundant, basic soluble DNA binding proteins have been found in the hyperthermophilic sulfur-dependent *Sulfolobus* species and have been categorized by
their molecular weight. In *Sulfolobus acidocaldarius* and *S. solfataricus*, proteins of 7, 8 and 10 kDa have been isolated. *S. acidocaldarius* has 5 proteins in the 7 kDa group namely 7a, b, c, d and e, and 2 each in the 8 and 10 kDa group namely 8a and b and 10a and b, respectively (Grote et al., 1996, Derkacheva and Kagramanova, 1994). In *S. solfataricus*, a single major 7 kDa nucleoid protein exists named Sso7d that is homologous but more basic than Sac7d, and in *S. shibatae*, Ssh7 has been isolated that bears resemblance to this group (Lopez-Garcia et al., 1998, Mai et al., 1998).

Sac7d binds non-specifically with micromolar affinity to DNA and increases its melting temperature by ~ 40°C (McAfee et al., 1996). Hence, it is believed to protect double stranded DNA from thermal denaturation and degradation, in vivo, at growth temperatures of ~ 80°C (Dijk and Reinhardt, 1986). The tertiary structure of Sac7d shows an incomplete 5-stranded β-barrel and a short C-terminal amphipathic α-helix (Edmondson et al., 1995). Both Sac7d and Sso7d unfold at temperatures > 90°C at pH 7.5, and are stable at pH 0 with Tm > 60°C (McAfee et al., 1995). Small angle X-ray scattering studies and the crystal structure of the Sac7d-DNA complex show Sac7d binding in the minor groove that generates a large kink in the DNA of ~ 70° by intercalation of hydrophobic side chains between DNA bases. However, the multiple Sac7d-induced bends do not produce DNA supercoiling but result in tilting of the bases (Robinson et al., 1998, Kreuger et al., 1999). Sso7d also binds to DNA in a sequence-independent manner and in the micromolar affinity range, introduces negative supercoils into the DNA and increases the melting temperature of DNA by 39°C (Lundback et al., 1996, Baumann et al., 1994). NMR spectroscopy shows Sso7d as a globular structure with a triple-stranded anti-parallel β-sheet (β3, β4, β5), packed against an orthogonal
double stranded β sheet (β1 and β2), and a small C-terminal α-helix. Sso7d is
homologous to the SH3 domains in eukaryal signal transduction proteins (Baumann et
al., 1994). The triple-stranded β-sheet spans the minor groove and makes non-specific
electrostatic and hydrogen bonds with DNA phosphates and hydrophobic interactions
with deoxyribose groups (Agback et al., 1998). The crystal structure of the Sso7d protein
complexed with DNA shows that binding in the minor groove introduces a sharp bend of
60° in the DNA due to intercalation of hydrophobic amino acids into DNA bases (Gao et
al., 1998). Furthermore, both Sso7d and Sac7d bind to T-G mismatched base pairs in the
minor groove intercalation site (Su et al., 2000).

Another group of acid soluble proteins are found in S. acidocaldaricus DSM 639
that are helix stabilizing proteins (HSNP-A, HSNP-C, HSNP-C') or DNA binding
nucleoid proteins (DBNP-B) (Reddy and Suryanarayana, 1988). HSNP-C' is identical to
Sac7d from S. acidocaldaricus (Celestina and Suryanarayana, 1995). The helix
stabilizing proteins bind non-cooperatively to double-stranded DNA and increase the
melting temperature by 25°C while DBNP-B proteins have a higher affinity for single-
stranded DNA and, furthermore, unstack dsDNA (Reddy and Suryanarayana, 1988,

Recently, a new DNA-binding protein has been identified in S. solfataricus called
Smj12 that is highly thermostable and basic, binds non-specifically to dsDNA and
protects it from thermal denaturation. However, unlike the other Sulfolobus proteins
categorized to date, Smj12 is not an abundant protein, it does not bend DNA, and
induces positive supercoiling in the DNA. Hence, it is not believed to organize DNA like
histones (Napoli et al., 2001).
**HMf family of archaeal nucleosomes**

Amino acid sequence alignments of the HMf family of archaeal histones with the eukaryal core histones reveal primary sequence homologies of ~45%. In fact, the HMf family of archaeal histones are more similar to eukaryal core histones than the eukaryal histones are to each other (Grayling et al., 1996). Hence, the archaeal histones and eukaryal histones are believed to have evolved from a common ancestor that must have existed before the divergence of the Eukarya and Archaea (Sandman et al., 1994). To date, approximately 30 archaeal histones have been identified that all share the conserved histone fold motif and lack N and C terminal extensions and tails of eukaryal histones (www.biosci.ohio-state.edu/microbio/ArchaealHistones/Alignments/alignments.html). (Figure 1.5). Least-squares superimposition of the histone folds of the most well-characterized archaeal histone HMfB (histone from the hyperthermophilic methanogen *Methanothermus fervidus*) homodimer and the (H3 + H4) heterodimer show deviation in α carbon atoms of 2.1 Å. In fact, (H2A + H2B) and (H3 + H4) histone folds also deviate by 2.1 Å (Luger and Richmond, 1998a).

HMf was the first archaeal histone identified and is the most well-characterized archaeal histone (Sandman et al., 1990). The genome of *M. fervidus* is 33% mol G+C and contains ~980 mM K+ and 300 mM 2',3' cyclic diphosphoglycerate (K2cDPG) which protects enzymes from heat denaturation (Stetter et al., 1981, Hensel and König, 1988). HMf contains homodimers and heterodimers of small (~7.5 kDa), basic (pI~9-10) polypeptides HMfA and HMfB that are 84% identical (they differ by 11 amino acids) (Figure 1.5) are localized in the *M. fervidus* nucleoid and bind to DNA in a
Figure 1.5. Primary sequence alignments of the archaeal histones with the 4 eukaryal core histone consensus sequences. The positions of the α-helices and loop regions are indicated above the sequences, determined by solution and crystal structures of HMfB and HMfA homodimers (Starich et al., 1996, Decanniere et al., 2000) and the eukaryal core histone structures (Luger et al., 1997). The numbers in parentheses, preceding and following the histone fold regions of the eukaryal histones, indicate the number of amino acids that were omitted in this alignment. Gaps in the sequence, indicated by hyphens, were introduced to improve alignments. Stop codons in the sequences are denoted by asterisks and vertical lines indicate conserved hydrophobic residues in monomer-monomer interactions. The archaeal histone gene ORFs shown are: HMf from *Methanothermus fervidus* (Krzycki et al., 1990, Sandman et al., 1990); MT from *Methanobacterium thermoaototrophicum* (Tabassum et al., 1992, Smith et al., 1997); HFo from *Methanobacterium formicicum* (Darcy et al., 1990); HTz from *Thermococcus zilligii* (Rominus and Musgrave, 1996); HPy from *Pyrococcus* strain GB3a (Sandman et al., 1994b, Soares et al., 1998); PF from *Pyrococcus furiosus* (Robb et al., 2001); PH from *Pyrococcus horikoshii* (Kawarabayashi et al., 1998); HPk from *Pyrococcus kodakaraensis* (Higashibata et al., 1999); PAB from *Pyrococcus abyssi* (www.genoscope.cns.fr/cgi-bin/Pab.cgi); AF from *Archaeoglobus fulgidus* (Klenk et al., 1998); MJ from *Methanococcus jannaschii* (Bult et al., 1996, Li et al., 1999b); HMv from *Methanococcus voltae* (Agha-Amiri and Klein, 1993); Mka (N and C terminal domains) from *Methanopyrus kandleri* (Slesarev et al., 1998) and HHb (N and C terminal domains) from *Halobacterium NRC-1* (Ng et al., 2000). This figure has been adapted from Sandman et al., 1998 and (www.biosci.ohio-state.edu/microbio/Archaealhistones/Alignments/alignments.html)
Fig. 1.5.
non-sequence specific manner (Sandman et al., 1990, Bohrmann et al., 1994, Grayling et al., 1997). The NMR structure of rHMfB and the high resolution crystal structures of rHMfA and rHMfB homodimers have confirmed that these proteins contain the conserved tertiary "histone fold" motif (Starich et al., 1996, Decanniere et al., 2000) (Figure 1.2). Electron microscopy shows that HMf binds to DNA, in vivo, to form nucleosome-like structures that resemble "beads on a string" and these proteins decrease the length of the bound DNA (Sandman et al., 1990, Howard et al., 1992). HMfA and HMfB have been overexpressed in E. coli and the recombinant proteins have been found to be identical to their native counterparts in M. fervidus (Sandman et al., 1995). Growth-phase dependent differences occur in the amounts of native HMfA and HMfB in M. fervidus, which suggest different roles of these proteins for genome compaction in vivo. HMfA predominates during exponential growth phase (70-80% of total HMf protein), where rapid access to genomic information is needed, while the amount of HMfB synthesized increases to 50% in stationary phase. Consistent with these observations is the fact that rHMfB forms more compact complexes with linearized DNA > 2 Kb than rHMfA in agarose electrophoretic mobility shift assays and requires a higher protein-DNA ratio than rHMfA for DNA binding (Sandman et al., 1994). Since HMf can form both homodimers and heterodimers, different combinations of these dimers could function in gene regulation in M. fervidus. Similarly, all the archaea which are known to contain histones, to date, also have more than one histone and hence can fine-tune genome compaction with gene regulation. For example Methanococcus jannaschii has 5 histones which can form 15 different dimers (Bult et al., 1996) (Figure 1.5).
HMf nucleosomes are constrained in 1.5 positive toroidal supercoils. Topology experiments with HMf and relaxed circular DNAs have shown that negative supercoils are introduced into the DNA at a low histone to DNA ratio, while positive supercoiling occurs at higher ratios. The switch from negative to positive supercoiling occurs when one dimer reorients itself with respect to the second dimer in a tetrameric complex (Reeve et al., 1997).

Gel filtration chromatography and covalent crosslinking results indicate that HMf molecules are dimers in solution but they form tetramers when bound to DNA (Grayling et al., 1995, Grayling et al., 1996). Micrococcal nuclease digestion shows that HMf protects a 60 bp DNA molecule and immunoprecipitation of complexes, in vivo, indicate that almost all of the *M. fervidus* genome might be packaged by HMf tetramers (Grayling et al., 1997, Pereira et al., 1997,). Polyacrylamide EMSA experiments have shown that a minimum DNA length of 85 bp is needed to form HMfB nucleosomes in vitro. Also, a minimum length of 88 bp DNA was required to generate a covalently closed monomer circle with HMfB, following treatment with DNA ligase. Topology analysis of the ligated product showed a positively supercoiled DNA, and HMfB forms nucleosomes with the circular DNAs at a lower histone to DNA ratio than with linear DNA of the same length (Bailey et al., 1999). Furthermore, stoichiometry experiments with ¹²⁵I labeled rHMfB-YY (I131Y and M35Y rHMfB variant) have shown that a 100 bp of DNA can accommodate a tetrameric protein core (Bailey et al., 1999).

HMfA and HMfB also respond to certain positioning signals and preferentially assemble on specific DNA sequences. Immunoprecipitation of *in vivo* HMf-DNA complexes shows preferential assembly of HMf with 7S rRNA encoding sequences and
these complexes are translationally and rotationally positioned \textit{in vitro} (Pereira \textit{et al.}, 1997, Pereira and Reeve, 1999). Translational positioning is also seen with the CTG repeated DNA sequences and HMf \textit{in vitro}. It is interesting to note that CTG repeated elements are strong positioning signals for eukaryal nucleosomal assembly (Wang \textit{et al.}, 1995, Godde and Wolffe, 1996) but these sequences are not found in archaeal genomes (Bailey and Reeve, 1999). HMf also associates with \textit{C. fasiculata} highly curved DNA fragments that have 4 or 6 adenine tracts, repeated in phase with the DNA helix (Howard \textit{et al.}, 1992). Artificial DNA sequences with 5'-TTTAAAGCCG-3', that have these TA stretches repeated every 10 bp, are some of the highest affinity binding sequences seen with HMf to date (Bailey and Reeve, 1999, K. Bailey and J.N. Reeve, unpublished results). These DNA molecules can better accommodate the HMf core as the T/A and G/C regions are placed in minor and major groove compression sites, respectively.

Similarly, higher affinity sequences, selected from a random-sequence synthetic library of \(10^{14}\) DNA molecules that bound to HMfB using a SELEX approach, have patterns of alternating A/T and G/C dinucleotides every 5 bp. Furthermore, these sequences are abundant in the genomes of the histone-containing Euryarchaeota and Eukarya but not in the Crenarchaeota and Bacteria (Bailey \textit{et al.}, 2000). Since archaeal histones are most homologous to the eukaryal H4, and HMf nucleosomes appear to share several features with the (H3 + H4) tetramer such as the length of DNA protected from micrococcal nuclease digestion and circularized by ligase, the nature of supercoiling and positioning, it is believed that the archaeal nucleosome is the predecessor of the eukaryal tetramer (Sandman \textit{et al.}, 1998, Sandman and Reeve, 2000).
The histone fold structure of eukaryal nucleosomes

The 4 eukaryal nucleosome core histones H3, H4, H2A and H2B have almost identical histone fold structures comprising two short 3-turn α-helices (α1 and α3) and a long 8-turn central α-helix (α2), separated by β-strand loops (L1 and L2) arranged as α1-L1-α2-L2-α3. Dimerization is required for histone fold stability, formed by extensive hydrophobic interactions between the two α2 helices (Arents and Moudrianakis, 1995, Luger et al., 1997). The anti-parallel orientation of α2-α2a in eukaryal histone dimers places L1 in parallel position with L2a and L2 with L1a (the other partner in the dimer is denoted as “a”) (Figure 1.2). However, subtle variations in the L1-L2 structures in the eukaryal histone heterodimers result in pseudo-symmetry for these dimers with the symmetry axis crossing between the two α2 helices (Luger et al., 1997). The offset in α2-α2a interaction also places the N-terminal regions of α1 and α1a in contact with each other but α3 and α3a are not in close proximity on the opposite end of the dimer. However, both α1 and α3 helices interact with the central α2 helices. The C terminus of the α2 helix extends further than the N terminus of the α2 helix in the paired monomer and hence the L1-L2 surfaces are tapered at each end of the dimer (Luger et al., 1997) (Figures 1.6 and 1.7).

The exclusive pairing of (H3 + H4) and (H2A + H2B) heterodimers results from differences in histone fold packing and N and C terminal extensions. For example, H4 and H2A have 6 and 7 amino acids in L1 while H3 and H2B have 8 residues, respectively (Figure 1.4). To compensate for longer L1 loops in H3 and H2B, these histones have
**Figure 1.6.** The (H3+H4) histone-fold-DNA interactions. H3 and H4 chains are shown in blue and green, respectively, while 27-28 bp of DNA, arched around the histone fold, is shown in turquoise and brown. A pseudodyad symmetry axis intersects at SHL 1.5. The α1, α2 and α3 helices and L1 and L2 regions are shown. Histone main-chain hydrogen bond contacts with DNA phosphates are indicated in magenta and side-chain hydrophobic interactions, hydrogen bonds and the arginine residues that penetrate the DNA minor groove are also shown. Hydrogen bonds between arginine and threonine pairs are shown in blue-green, while a hydrophobic interaction of H3-L65 with a thymidine base is shown in yellow. This figure has been adapted from Luger *et al.*, 1997.
**Figure 1.7.** The (H2A+H2B) histone fold pair interactions with DNA. The H2A monomer is shown in yellow and its partner H2B is shown in red, while the 27-28 bp of DNA that surrounds the histone fold is shown in turquoise and brown. A pseudodyad axis of symmetry runs through SHL 4.5. The α helices and loops of the histone fold are labeled. Histone side chain hydrogen bonds and hydrophobic interactions with the DNA backbone are shown as well as arginine residues that insert into the DNA minor groove. Hydrogen bonds between main chain histones and DNA phosphates are indicated in magenta. A hydrogen bond between an H2A arginine in L1 and H2B threonine in L2 is shown in blue-green. This figure has been adapted from Luger *et al.*, 1997.
longer α1 helices. This causes the C terminal regions of the α2 helices of H2A and H4 to bend toward the longer H2B and H3 L1 loops, while on the opposite end of the heterodimer the α2 helices of H2B and H3 are straight (Figures 1.6 and 1.7). β-sheet interactions are found in the highly conserved L1 and L2 regions of all 4 eukaryal histones. A parallel β-conformation is formed with 3 hydrogen bonds between H3 amino acids R83, F84 and Q85 in L1 and K79, T80, and V81 in H4 L2 (Figure 1.6). Hydrophobic interactions also occur at the base of the loops between F84 in H3 and V81 in H4. Additionally, a buried salt bridge is found between R78 and D85 in H4. On the other hand, only 1 hydrogen bond is evident in the β-sheets of the other 3 L1-L2 pairs. Specifically, S47 in H4-L1 is hydrogen bonded to I119 in H3-L2. A buried salt bridge is also present between R116 and D123 in H3. A hydrophobic interaction occurs between I46 in H4-L1 and I119 in H3-L2 (Figures 1.4 and 1.6). In H2A, R42 in L1 is hydrogen bonded to S84 in H2B-L2 and in H2B-L1, S52 is H-bonded to I78 in H2A-L2 (Figures 1.4 and 1.7). Hydrophobic interactions are also found between V43 in H3-L1 and I86 in H2B-L2 as well as I51 in H2B-L1 and I78 in H2A-L2. A buried salt link between L2 and α3 helix is absent in H2A and H2B even though the charged residues are conserved in H2B, namely R83 and E90 (Figure 1.4). Additionally, a conserved arginine-threonine pair is hydrogen bonded in 3 out of the 4 histone paired L1-L2 regions namely R83 in L1-H3 and T80 in H4-L2, R45 in H4-L1 and T118 in H3-L2, R43 in H2A-L1 and T85 in H2B-L2 ((Luger et al., 1997, Luger and Richmond, 1998) (Figures 1.4, 1.6 and 1.7).

Mutagenesis studies have shown that L1-L2 regions are important for nucleosomal stability. SWI/SNF complexes act as transcription activators by removing or modifying nucleosomes and allowing TFIID access to the TATA box DNA but the
exact mechanism remains unknown (Côté et al., 1994, Imbalzano et al., 1994, Kwon et al., 1994). Sin (switch independent) mutants bypass the need for these SWI/SNF complexes in transcription activation. In yeast H3, Sin mutants R116H and T118I result in nucleosomes with an increased sensitivity to micrococcal nuclease and the loss of rotational positioning, seen by digestion with DNase I, and these nucleosomes might allow transcriptional activators to bind DNA (Kruger et al., 1995, Kurumizaka et al., 1997). In H3, E105K in α2 helix is also associated with the Sin phenotype but the nucleosomes are almost as stable as wild type H3 (Kurumizaka et al., 1997). In yeast H4, R45C and R45H variants disrupt the β-sheet interactions with H3-T118 and result in nucleosome lability. Furthermore, V43I in H4 has also been implicated with destabilization of nucleosomes, possibly due to the introduction of a bulky hydrophobic isoleucine near the loop regions that disrupts the hydrophobic interaction made with the isoleucine in H3 L2 (Kruger et al., 1995, Wechser et al., 1997). In yeast H4, substituting an isoleucine for threonine at position 82 is lethal, as the variant is defective in nuclear division and chromosome transmission, but addition of A89V acts as an intragenic suppressor (Smith et al., 1996).

The H3'-H3 4 helix bundle, which stabilizes the H3 + H4 tetramer, contains buried interactions between R116 + D123 and H113 + D123 in H3 (Figures 1.3 and 1.4). Additionally, hydrophobic interactions also occur between H3 C110, H113, L126 and I130 residues. The structurally homologous H2B-H4 4 helix bundle, which has an r.m.s.d. value of 1.85 Å when its α-carbon atoms are superimposed on H3'-H3, shows a hydrogen bond between H75 in H4 and H2B-E90. Also, hydrophobic interactions occur between Y72 and Y88 in H4 and Y80 in H2B (Luger et al., 1997) (Figure 1.4).
The histone fold structure of rHMfB

The rHMfB monomer, which extends 69 residues, contains α1 helix extending from I5 to A15 in 3 turns, followed by a central α2 helix, extending from D22 to A50 in 8 turns, and finally an α3 helix extending from A57 to R66 in 3 turns (Figure 1.2). The α1 and α2 helices are separated by L1 (G16 to S21), and L2 (G51 to K56) separates the α2 and α3 helices. In L1, residues R19, V20 and S21 and in L2, amino acids K53, T54 and I55 form β-sheet structures. As in the eukaryal histone folds, L1 is juxtaposed parallel to L2a and L2 is next to L1a in an HMfB dimer. However, unlike the eukaryal histone heterodimers, HMfB and all other archaeal histones, known to date, have symmetric homo and heterodimers. Three hydrogen bonds link β1 and β2a to form an intermonomer β-ladder, namely residues R19 and K53a, R19 and I55a, and S21 and I55a. The α1 helix follows a 4-residue-long unstructured N terminal extension, and a 3-amino acid-long C-terminal extension follows the α3 helix (Figure 1.5). Both the N termini and the C termini from each monomer are close to the dyad axis. The histone folds of HMfA and HMfB are remarkably similar and superimposition of the rHMfA homodimer on rHMfB homodimer provides a root-mean-squared deviation of 0.754 Å. These two histones are most homologous in the central core region and differ slightly in the more unstructured N and C terminal regions and in the orientation of α3 helices (Starich et al., 1996, Decanniere et al., 2000).

The two α2 helices in a homodimer are arranged in anti-parallel orientation and cross each other at an angle of ~35° near M35 and M35a and the α1 and α3 helices fold on the same side of α2 helix. The α2-α2a packing places α1 and α1a in contact with
each other but the α3 helices do not interact. A total surface area of 1647 Å² is buried at
the dimer interface of rHMfB, formed mainly via hydrophobic interactions of apolar side
chains in α2 and α2a (Starich et al., 1996, Decanniere et al., 2000). Also, nonpolar
residues are conserved within all archaeal histones known to date at positions 8, 12, 17,
20, 28, 29, 32, 36, 39, 40, 43, 55 and 60, consistent with a common mechanism of
hydrophobic packing in dimer formation (Zhu et al., 1998).

Hydrophobic interactions are responsible for dimer stabilization. Mutational
analysis and comparison of the hyperthermophilic archaeal histone rHMfB with the
mesophilic histone rHFoB have identified several residues that contribute to structural
stability (Li et al., 2000). Substituting larger hydrophobic residues, as that in rHMfB, in
the α2 helix of rHFoB generates variants with greater thermostability. For example the
rHFoB-A3II and K35M variants have increased $T^\circ$ values of +11 and +15°C,
respectively. $T^\circ$ are the thermal unfolding transition midpoint values, determined by
circular dichroism measurements that have been extrapolated to standard state. Similarly,
the reciprocal changes in rHMfB generating the I31A and M35K variants, results in
changes in $T^\circ$ values of −5 and −19°C, respectively. Also, combining A3II and K35M
variants in rHFoB increases the $T^\circ$ value by 25°C, while the reverse changes in rHMfB
result in a variant that does not fold properly and is degraded in E. coli. Conserved
alanine residues are found at position 36 in many of the archaeal histones (Figure 1.5)
and this residue appears to decrease the size of the internal cavity in the hydrophobic core
that is exposed to solvent and consequently increases thermostability. HMfA and HMfB
both have smaller glycine residues at this position and hence have larger cavities in the
hydrophobic core. The HMfB-G36A variant has an increased $T^\circ$ value of 5°C (Li et al., 2000).

A network of electrostatic interactions and hydrogen bonds also contributes to protein stability and is predicted to play a role in HMf-DNA binding. In most cases, amino acid substitutions that create or disrupt favorable ionic networks, generate variants with increased or decreased thermostability, respectively (Li et al., 2000). One intermonomer salt bridge E18 and K53a and two intramonomer salt bridges, D22 and R25 as well as D59a and R52a, a hydrogen bond between R19 and T54a, and β-ladder interactions between L1 and L2a, by hydrogen bonding, maintain the stability of the loops. These residues, especially R19, K53 and T54 appear to be conserved in almost all archaeal histones to date (Decanniere et al., 2000) (Figure 1.5).

In HMfB, arginine at position 37a forms an electrostatic interaction with E33a (3.2 Å) but a more direct salt bridge with D14 in the opposite monomer (3.8 Å) (Figure 1.8). This results in a different orientation for the negatively charged E2a side chain which is hydrogen-bonded to the side chain of R10. In HMfA, position 14 is occupied by the neutral charge asparagine residue which is hydrogen-bonded to E2a, while position 37 contains a glutamate that faces away from N14 and into solution, on the opposite monomer (Decanniere et al., 2000) (Figure 1.9). When R37 in rHMfB is replaced with E37 as in HMfA, by site-directed mutagenesis, the rHMfB-R37E variant has a lower α-helical content and a $T^\circ$ value that is 6°C lower and it does not bind DNA. However, when the R37E mutation is combined with D14N in rHMfB to generate the rHMfB-D14N+R37E variant, it is fully folded and forms wild-type complexes with DNA (Li et al., 1998, Li et al., 2000 and Decanniere et al., 2000). These results can be explained in a
Figure 1.8. Schematic representation of the HMfB crystal structure showing favorable salt bridge and/or hydrogen bonds interactions. Large dotted lines show ionic and/or hydrogen bond interactions, while the thin dotted lines indicate a potentially repulsive ion pair. This figure is adapted from Decanniere et al., 2000.
Figure 1.8.
Figure 1.9. Schematic drawing of the (HMfA)$_2$ crystal structure showing ionic and/or hydrogen bond interactions. Bold dotted lines show favorable interactions. This figure was adapted from Decanniere et al., 2000.
structural context. The rHMfB-R37E variant has disrupted salt bridge interactions with D14 in the opposite monomer and E33 in the same polypeptide chain, and this excess negative charge destabilizes the protein. The loss of DNA binding in rHMfB-R37E can be explained by the change in the orientation of the R10 side chain that is predicted to interact with the DNA. These changes in stability and DNA binding are rescued in the rHMfB-D14N+R37E variant because the hydrogen-bonding network is restored. It is interesting to note that all archaeal histones, known to date, have either N14 and E37, as found in HMfA or have a lysine at position 14 and a large, apolar side chain like M, I or L at position 37 on the opposite monomer (Decanniere et al., 2000) (Figure 1.5). These structural variations could also explain the differences in the complexes formed between HMfA and HMfB with DNA.

Another difference between HMfA and HMfB is that HMfB has a hydrophobic valine residue at position 64 that points toward I39 and I31a in the hydrophobic core (Figure 1.10). Also E34a, in the vicinity of these residues, has only one orientation for its side chain and is salt linked to H49. In HMfA, R64 replaces the valine found in HMfB and the guanidinium side chain is hydrogen bonded to E34a, which can adopt two conformations (Figure 1.11). These differences in hydrophobic packing between HMfA and HMfB can explain the differences in the orientation of α3 helices in these two archaeal histones (Decanniere et al., 2000). It is interesting to note that the wt rHMfA dimer unfolds at a temperature that is 10°C lower than wt rHMfB and the rHMfA-R64V variant has a 10°C higher T° value. HMfB also has an additional lysine residue at position 69 that is not found in HMfA (Figure 1.5). Removal of this C terminal lysine in rHMfB decreases the T° value by 8°C. This residue (K68, K68a) as well as K69, K69a,
Figure 1.10. Schematic representation of the (HMfB)$_2$ crystal structure in the vicinity of V64. Two conformations are indicated for V64 and hydrophobic interactions between this residue and I31a and I39 are shown. This figure was adapted from Decanniere et al., 2000.
Figure 1.10.
Figure 1.11. Schematic drawing of the (HMfA)$_2$ crystal structure in the region surrounding R64. One of two conformations of the acidic E34a side chain forms a salt bridge with the guanidinium group of the R64 side chain and is indicated by a dotted line. This figure has been adapted from Decanniere et al., 2000.
Figure 1.11.
D38, D38a, V64 and V64a appear to increase thermostability of rHMfB by shielding it from solution. HMfA has a lysine at position 38 but has an arginine at position 64, lacks K69 and has a glutamate at position 38 (Li et al., 2000) (Figure 1.5).

A unique proline tetrad structure is seen at the N terminus that involves van der waals intramonomer and intermonomer interactions between C' atoms of P4 and P7 and between C^ of P4 and C' of P7a, respectively (Figure 1.12). The proline cap orients the side chains of R10 and R10a that are predicted to contact the DNA. This structure is not present in the eukaryal histones but appears to be conserved in all the archaeal histones available to date (Decanniere et al., 2000) (Figure 1.5).

**Histone fold-DNA contacts in eukaryal nucleosomes**

The crystal structure of the eukaryal nucleosome core particle shows the specific contacts between the histone folds and DNA. Each histone dimer contacts 27-28 bp over 2.5 turns of the DNA helix, exclusively on the inner face of the DNA and hence 121 bp of DNA, in 12 of 14 minor grooves, interact with the histone fold domains of the octameric complex (Figures 1.6 and 1.7). Specifically, 2 hydrogen bond interactions are made between consecutive phosphates and histone folds in each helical turn. Three independent histone-fold DNA binding sites are evident in each dimer namely one central a1-a1a site and two paired loop regions, L1-L2a and L2-L1a, at the ends of each dimer (Luger et al., 1997). Hence there are 4 a1-a1a and 8 L1-L2 interfaces in the eukaryal nucleosome. The L1-L2 DNA binding interfaces are larger (1400 Å^2) compared to a1-a1a sites (1000 Å^2) in terms of number of phosphate atoms contacted. Similarly, not all
Figure 1.12. Schematic representation of the N terminal proline tetrad in the (HMfA)$_2$ crystal structure. This structure is identical in (HMfB)$_2$. Intramonomer and intermonomer van der Waals contacts are shown by dotted lines. This figure was adapted from Decanniere et al., 2000.
Figure 1.12.
the L1-L2 motifs in the eukaryal heterodimer pairs are identical and subtle differences occur in the number and strength of DNA contacts in these regions. Based on the number of histone-DNA contacts it appears that the superhelix locations differ in binding as follows: +/- 0.5 > or = +/- 3.5 > +/- 4.5 > +/- 1.5 - +/- 2.5 - +/- 5.5 > +/- 6.5. Binding at SHL +/- 0.5 site is enhanced by the N terminal α-helical extension and tail region of H3 (Figure 1.4). Also, several residues in the H2A and H2B tail regions increase DNA binding at SHL +/- 4.5 (Luger and Richmond, 1998a).

Five types of histone-DNA interactions are made at these interfaces. The helix-dipole moments of the H3, H4 and H2B α1 helices and α2 helices of all 4 core histones fix the position of DNA phosphates. Main-chain amide nitrogen atoms in the last turn of the α1 and α2 helices hydrogen bond to DNA phosphate oxygen atoms. Furthermore, histone fold arginine side chains contact the minor groove DNA phosphate backbone, 10 of the 14 possible times, while the other 4 interactions in the minor groove occur with arginine side chains from histone tail regions. Hydrophobic contacts are made between deoxyribose groups and histones. Finally, electrostatic interactions and hydrogen bonds form between basic and hydroxyl side chains and DNA phosphate oxygen atoms. Most of the DNA binding residues are conserved in the histone folds of the four eukaryal core histones. However, additional DNA contacts occur in the N-terminal extensions and tails of the 4 core histones (Luger et al., 1997).

Although eukaryal histones have an abundance of positively charged amino acids to neutralize the phosphodiester DNA backbone, only half of the hydrogen bond interactions with the DNA phosphates are with these histone-fold residue side chains. The histone main-chain hydrogen bonds to the DNA phosphates are more effective in
locking the DNA into position than flexible basic side chains (Luger and Richmond, 1998a). Also, histones do not make many interactions with DNA bases as expected for sequence-independent binding proteins. Therefore, the only specificity associated with DNA binding occurs with curved or flexible DNA molecules that can easily adapt to histone structure.

In the α1-α1 DNA binding region, the N termini of the α1 helices of H3, H4 and H2B are oriented toward individual DNA phosphates (Figures 1.6, 1.7 and 1.13). In H3, the main chain amide groups of K64 and L65 are hydrogen-bonded to phosphate oxygen atoms at SHL 1.5 while in H2B, the main chain amide N atom of S33 forms a hydrogen bond with a DNA phosphate at SHL 4.5. Contacts are made to deoxyribose groups by H3-P66, H4-P32 and H2B-I36 side chains. A conserved arginine residue in the α1 helix, H3-R69, H4-R35 and H2A-R32, has its side chain hydrogen-bonded to a DNA phosphate in the major groove. At the apex of the (H3 + H4) dimer, an H3-R63 residue inserts itself between phosphate groups in the minor groove but is hydrogen-bonded to H4-T30, which prevents the arginine from reaching further into the DNA structure and contacting the bases (Figures 1.6 and 1.13). A hydrophobic contact is also made by H3-L65 with a DNA thymidine base in the major groove (Figure 1.6). In H2A, R29 and R32 contact DNA phosphates as the α1 helix has a different orientation due to the presence of a large, hydrophobic Y39 in L1 (Luger et al., 1997) (Figures 1.7 and 1.13).

The L1-L2 DNA binding site is highly conserved between the 2 pairs of eukaryal histone heterodimers. The N terminal regions of α2 helices of H3, H4 and H2A are oriented toward an individual phosphate group and main chain amide groups are hydrogen bonded to DNA phosphates at the L1-α2 junction. The longer L1 loops of H3
Figure 1.13. Primary sequence alignments of the eukaryal histone fold dimer pairs based on structure. Two DNA binding regions are shown labeled α1α1 and L1L2 motifs. The α helices are indicated by boxed regions, gaps in the alignment are denoted by hyphens, while dots show continuations of the sequences. Open circles represent lysine side chains that contact the DNA major groove, closed circles show histone side chain-DNA contacts and closed squares show residues that make main chain contacts with the DNA. Small dashed lines show hydrogen bonds between L1 and L2 in the β-sheet structures, while ovals show hydrophobic contacts at the base of the L1L2 interactions. Arginine-threonine pairs are shown by the corresponding underlined residues. Buried salt bridges are shown by solid brackets while the interrupted brackets indicate the absence of an ion pair despite conservation of charged residues at these positions. This figure has been adapted from Luger and Richmond, 1998a.
Figure 1.13.
and H2B are further away from the DNA at SHL 2.5 and 5.5, respectively, while the shorter L1 loops of H4 and H2A make more intimate contact with the DNA at SHL 0.5 and 3.5, respectively (Figures 1.6 and 1.7). Main chain amide groups and side chains in the more highly conserved L2 loops interact with the ends of a 27-28 bp stretch of DNA. The length of the DNA contacted by the histone folds is extended by half a helical turn in both directions by conserved lysine residues in L2 namely H3-K115, H4-K77, H2A-K74 and H2B-K82 that are salt linked to the major groove of the DNA (Figures 1.6, 1.7 and 1.13). A conserved arginine residue in the L1 of H3, H4 and H2A namely R83, R45 and R42, respectively, inserts itself into the minor groove by hydrogen bonding to a DNA phosphate, and a hydroxyl side chain of a nearby conserved threonine residue in L2 namely H4-T80, H3-T118 and H2B-T85, respectively, is hydrogen-bonded to the guanidinium side chain of the arginine and restrains this residue from interacting with the DNA bases (Figures 1.6, 1.7 and 1.13). However, this is not the case for H4-R45 which interacts with the thymidine bases of the DNA. Mutations in H4-R45 and H3-T118 are able to suppress the loss of the nucleosomal remodeling complex, SWI/SNF, and change the yeast nucleosome structure thereby indirectly implicating these residues in DNA binding (Kruger et al., 1995, Kurumizaka and Wolffe, 1997). H2B has a glycine in L1 (H2B-G50) and H2A has an arginine (H2A-R77) instead of the conserved arginine and threonine pair, respectively and H2A-R77 does interact with the minor groove at SHL 5.5 (Figures 1.13). But this arginine does not insert itself too far into the DNA because the DNA straightens at this site. Another exception with H2B is that its R83 side chain, which is not salt linked to E90 as seen with H3 and H4, is hydrogen bonded to a DNA phosphate (Luger et al., 1997) (Figure 1.13).
Interactions of histone extensions and tails

The histone tails that extend outside the histone fold domains of the nucleosome core are very basic in nature, containing a high percentage of arginine and lysine residues, and appear to leave the nucleosome at the minor groove of the DNA. Outside the nucleosome, the flexible histone tails appear to make only weak contacts with the DNA (Luger et al., 1997, Luger and Richmond 1998b). Crosslinking studies have revealed two H2A tail contacts with the outside of the DNA in a nucleosome and the N-terminal residue is more flexible in the number of DNA bases it interacts with (Lee et al., 1997). This weak DNA binding appears to prevent transcription factors from accessing the DNA, and conversely the acetylation of histone tails enhances the binding of transcription factors (Luger et al., 1998b, Lee et al., 1993, Godde et al., 1995, Vettesse-Dadey et al., 1996).

The crystal structure of the nucleosome core shows several sites of interaction between the histone tails and DNA but some of the N terminal regions are excluded from the model because of weak electron density (Figures 1.4 and 1.14). In H3, amino acids 39-43 (HRYPY) pass between the juxtaposed minor grooves that create a channel in the DNA superhelix, formed by SHL 1 at the center and SHL −7 at the end of the DNA (Figure 1.4). These minor grooves are aligned as a result of overwinding of the DNA in this region. At SHL 4.5 and −2.5 to −3, a stretch of 8 basic amino acids of the H2B tail (KKRRKTRK) also passes between the DNA superhelix at the adjacent minor grooves and at SHL 4.5, amino acids 13-16 (KAKT) of the H2A tail interact with the outer surface of the minor groove (Figure 1.4). The tail and αN terminal extension of H3 contact 13 bp of DNA at each end and also interact with the H2A' C-terminal tails at
Figure 1.14. The eukaryal nucleosome crystal structure showing the N and C terminal extensions and tails. This structure is viewed along the DNA superhelical axis and histone tail and extension regions, which contact the DNA, are shown by solid lines while residues that could not be defined in the model are shown at random positions by broken lines. The first or last residue of each histone tail, incorporated in the crystal structure model is specified. Acetylation sites are labeled with filled arrows for those included in the crystal structure, while open arrows mark those which are outside. Sites of ubiquitination are labeled by “U” and the knuckle region of H2A that represents the single-turn α helix, preceding the H2A α1 helix, are shown. This figure was adapted from Luger and Richmond, 1998b.
Figure 1.14.
residues 105-117. Amino acids 92-108 of the H2A C-terminal tail form a buried docking domain with the α3 helix of H2A and make β-sheet interactions with the H4' C-terminal tail (residues 95-102) (Figure 1.14). These histone tail and extension contacts appear to stabilize the nucleosomal ramp (Luger et al., 1997).

The H2B α3 and αC extensions are in anti-parallel orientation and form a V-shaped structure (Figure 1.15). The H2B αC helix interacts with 4 amino acids in the α3 helix, forms hydrophobic contacts with the H2A α2 helix, and interacts with amino acids 17-20 in the H2A tail. The N terminal tails of H4 (amino acids 20-26) contact the exposed surface of the (H2A + H2B) heterodimer in an adjacent nucleosome via salt bridges and hydrogen bonds between K16, R19, K20 and R23 in H4 and the clustered acidic amino acids namely E56, E61, E64, D90, E91 and E92 in H2A and H2B-E110 (Figure 1.15). These interactions provide evidence of inter-nucleosome contact in chromatin fiber (Luger et al., 1997).

**Eukaryal transcriptional regulation by histones**

Chromatin is a dynamic structure that provides flexibility for transcriptional regulation. Studies with yeast promoters show that removal of nucleosomes activates transcription and that the N terminal tails of the 4 core histones are needed for transcriptional repression (Grunstein, 1990, Grunstein, 1997). Histone acetylation transferases (HATs) acetylate H4-K8 and K16 and H3-K9 and K14 in the N terminal tail regions as well as other lysine residues in the amino terminal regions of H2A and H2B (Figures 1.4 and 1.14). These modifications facilitate the binding of transcription factors to DNA and destabilize interactions between adjacent nucleosomes in chromatin fiber.
Figure 1.15. Inter-nucleosomal contacts between the H4 tail and the (H2A+H2B) dimer in the crystal structure. The H4 tail is shown in dark shading while the (H2A+H2B) dimer is depicted with light shading. Italicized residues represent those in the H4 tail that contribute to (H2A+H2B) binding and labeled plain residues belong to H4. Hydrogen bonds between nucleosomes are indicated by dashed lines. The H4 short 3-10 helix (residues 25-28) is boxed and is part of the region that interacts with SIR3 protein in silencing transcription in yeast. This figure has been adapted from Luger and Richmond, 1998b.
Figure 1.15.
(Luger and Richmond, 1998b). Some mutations in lysine residues in the tail regions of H3 and H4 to glutamine bypass the need for histone acetyltransferases in transcription activation (Zhang et al., 1998). Although the exact mechanism by which HATs alter chromatin structure is unknown, it is clear that these modifications affect transcription. Acetylated N-terminal histone tails have a reduced affinity for DNA and wrap it less tightly than non-acetylated histones, which might decrease the DNA superhelical writhe (Hong et al., 1993, Bauer et al., 1994, Krajewski and Becker, 1998). In yeast, GCN5p/ADA2p/ADA3p is a transcription coactivator complex that acetylates K16 in the H4 N terminal tail and hence this coactivator is directly able to modify chromatin structure to enhance transcription (Brownell et al., 1996). Similarly, TAF\(_{\mu}250\), a component of the basal transcription factor TFIID, also has histone acetyltransferase activity (Mizzen et al., 1996). Twelve acetylated lysine residues in a histone octamer have been shown to enhance transcription more than 15-fold \textit{in vitro}. However, the amount of charge neutralization needed to destabilize chromatin is so high that other factors are needed to complement acetylation in transcriptional regulation. Serine residues in the H3 amino terminal tail can be phosphorylated and have been implicated with nucleosomal structural defects comparable to those seen with histone acetylation. Also, phosphorylation of linker histone H1 tails weakens the interactions between nucleosomes and thus regulates gene expression (Wolffe and Guschin, 2000). The C-terminal H2A tail is often linked to a 76-residue long ubiquitin peptide in transcriptionally active genes (Figures 1.4 and 1.14). Since the H2A C terminus contacts DNA at the nucleosomal dyad axis, it is thought that the ubiquitinated tail disrupts linker
histone interactions with the nucleosomal core and hence destabilizes higher order chromatin structures (Usachenko et al., 1994).

Many genes require chromatin remodeling complexes to activate transcription. SWI/SNF is a multi-subunit complex that regulates a group of inducible genes in yeast by altering histone-DNA interactions in chromatin (Côte et al., 1994). Although the precise mechanism of SWI/SNF action is not known, one model suggests that the histone octamer might be moved to an upstream position ahead of a tracking SWI/SNF complex (Cairns, 1998). An alternative hypothesis is that the (H2A+H2B) dimer and the (H3+H4) tetramer are removed from the DNA by SWI/SNF which facilitates binding of transcription factors to DNA (Wolffe and Guschin, 2000). Furthermore, it has been suggested that other histone binding proteins act as "histone sinks" to sequester histones that have been released from the DNA by SWI/SNF, and specifically amino acids 3-19 in the H2A N-terminal tail are believed to be recognized by one such protein to remove the (H2A+H2B) dimer from the DNA (Luger and Richmond, 1998b) (Figure 1.14).

Experiments with other nucleosomal remodeling complexes like the chromatin remodeling factor (CHRAC) and nucleosome remodeling factor (NURF) show increased access of transcription factors for chromatin by actively promoting nucleosome mobility (Langst et al., 1999, Mizgucji et al., 1997, Hamiche et al., 1999). This is thought to occur by local disruption of histone-DNA interactions at the edge of the nucleosome core (Wolffe and Guschin, 2000).
Goals of this study

Structural, biochemical and biophysical studies of the HMf family of archaeal histones conducted both in vitro and in vivo, thus far, show that they are true homologs of the eukaryal histones. The crystal structure of the eukaryal nucleosome identifies specific contacts between the histones and DNA, and the NMR solution structure and the recent high-resolution crystal structure of rHMfB suggest that homologous residues in HMf might play a role in DNA binding. This work uses a mutagenesis approach to test these predictions between the prototypic archaeal histone, HMfB, and DNA in vitro by EMSAs, DNA circularization and circular dichroism assays. Furthermore, it shows that higher affinity variants of rHMfB can be constructed and DNA binding of these histones is compared with wild type rHMfB. Finally, the affinities of wild type rHMfA, rHMfB and its variants for DNA are examined.
CHAPTER 2
IDENTIFICATION OF SPECIFIC IN VITRO DNA CONTACTS OF THE ARCHAEAAL HISTONE, rHMFB

INTRODUCTION

As discussed in Chapter 1, archaeal nucleosomes are most similar to the eukaryal (H3+H4) tetramer at the center of a nucleosome. Archaeal nucleosomes contain tetramers that protect ~ 60 bp of DNA from micrococcal nuclease digestion while the (H3+H4) tetramer protects ~70 bp of DNA (Pereira et al., 1997, Reeve et al., 1997, Hayes et al., 1991). Both these tetramers position DNA (Pereira and Reeve, 1999, Sandman and Reeve, 1999, Dong and Van Holde, 1991) and wrap DNA in an overwound conformation of ~10.2 bp/ helical turn in either negative or positive supercoils (Pereira and Reeve, 1999, Musgrave et al., 1991, Hamiche and Richard-Foy, 1998, Alilat et al., 1999). Archaeal and eukaryal histones also share a common histone fold structure comprising a central α2 helix, separated from the shorter α1 and α2 helices by L1 and L2 β-sheet regions, respectively. The antiparallel arrangement of the two central α2 helices in a dimer, places L1 in contact with L2a and L2 in contact with L1a at the opposite end of the dimer. Also, the N termini of the α1 helices are in close proximity at one end of the molecule but at the opposite side, the α3 helices are not in contact (Starich et al., 1999).
The crystal structure of the eukaryal nucleosome shows three histone fold regions of each heterodimer contacting 27-28 bp of DNA over 2.5 consecutive helical turns, namely at the N termini of the α1-α1a site and on the outer surface of the α1 helices, and in the two paired L1-L2 interfaces on the ends of the dimer (Luger et al., 1997).

Several specific interactions occur between histone fold residue side chains and main chains with the phosphodiester DNA backbone, but the main chain interactions are more rigid and fix the DNA position than the flexible side chain interactions (Luger and Richmond 1998a). Most of the interactions with the DNA are non-sequence specific and occur through hydrogen bonds with histone fold main-chain amide groups as well as hydrogen bonds and electrostatic interactions of histone fold side chains with the DNA phosphate oxygen atoms. Interactions with DNA bases are believed to result from DNA structure rather than sequence. An arginine residue inserts into the DNA minor groove at all sites where the minor groove faces the histone octamer and is hydrogen-bonded to the phosphodiester DNA backbone. Additionally, hydrophobic interactions also occur between histone fold side chains and deoxyribose groups. We presumed that most of these histone fold residues would be important to DNA binding, nucleosome formation and integrity but it is not possible to test the roles of individual residues because of the required participation of 4 different eukaryal histones in a nucleosome. Hence, even if a change were introduced into one of the core histones by reverse genetics, its effect might be masked or suppressed by the other 3 wild-type histones at this position. Archaeal histones, on the other hand, do not have this complexity as they can also form
homodimers. Hence, an intensive site-directed mutagenesis approach was used to identify specific DNA contacts in the prototypic archaeal histone, rHMfB. Also, since the archaeal nucleosome has a tetrameric protein core, a variant at one position would be amplified 4 times in a homotetramer, which could identify more subtle DNA interactions, which might have otherwise gone undetected in the eukaryal nucleosome.

MATERIALS AND METHODS

Chemicals, reagents and kits

Chemicals and reagents were purchased from Sigma Chemical Co. (St. Louis, MO). All DNA modifying and restriction enzymes were purchased from Gibco-BRL (Life Technologies Inc.; Gaithersburg, MD), New England BioLabs (NEB, Beverly, MA) or Boehringer Mannheim (Indianapolis, IN). Oligonucleotides were purchased from Ransom Hill Bioscience (Ramona, CA). Pfu DNA polymerase and E. coli XL1 Blue supercompetent cells were purchased from Stratagene (La Jolla, CA). QIAGEN DNA prep kits were purchased from QIAGEN, Inc. (Valencia, CA) and radioactively-labeled chemicals from ICN (Costa Mesa, CA).

Site-specific mutagenesis and purification of rHMfB variants

Mutations were introduced into the pKS323 construct, that has hmfB cloned between the EcoRI and HindIII sites of pKK223-3, using a Quik-Change site-directed mutagenesis kit (Stratagene, La Jolla, CA) and two overlapping mutagenic
oligonucleotide primers in PCR reactions (Table 2.1). The mutagenesis PCR reactions included 50 ng pKS32 template DNA, 1 pmole each of two mutagenic oligonucleotides, 200 μM dNTPs, 10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 20 mM Tris-HCl (pH 8.8), 0.1% Triton X-100, 0.1 μg/μL BSA and 2.5 U of either cloned or native Pfu DNA polymerase (Stratagene) in a 50 μL reaction. The PCR reactions were performed using 18 cycles of the following program: denaturation at 95°C for 30 sec, followed by annealing at 55°C for 1 min and extension at 68°C for 12 min. The methylated parental DNA was digested with Dpn I (5 U) at 37°C for 3 hours. E. coli XL1 Blue supercompetent cells were transformed with the mutagenic DNA and colonies were selected on LB-ampicillin plates. The presence of the desired mutation was confirmed by DNA sequencing and E. coli JM105 was transformed with the mutated plasmid for rHMfB synthesis (Sandman et al., 1995).

A similar protocol to the one described by Sandman et al., (1994) was used to synthesize and purify rHMfB variants. Isopropyl β-thiogalactopyranoside (400μM) was added to an exponentially growing transformant culture at 37°C in LB medium containing ampicillin (100 μg/mL) to induce rHMfB variant synthesis. After 3 hours of incubation at 37°C, the E. coli cells were resuspended in low salt buffer (100 mM NaCl, 50 mM Tris-HCl pH 8.0, 2 mM Na₂HPO₄ pH 8.0) and lysed by passage through a French pressure cell at 20,000 psi. The lysate was separated by centrifugation at 26,000 x g for 30 min in a Sorvall SS34 rotor, followed by ultracentrifugation at 125,000 x g for 90 min at 4°C in a Ti60 rotor. The supernatant was incubated with DNase I at 37°C for 2 hours. The lysate was subsequently boiled in 3M NaCl and allowed to cool to room temperature.
<table>
<thead>
<tr>
<th>Variant</th>
<th>Primer Sequence</th>
<th>Primer #</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2D</td>
<td>GTA AGT ATG GAC TTA CCA ATA GCA CCA ATT GG CAT TCA TAC CTG AAT GGT TAT CGT GGT TAA CC</td>
<td>1069 1070</td>
</tr>
<tr>
<td>E2K</td>
<td>GTA AGT ATG AAA TTA CCA ATA GCA CCA ATT GG CAT TCA TAC TTT AAT GGT TAT CGT GGT TAA CC</td>
<td>1071 1072</td>
</tr>
<tr>
<td>L3C</td>
<td>GTA AGT ATG GAA TGT CCA ATA GCA CCA ATT GGT AG CAT TCA TAC CTT ACA GGT TAT CGT GGT TAA CCA TC</td>
<td>1101 1102</td>
</tr>
<tr>
<td>L3I</td>
<td>GTA AGT ATG GAA ATA CCA ATA GCA CCA ATT GG CAT TCA TAC CTT AAT GGT CAT CGT GGT TAA CC</td>
<td>1073 1074</td>
</tr>
<tr>
<td>P4C</td>
<td>GTA AGT ATG GAA TTA TGT ATA GCA CCA ATT GGT AG CAT TCA TAC CTT AAT ACA TAT CGT GGT TAA CCA TC</td>
<td>1103 1104</td>
</tr>
<tr>
<td>P4S</td>
<td>GTA AGT ATG GAA TTA TCA ATA GCA CCA ATT GG CAT TCA TAC CTT AAT AGT TAT CGT GGT TAA CC</td>
<td>1079 1080</td>
</tr>
<tr>
<td>I5V</td>
<td>GTA AGT ATG GAA TTA CCA GCA GCA CCA ATT GG CAT TCA TAC CTT AAT GAT CAT CGT GGT TAA CC</td>
<td>1083 1084</td>
</tr>
<tr>
<td>A6P</td>
<td>GTA AGT ATG GAA TTA CCA ATA CCA CCA ATT GGT AG CAT TCA TAC CTT AAT GGT TAT GGT GGT TAA CCA TC</td>
<td>1097 1098</td>
</tr>
</tbody>
</table>

^ primer sequences listed on the top of each pair are 5' -> 3'
primer sequences listed on the bottom of each pair are 3' -> 5'
bases that were changed are underlined

Table 2.1 Oligonucleotide sequences used to generate mutations in *hmFB*
(CONTINUED)
<table>
<thead>
<tr>
<th>Variant</th>
<th>Primer Sequence</th>
<th>Primer #</th>
</tr>
</thead>
<tbody>
<tr>
<td>P7A</td>
<td>G GAA TTA CCA ATA GCA GCA ATT GGT AGA ATT ATA AAA G</td>
<td>1099</td>
</tr>
<tr>
<td></td>
<td>C CTT AAT GGT TAT CGT CGT TAA CCA TCT TAA TAT TTT C</td>
<td>1100</td>
</tr>
<tr>
<td>P7C</td>
<td>G GAA TTA CCA ATA GCA GCA TGT ATT GGT AGA ATT ATA AAA G</td>
<td>1133</td>
</tr>
<tr>
<td></td>
<td>C CTT AAT GGT TAT CGT ACA TAA CCA TCT TAA TAT TTT G</td>
<td>1134</td>
</tr>
<tr>
<td>I8V</td>
<td>CCA ATA GCA GCA GGT AGA ATT ATA AAA GAT G</td>
<td>1095</td>
</tr>
<tr>
<td></td>
<td>GGT TAT CGT GGT CAA CCA TCT TAA TAT TTT CTA C</td>
<td>1096</td>
</tr>
<tr>
<td>R10S</td>
<td>CCA ATA GCA CCA ATT GGT AGC ATT ATA AAA GAT G</td>
<td>1087</td>
</tr>
<tr>
<td></td>
<td>GGT TAT CGT GGT TAA CCA TC G TAA TAT TTT CTA C</td>
<td>1088</td>
</tr>
<tr>
<td>R10G</td>
<td>CA ATA GCA CCA ATT GGT GGA ATT ATA AAA GAT G</td>
<td>1089</td>
</tr>
<tr>
<td></td>
<td>GT TAT CGT GGT TAA CCA CCT TAA TAT TTT CTA C</td>
<td>1090</td>
</tr>
<tr>
<td>K13E</td>
<td>CCA ATT GGT AGA ATT ATA GAA GAT GCA GGA GC</td>
<td>1007</td>
</tr>
<tr>
<td></td>
<td>GGT TAA CCA TCT TAA TAT CTT CTA CGT CCT CG</td>
<td>1008</td>
</tr>
<tr>
<td>K13Q</td>
<td>CCA ATT GGT AGA ATT ATA CAA GAT GCA GGA GC</td>
<td>1025</td>
</tr>
<tr>
<td></td>
<td>GGT TAA CCA TCT TAA TAT CTT CTA CGT CCT CG</td>
<td>1026</td>
</tr>
<tr>
<td>K13T</td>
<td>CCA ATT GGT AGA ATT ATA AGA GAT GCA GGA GC</td>
<td>1021</td>
</tr>
<tr>
<td></td>
<td>GGT TAA CCA TCT TAA TAT CTT CTA CGT CCT CG</td>
<td>1022</td>
</tr>
</tbody>
</table>

^ primer sequences listed on the top of each pair are 5' -> 3' 
primer sequences listed on the bottom of each pair are 3' -> 5' 
bases that were changed are underlined

(CONTINUED)
<table>
<thead>
<tr>
<th>Variant</th>
<th>Primer Sequence</th>
<th>Primer #</th>
</tr>
</thead>
<tbody>
<tr>
<td>D14E</td>
<td>GT AGA ATT ATA AAA GAA GCA GGA GCC GAA AG</td>
<td>1005</td>
</tr>
<tr>
<td></td>
<td>CA TCT TAA TAT TTT CTT CGT CCT CGG CTT TC</td>
<td>1006</td>
</tr>
<tr>
<td>D14H</td>
<td>GGT AGA ATT ATA AAA CAT GCA GGA GCC GAA AG</td>
<td>1023</td>
</tr>
<tr>
<td></td>
<td>CCA TCT TAA TAT TTT GTA CGT CCT CGG CTT TC</td>
<td>1024</td>
</tr>
<tr>
<td>A15S</td>
<td>GGT AGA ATT ATA AAA GAT TCA GGA GCC GAA AG</td>
<td>1013</td>
</tr>
<tr>
<td></td>
<td>CCA TCT TAA TAT TTT CTA AGT CCT CGG CTT TC</td>
<td>1014</td>
</tr>
<tr>
<td>A15G</td>
<td>GGT AGA ATT ATA AAA GAT GGA GGA GCC GAA AG</td>
<td>1009</td>
</tr>
<tr>
<td></td>
<td>CCA TCT TAA TAT TTT CTA AGT CCT CGG CTT TC</td>
<td>1010</td>
</tr>
<tr>
<td>A15P</td>
<td>GGT AGA ATT ATA AAA GAT CCA GGA GCC GAA AG</td>
<td>1011</td>
</tr>
<tr>
<td></td>
<td>CCA TCT TAA TAT TTT CTA AGT CCT CGG CTT TC</td>
<td>1012</td>
</tr>
<tr>
<td>G16K</td>
<td>GGT AGA ATT ATA AAA GAT GCA AAG GCC GAA AGA G</td>
<td>1131</td>
</tr>
<tr>
<td></td>
<td>CCA TCT TAA TAT TTT CTA AGT CCT CGG CTT TCT C</td>
<td>1132</td>
</tr>
<tr>
<td>A17V</td>
<td>GAT GCA GGA GTC GAA AGA GTC AGT GAT GAT G</td>
<td>1121</td>
</tr>
<tr>
<td></td>
<td>CTA CGT CCT CAG TTT TCT CAG TCG CTA CTA C</td>
<td>1122</td>
</tr>
<tr>
<td>A17I</td>
<td>GAT GCA GGA ATC GAA AGA GTC AGT GAT GAT G</td>
<td>1123</td>
</tr>
<tr>
<td></td>
<td>CTA CGT CCT TCG TTT TCT CAG TCA CTA CTA C</td>
<td>1124</td>
</tr>
</tbody>
</table>

^ primer sequences listed on the top of each pair are 5' -> 3' primer sequences listed on the bottom of each pair are 3' -> 5' bases that were changed are underlined
<table>
<thead>
<tr>
<th>Variant</th>
<th>Primer Sequence</th>
<th>Primer #</th>
</tr>
</thead>
<tbody>
<tr>
<td>E18D</td>
<td>GAT GCA GGA GCC GAT AGA GTC AGT GAT G</td>
<td>1003</td>
</tr>
<tr>
<td></td>
<td>CTA CGT CCT CGG CTA TCT CAG TCA CTA C</td>
<td>1004</td>
</tr>
<tr>
<td>E18P</td>
<td>GAT GCA GGA GCC CAA AGA GTC AGT GAT G</td>
<td>1019</td>
</tr>
<tr>
<td></td>
<td>CTA CGT CCT CGG GTT TCT CAG TCA CTA C</td>
<td>1020</td>
</tr>
<tr>
<td>E18K</td>
<td>GAT GCA GGA GCC AAA AGA GTC AGT GAT G</td>
<td>1015</td>
</tr>
<tr>
<td></td>
<td>CTA CGT CCT CGG TTT TCT CAG TCA CTA C</td>
<td>1016</td>
</tr>
<tr>
<td>R19K</td>
<td>GAT GCA GGA GCC GAA AAA GTC AGT GAT G</td>
<td>1017</td>
</tr>
<tr>
<td></td>
<td>CTA CGT CCT CGG TTT TCT CAG TCA CTA C</td>
<td>1018</td>
</tr>
<tr>
<td>R19S</td>
<td>GCA GGA GCC GAA AGT GTC AGT GAT GAT GC</td>
<td>1055</td>
</tr>
<tr>
<td></td>
<td>CGT CCT CGG CTT CTA CAG TCA CTA CTA CG</td>
<td>1056</td>
</tr>
<tr>
<td>R19G</td>
<td>GCA GGA GCC GAA GGA GTC AGT GAT GAT GC</td>
<td>1057</td>
</tr>
<tr>
<td></td>
<td>CGT CCT CGG CTT CCT CAG TCA CTA CTA CG</td>
<td>1058</td>
</tr>
</tbody>
</table>

^ primer sequences listed on the top of each pair are 5' -> 3'  
primer sequences listed on the bottom of each pair are 3' -> 5'  
bases that were changed are underlined
<table>
<thead>
<tr>
<th>Variant</th>
<th>Primer Sequence*</th>
<th>Primer #</th>
</tr>
</thead>
<tbody>
<tr>
<td>V20I</td>
<td>GCA GGA GCC GAA AGA ATC AGT GAT GAT GC</td>
<td>1029</td>
</tr>
<tr>
<td></td>
<td>CGT CCT CGG CTT TCT TAT GCA CTA CTA CG</td>
<td>1030</td>
</tr>
<tr>
<td>V20C</td>
<td>GCA GGA GCC GAA AGA TGC AGT GAT GAT GC</td>
<td>1091</td>
</tr>
<tr>
<td></td>
<td>CGT CCT CGG CTT TCT TAT CTA CTA CG</td>
<td>1092</td>
</tr>
<tr>
<td>V20D</td>
<td>CA GGA GCC GAA AGA GAC AGT GAT GAT GC</td>
<td>1031</td>
</tr>
<tr>
<td></td>
<td>GT CCT CGG CTT TCT CTG TCA CTA CTA CG</td>
<td>1032</td>
</tr>
<tr>
<td>R25K</td>
<td>GA GTC AGT GAT GAT GCA AAA ATA ACT CTA GC</td>
<td>1001</td>
</tr>
<tr>
<td></td>
<td>CT CAG TCA CTA CTA CGT TTT TAT TGA GAT CG</td>
<td>1002</td>
</tr>
<tr>
<td>G51K</td>
<td>GCA AGA CAC GCC AAA AGG AAA ACA ATA AAA G</td>
<td>1119</td>
</tr>
<tr>
<td></td>
<td>CGT TCT GTG CGG TTT TCC TTT TGT TAT TTT C</td>
<td>1120</td>
</tr>
<tr>
<td>G51A</td>
<td>GCA AGA CAC GCC GCA AGG AAA ACA ATA AAA G</td>
<td>1061</td>
</tr>
<tr>
<td></td>
<td>CGT TCT GTG CGG CTT TTT TGT TAT TTT C</td>
<td>1062</td>
</tr>
<tr>
<td>R52K</td>
<td>CA AGA CAC GCC GGA AAG AAA ACA ATA AAA GC</td>
<td>1033</td>
</tr>
<tr>
<td></td>
<td>GT TCT GTG CGG CTT TTT TGT TAT TTT CG</td>
<td>1034</td>
</tr>
<tr>
<td>R52H</td>
<td>CA AGA CAC GCC GGA CAT AAA ACA ATA AAA GCA G</td>
<td>1125</td>
</tr>
<tr>
<td></td>
<td>GT TCT GTG CGG CTT GTA TTT TGT TAT TTT CGT C</td>
<td>1126</td>
</tr>
</tbody>
</table>

* primer sequences listed on the top of each pair are 5' -> 3'
* primer sequences listed on the bottom of each pair are 3' -> 5'
* bases that were changed are underlined
<table>
<thead>
<tr>
<th>Variant</th>
<th>Primer Sequence*</th>
<th>Primer #</th>
</tr>
</thead>
<tbody>
<tr>
<td>K53R</td>
<td>GA CAC GCC GGA AGG AGA ACA ATA AAA GCA G</td>
<td>1037</td>
</tr>
<tr>
<td></td>
<td>CT GTG CGG CCT TCC TCT TGT TAT TTT CGT C</td>
<td>1038</td>
</tr>
<tr>
<td>K53T</td>
<td>GA CAC GCC GGA AGG ACA ACA ATA AAA GCA G</td>
<td>1035</td>
</tr>
<tr>
<td></td>
<td>CT GTG CGG CCT TCC TGT TGT TAT TTT CGT C</td>
<td>1036</td>
</tr>
<tr>
<td>T54A</td>
<td>CAC GCC GGA AGG AAA GCA ATA AAA GCA GAG</td>
<td>1059</td>
</tr>
<tr>
<td></td>
<td>GTG CGG CCT TCC TTT CGT TAT TTT CGT CTC</td>
<td>1060</td>
</tr>
<tr>
<td>T54K</td>
<td>CAC GCC GGA AGG AAA AAA ATA AAA GCA GAG</td>
<td>1047</td>
</tr>
<tr>
<td></td>
<td>GTG CGG CCT TCC TTT TTT TAT TTT CGT CTC</td>
<td>1048</td>
</tr>
<tr>
<td>T54R</td>
<td>CAC GCC GGA AGG AAA AGA ATA AAA GCA GAG</td>
<td>1045</td>
</tr>
<tr>
<td></td>
<td>GTG CGG CCT TCC TTT TCT TAT TTT CGT CTC</td>
<td>1046</td>
</tr>
<tr>
<td>T54I</td>
<td>CAC GCC GGA AGG AAA ATA ATA AAA GCA GAG</td>
<td>1127</td>
</tr>
<tr>
<td></td>
<td>GTG CGG CCT TCC TTT TAT TAT TTT CGT CTC</td>
<td>1128</td>
</tr>
</tbody>
</table>

* primer sequences listed on the top of each pair are 5' -> 3'
primer sequences listed on the bottom of each pair are 3' -> 5'
bases that were changed are underlined
<table>
<thead>
<tr>
<th>Variant</th>
<th>Primer Sequence*</th>
<th>Primer #</th>
</tr>
</thead>
<tbody>
<tr>
<td>I55L</td>
<td>C GCC GGA AGG AAA ACA CTA AAA GCA GAG G</td>
<td>1039</td>
</tr>
<tr>
<td></td>
<td>G CGG CCT TCC TTT TGT GAT TTT CGT CTC C</td>
<td>1040</td>
</tr>
<tr>
<td>I55T</td>
<td>C GCC GGA AGG AAA ACA ACA AAA GCA GAG G</td>
<td>1043</td>
</tr>
<tr>
<td></td>
<td>G CGG CCT TCC TTT TGT TGT TTT CGT CTC C</td>
<td>1044</td>
</tr>
<tr>
<td>I55M</td>
<td>C GCC GGA AGG AAA ACA ATG AAA GCA GAG G</td>
<td>1041</td>
</tr>
<tr>
<td></td>
<td>C CGG CCT TCC TTT TGT TAC TTT CGT CTC C</td>
<td>1042</td>
</tr>
<tr>
<td>I55C</td>
<td>C GCC GGA AAG AAA ACA TGT AAA GCA GAG GAC</td>
<td>1093</td>
</tr>
<tr>
<td></td>
<td>G CGG CCT TTC TTC TCT ACA TTT CGT CTC CTG</td>
<td>1094</td>
</tr>
<tr>
<td>K56R</td>
<td>GCC GGA AGG AAA ACA ATA AGA GCA GAG GAC</td>
<td>1063</td>
</tr>
<tr>
<td></td>
<td>CGG CCT TTC TTC TGT TAT TGT CGT CTC CTG</td>
<td>1064</td>
</tr>
<tr>
<td>K56T</td>
<td>GCC GGA AGG AAA ACA ATA ACA GCA GAG GAC</td>
<td>1067</td>
</tr>
<tr>
<td></td>
<td>CGG CCT TTC TTC TGT TAT TGT CGT CTC CTG</td>
<td>1068</td>
</tr>
<tr>
<td>K56E</td>
<td>GCC GGA AGG AAA ACA ATA GAA GCA GAG GAC</td>
<td>1065</td>
</tr>
<tr>
<td></td>
<td>CGG CCT TTC TTC TGT TAT CTT CGT CTC CTG</td>
<td>1066</td>
</tr>
<tr>
<td>K56I</td>
<td>GCC GGA AGG AAA ACA ATA ATA GCA GAG GAC</td>
<td>1109</td>
</tr>
<tr>
<td></td>
<td>CGG CCT TTC TTC TGT TAT TAT CGT CTC CTG</td>
<td>1110</td>
</tr>
<tr>
<td>K56Q</td>
<td>GCC GGA AGG AAA ACA ATA CAA GCA GAG GAC</td>
<td>1111</td>
</tr>
<tr>
<td></td>
<td>CGG CCT TTC TTC TGT TAT CTT CGT CTC CTG</td>
<td>1112</td>
</tr>
</tbody>
</table>

* primer sequences listed on the top of each pair are 5' -> 3'
  primer sequences listed on the bottom of each pair are 3' -> 5'
  bases that were changed are underlined
Denatured proteins were separated by centrifugation and filtration through a 0.45 μm (pore-size) membrane (Millipore). The extract was dialyzed against an appropriate buffer (Table 2.2) used for loading onto heparin affinity columns. The proteins were adsorbed to Hi-Trap heparin-sepharose affinity columns (Pharmacia, Piscataway, NJ) that had been pre-equilibrated with the same concentration of potassium citrate as the loading buffer in 50 mM Tris-HCl (pH 8.0). A wash was also performed with the same buffer. The proteins were eluted with a linear increasing potassium citrate gradient in 50 mM Tris-HCl (pH 8.0) (Table 2.2). Fractions from the columns that contained the rHMfB variants were identified by Tricine SDS-PAGE (16.5%T, 3% C acrylamide resolving gel and a 10%T, 3% C acrylamide spacer gel), run at 2-2.5 V/cm for 1 h and then at 8.5-10.5 V/cm for 2.5-3.5 h. Fractions containing rHMfB variants were pooled and concentrated in Microcon ultra-filtration centrifugal filters (5K MWCO). Concentrations of the rHMfB variants were determined by amino acid composition analysis.

**Folding studies of rHMfB variants using circular dichroism**

Solutions of wild-type rHMfB and its variants (200 μg/mL) were prepared in 50 mM K2SO4, 25 mM MES (pH 6.0). Circular dichroism spectra of these proteins were measured with an AVIV 62A-DS spectropolarimeter (Aviv, Lakewood, NJ) using a similar protocol to that developed by Li et al., (1998). A 1 cm pathlength cylindrical quartz cuvette (Sigma Chemical Co, St. Louis, MO) was used to measure the spectra between 200 and 300 nm at 1 nm intervals with an average data collecting time of 5 seconds at 25°C. Data analysis was performed using Kaleidagraph software. CD
<table>
<thead>
<tr>
<th>rHMfB variant</th>
<th>Isoelectric point</th>
<th>loading buffer*</th>
<th>elution buffer**</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2D</td>
<td>10.38</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>E2K</td>
<td>10.76</td>
<td>45</td>
<td>45-200</td>
</tr>
<tr>
<td>L3C</td>
<td>10.02</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>L3I</td>
<td>10.38</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>P4C</td>
<td>10.02</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>P4S</td>
<td>10.38</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>L5V</td>
<td>10.38</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>P7A</td>
<td>10.38</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>P7C</td>
<td>10.02</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>IBV</td>
<td>10.38</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>R10S</td>
<td>10.02</td>
<td>25</td>
<td>25-200</td>
</tr>
<tr>
<td>K13E</td>
<td>9.25</td>
<td>15</td>
<td>15-200</td>
</tr>
<tr>
<td>K13Q</td>
<td>10.09</td>
<td>20</td>
<td>20-200</td>
</tr>
<tr>
<td>K13T</td>
<td>10.09</td>
<td>20</td>
<td>20-200</td>
</tr>
<tr>
<td>D14E</td>
<td>10.38</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>A15S</td>
<td>10.38</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>A17V</td>
<td>10.38</td>
<td>20</td>
<td>20-200</td>
</tr>
<tr>
<td>A17I</td>
<td>10.38</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>E18D</td>
<td>10.38</td>
<td>20</td>
<td>20-200</td>
</tr>
<tr>
<td>E18P</td>
<td>10.64</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>F19K</td>
<td>10.76</td>
<td>45</td>
<td>45-200</td>
</tr>
<tr>
<td>K19R</td>
<td>10.31</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>R19I</td>
<td>10.02</td>
<td>25</td>
<td>25-200</td>
</tr>
<tr>
<td>R19Q</td>
<td>10.02</td>
<td>25</td>
<td>25-200</td>
</tr>
<tr>
<td>R19S</td>
<td>10.02</td>
<td>20</td>
<td>20-200</td>
</tr>
<tr>
<td>R19G</td>
<td>9.21</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>V20C</td>
<td>10.02</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>V20I</td>
<td>10.38</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>S21C</td>
<td>10.02</td>
<td>15</td>
<td>15-200</td>
</tr>
<tr>
<td>S21T</td>
<td>10.38</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>R25K</td>
<td>10.38</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>G51K</td>
<td>10.55</td>
<td>35</td>
<td>35-200</td>
</tr>
<tr>
<td>R52K</td>
<td>10.31</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>K53R</td>
<td>10.47</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>K53T</td>
<td>10.09</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>T54A</td>
<td>10.38</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>T54K</td>
<td>10.55</td>
<td>35</td>
<td>35-200</td>
</tr>
<tr>
<td>T54R</td>
<td>10.64</td>
<td>35</td>
<td>35-200</td>
</tr>
<tr>
<td>T54I</td>
<td>10.38</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>L55L</td>
<td>10.38</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>L55T</td>
<td>10.38</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>L55M</td>
<td>10.38</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>K56T</td>
<td>10.09</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>K56E</td>
<td>9.25</td>
<td>15</td>
<td>15-200</td>
</tr>
<tr>
<td>K56R</td>
<td>10.47</td>
<td>30</td>
<td>30-200</td>
</tr>
<tr>
<td>R19S+T54K</td>
<td>10.31</td>
<td>40</td>
<td>40-200</td>
</tr>
<tr>
<td>K13T+R19S+T54K</td>
<td>10.02</td>
<td>20</td>
<td>20-200</td>
</tr>
</tbody>
</table>

* loading buffer includes 50 mM Tris-HCl (pH 8.0)
** elution buffer includes a linear increasing gradient of [potassium citrate] indicated in 50 mM Tris-HCl (pH 8.0)

Table 2.2. Purification conditions of rHMfB variants using heparin affinity chromatography.
measurements at 222 nm (θ222) provide an estimate of α-helicity, and the α-helical content of the rHMfB variants was calculated using these ellipticity measurements at 222 nm based on the assumption that wt rHMfB was completely folded (i.e., 100% folded).

Agarose EMSA with rHMfB variants

pBR322 DNA substrates were purified from E. coli DH5α using QIAGEN Maxiprep kit according to instructions supplied by the manufacturer and were linearized with EcoRI. Increasing amounts of the rHMfB variants were incubated with 50 ng amounts of the linearized pBR322 DNA in a 15 μl reaction volume containing 50 mM K2SO4, 25 mM MES (pH 6.0) for 15 min at room temperature (~20°C). Following the addition of 3 μL 5X GLB (40% sucrose w/v, 0.4% bromophenol blue, 0.4% xylene cyanol), the reaction products were subjected to electrophoresis in 0.8% (w/v) agarose (Amresco, Solon, OH) gels run in 1X TBE buffer (90 mM Tris-borate, 2 mM EDTA pH 8.0) for 16 hours. The reaction products were visualized by ethidium bromide staining.

Polyacrylamide EMSA with rHMfB variants

32P Exchange DNA labeling reaction. Linear DNA molecules were end-labeled with 32P using a 5' phosphate exchange reaction with 167 μCi of [γ-32P] ATP (specific activity 7000 Ci/mmol) and 5 U of T4 polynucleotide kinase (BRL) according to Sambrook et al., (1989). The labeling reactions were terminated by the addition of 20 mM EDTA. Unincorporated nucleotides were removed from the labeling reactions by Sephadex G-50 (Sigma Chemical Co., St.Louis, MO) spin column chromatography.
Specific activity measurements of the $^{32}$P-labeled DNA molecules were made using a Beckman model 7500 liquid scintillation counter (Palo Alto, CA) by taking aliquots of the labeling reaction using a protocol described by Mukhopadhyay et al., (1995).

Preparation of linear DNA substrates labeled with $^{32}$P at one end.

pBR322 DNA was purified from E. coli DH5α using the QIAGEN Maxiprep kit, following the instructions provided by the manufacturer. An 89 bp DNA molecule was isolated from pBR322 by digestion with SphI and Sall. To generate a DNA molecule labeled at one end, pBR322 was first linearized with SphI digestion and end-labeled with [$\gamma$-$^32$P] ATP using an exchange labeling reaction. The DNA was purified using QIAGEN PCR purification kit (QIAGEN, Valencia, CA) and then digested with Sall to generate an 89 bp pBR322 labeled fragment at the Sall half-site. The reaction products were separated on an 8% T, 0.11% C polyacrylamide gel in 1X TBE at 8 V/cm and the 89 bp band, visualized by wet gel autoradiography, was excised from the gel. The DNA was eluted from crushed gel pieces by incubation in 500 mM ammonium acetate, 0.1% SDS, 2 mM EDTA and with agitation at 37°C overnight and precipitated according to a procedure described by Sambrook et al., (1989). The specific activity of the purified DNA fragments were determined by extrapolating from the activity of the linearized plasmids in the labeling reactions, described earlier.
Polyacrylamide EMSA with rHMfB variants and 89 bp $^{32}$P-pBR322 DNA

A similar protocol was used for the polyacrylamide gel shift assay as described by Bailey et al., (1999). An 89 bp $^{32}$P-pBR322 DNA molecule (0.1 ng) was incubated with the rHMfB variants (10-150 ng) in a 10 µl reaction containing 50 mM K$_2$SO$_4$, 25 mM MES (pH 6.0) for 30 minutes at 25°C. Controls included protein-free DNA (0.1 ng) and reactions mixtures containing the DNA (0.1 ng) and wild-type rHMfB (75 ng). 5X GLB (2 µl) was added to the reaction mixtures prior to electrophoresis through 8% (w/v) polyacrylamide (acrylamide/bis-acrylamide, 8:0.13 (w/w) gels run in 1X TBE buffer at 8 V/cm for 1.5 hours. The gels were dried and the reaction products were visualized by autoradiography.

Preparation of the 110 bp DNA substrates for polyacrylamide EMSA

Selex cycle 8 clone 1 DNA was amplified from the pGEM T-easy construct using a similar protocol as described by Bailey et al., (2000). PCR amplification reactions included 2 ng of plasmid DNA, 10 pmol each of primers r60F (5' GCTGCAGATGCACGAATTCGAGCTC) and r60R (5' CAGCTCAGAAGCTTGGATCCTGTC), 200 µM dNTPs, 300 mM MgCl$_2$, 50 mM KCl, 20 mM Tris-HCl (pH 8.4) and 5 U Taq DNA polymerase (BRL) in a 100 µl reaction. PCR was performed in 30 cycles using this program: 95°C for 1 min for denaturation, followed by annealing and extension at 72°C for 1 min. The 110 bp amplified DNA was purified using the QIAGEN PCR purification kit and the eluted
DNA was quantitated by UV spectroscopy. The DNA was end-labeled with $[\gamma-^{32}P]$ ATP using an exchange reaction and specific activity measurements were made as described earlier for the 89 bp pBR322 DNA molecules.

Polyacrylamide EMSA with rHMfB variants and 110 bp $^{32}P$-Selex clone 1 DNA

A similar protocol was used for the polyacrylamide gel shift assay with the higher affinity $^{32}P$-Selex cycle 8 clone 1 DNA and the rHMfB variants as described for the polyacrylamide EMSA with $^{32}P$-pBR322 DNA except, in this assay with the Selex DNA, the rHMfB variants (5-50 ng) were added to reaction mixtures containing 0.1 ng of DNA and electrophoresis was performed at 8 V/cm for 2 hours.

Construction of (R19S+T54K) and (K13T+R19S+T54K) variants

Site-directed mutagenesis was performed on the pKS323 construct, containing the T54K substitution in $hmfb$, to introduce the R19S mutation using a similar protocol to the one described earlier for the single variants. The triple variant (K13T+R19S+T54K) was constructed by introducing a K13T mutation into the pKS323 construct, which contained the R19S+T54K mutation in $hmfb$. The double and triple variants were generated using the same mutagenic oligonucleotides that were used for creation of the single variants described earlier (Table 2.1). DNA sequencing was used to confirm the presence of these mutations and these variants were expressed in *E.coli* JM105, as previously described. These variants were purified using a similar protocol as that described for the single variants except that these variants were loaded onto heparin-sepharose columns in 30 mM
potassium citrate, 50 mM Tris-HCl (pH 8.0) and a linear increasing 30-200 mM potassium citrate, 50 mM Tris-HCl (pH 8.0) gradient was used to elute both these variants from the heparin columns.

**Preparation of DNA substrates for DNA circularization assays**

A similar protocol was followed for these experiments as described by Bailey et al., (1999). The pLITMUS28Δ10 DNA that contains a 10 bp deletion in the pLITMUS28 polylinker was a gift from K. Bailey. An 88 bp fragment, generated by restriction digestion of pLITMUS28Δ10 DNA with SpeI and XbaI, was end-labeled with $^{32}$P at the 5' terminus of the XbaI half-site using the exchange reaction as described earlier for preparation of the 89 bp $^{32}$P-pBR322 DNA molecule. The 88 bp pLITMUS28Δ10 DNA molecule, which contains complementary 4 nucleotide overhangs, was purified from an 8% polyacrylamide gel and its specific activity was determined as described for the 89 bp $^{32}$P-pBR322 DNA molecule.

**DNA circularization assays with rHMfB variants**

The 88 bp pLITMUS28Δ10 DNA molecule (0.5 ng) was incubated with wild-type rHMfB and the different rHMfB variants (10-200 ng) in ligase buffer [50 mM Tris-HCl (pH 7.8), 10 mM MgCl₂, 10 mM DTT, 1 mM ATP, 50 μg BSA/mL] in 10 μl reaction volumes for 25 min at 25°C. Phage T4 DNA ligase (80 U) (New England BioLabs, Beverly, MA) was added and incubation was continued at 16°C overnight (> 12 hours). The reactions were deproteinized by incubation with 0.5% (w/v) SDS and 2 μg of
proteinase K for 30 min at 37°C. The reaction products were subjected to electrophoresis at 8 V/cm in 8% native polyacrylamide gels (acrylamide/bis-acrylamide, 8:0.13 w/w) run in 1X TBE buffer. The wet gels were visualized by autoradiography. Ligation products were excised from the wet gels and digested with HindIII or BstWI. Following deproteinization of the reactions with proteinase K (2 µg) and 0.5% (w/v) SDS at 37°C for 30 min, the reaction products were electrophoresed through native 8% polyacrylamide (acrylamide/bis-acrylamide, 8:0.13 w/w) gels in 1X TBE for 1.5 h. The gels were dried and the products were visualized by autoradiography.

Development of circular dichroism assays for archaeal histones

Preparation of DNA substrates for circular dichroism assays: pUC19 DNA was purified from *E. coli* DH5α using QIAGEN Megaprep kit using instructions provided by the manufacturer. The plasmid was linearized by digestion with EcoRI.

CD assay for archaeal histone-DNA interactions

CD assays were developed for archaeal histone-DNA interactions using a similar procedure to that described by Oohara and Wada, (1987) for eukaryal histones. CD spectra of the wild-type rHMfB, rHMfB variants and protein-free pUC19 DNA (30 µg/ml), dissolved in 100 mM KCl, 25 mM K₃PO₄ pH 7.0, were performed at 25°C using an AVIV 62A-DS spectropolarimeter (Aviv, Lakewood, NJ). A 1 cm pathlength cylindrical quartz cuvette (Sigma Chemical Co., St. Louis, MO) was used to measure the spectra between 200 and 300 nm at 1 nm intervals with an average data collecting time of
5 seconds. Increasing amounts of rHMfB or rHMfB variants (0.1-1.3 mass ratios) were added to pUC19 DNA (30 µg/ml) and the solution was mixed several times prior to measuring the CD spectra of each histone addition. BSA (0.1-1.3 mass ratios) was also added to the pUC19 DNA as a negative control. Data analysis was performed using Kaleidagraph software. The relative ellipticities of the histone-DNA complexes at 275 nm were calculated as fractions of the ellipticity of the protein-free pUC19 DNA (θ275).

**Copper phenanthroline crosslinking of rHMfB cysteine variants**

The crosslinking of rHMfB cysteine variants was performed with copper phenanthroline using a protocol described by Bisaccia *et al.*, (1996) and adapted by K. Sandman (unpublished results). Each protein (20 µg) was dialyzed against crosslinking buffer [0.1 M NaCl, 50 mM HEPES (pH 7.5)] in a reaction volume of 50 µl. A 0.1 M stock of o-phenanthroline (Sigma Chemical Co., St. Louis, MO), dissolved in 95% ethanol, and a 0.2 M CuCl2 (Sigma) stock in water were prepared. A final concentration of 7.5 mM o-phenanthroline and 2.5 mM CuCl2 was added to each protein solution and incubation was performed at 25°C for 1 h. The crosslinking reactions were stopped by addition of 1 µl of 50 mM EDTA. The reactions were dialyzed against 0.1 M NaCl, 50 mM HEPES (pH 7.5), 1 mM EDTA at 25°C for 1 h. The proteins (2 µg each) were mixed with an equal volume of non-reducing Tricine gel sample buffer [0.3% (w/v) SDS, 4% (w/v) glycerol, 0.1 mM Tris-HCl pH 8.0, 0.01% bromophenol blue with pH adjusted to 6.8 with HCl]. Protein samples were loaded directly onto denaturing tricine SDS-PAGE gels, as described earlier but without boiling. Controls included protein samples
that were not treated with copper-phenanthroline (2 µg each) and were boiled prior to loading onto gels. Both untreated wild-type rHMfB and wild-type rHMfB (2 µg each), treated with copper-phenanthroline, were included as controls. Electrophoresis was performed at 40 V for 1h and 135 V for 3 hours. The reaction products were visualized following staining with Coomassie Blue.

RESULTS

Construction and purification of rHMfB variants

rHMfB variants were constructed at positions predicted to interact with the DNA at the N termini, in the α1 helices and in the L1 and L2 regions (Table 2.1). Amino acid replacements at these sites included conserved substitutions or changes to residues found in other archaeal histones or in the eukaryal core histones, which are predicted to have less impact on DNA binding, and the more dramatic charge disruptions or changes in hydrophobicity of the residues, predicted to alter amino acid side-chain interactions with the DNA. After each mutation was confirmed by DNA sequencing, the mutant hmfB was expressed in E. coli. The variants listed in Table 2.2 were purified from E. coli, while the rHMfB variants A6P, R10G, D14H, A15G, G16K, V20D, G51A, R52H, R52Q, I55C, K56I and K56Q were not synthesized in E. coli (Table 2.3). The presence of these mutations in hmfB was confirmed by DNA sequencing but since they did not accumulate in E. coli it was presumed that these variants were not properly folded after induction and were rapidly degraded. Their isolation was not further pursued. These results can be
<table>
<thead>
<tr>
<th>100 +/- 10%*</th>
<th>80-89%*</th>
<th>70%*</th>
<th>not produced in E. coli</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2K</td>
<td>E2D</td>
<td>I55M</td>
<td>A6P</td>
</tr>
<tr>
<td>L3C</td>
<td>K13Q</td>
<td></td>
<td>R10G</td>
</tr>
<tr>
<td>L3I</td>
<td>R19Q</td>
<td></td>
<td>D14H</td>
</tr>
<tr>
<td>P4C</td>
<td>R52K</td>
<td></td>
<td>A15G</td>
</tr>
<tr>
<td>P4S</td>
<td>K53T</td>
<td></td>
<td>A15P</td>
</tr>
<tr>
<td>I5V</td>
<td>K56T</td>
<td></td>
<td>G16K</td>
</tr>
<tr>
<td>P7A</td>
<td>K56E</td>
<td></td>
<td>V20D</td>
</tr>
<tr>
<td>P7C</td>
<td></td>
<td></td>
<td>G51A</td>
</tr>
<tr>
<td>I8V</td>
<td></td>
<td></td>
<td>R52H</td>
</tr>
<tr>
<td>R10S</td>
<td></td>
<td></td>
<td>R52Q</td>
</tr>
<tr>
<td>K13E</td>
<td></td>
<td></td>
<td>I55C</td>
</tr>
<tr>
<td>D14E</td>
<td></td>
<td></td>
<td>K56I</td>
</tr>
<tr>
<td>A15S</td>
<td></td>
<td></td>
<td>K56Q</td>
</tr>
<tr>
<td>A17V</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A17I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E18D</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E18P</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E18K</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R19K</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R19I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R19S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R19G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V20C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V20I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S21C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S21T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S21A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R25K</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G51K</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K53R</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T54A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T54K</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T54R</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T54I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I55L</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I55T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K56R</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*% folding was determined by CD measurements using circular dichroism, based on wt rHMfB being 100% folded.

Table 2.3. Folded α-helical contents of rHMfB variants and a list of variants that were not produced in E. coli.
explained by the NMR and crystal structures of rHMfB (Starich et al., 1996, Decanniere et al., 2000). Changes in the conserved Gly16 and Gly51 residues that terminate the α1 and α2 helices, respectively, are expected to have detrimental effects on rHMfB folding (U. Heinneman, personal communication). Similarly, introducing hydrophilic residues at positions containing the non-polar residues V20 and I55, predicted to maintain the stability of the L1-L2 interface through hydrophobic interactions, also results in misfolded variants. The A6P variant is also predicted to be misfolded because proline residues usually occur at sites of helix bending, for example at position 7 in the α1 helix, and hence another proline residue cannot be accommodated next to Pro7 (U. Heinneman, personal communication). All other variants were purified using conditions described earlier (Table 2.2) and a sample elution profile of one of the rHMfB variants namely (K13T+R19S+T54K) is shown in Figure 2.1.

Structural studies of archaeal histone folding

Ellipticity measurements at 222 nm (θ222) provide an estimate of native folding of proteins and the circular dichroism spectra of the rHMfB variants were compared to wt rHMfB (Hirst and Brooks, 1994, Li et al., 1998). An α-helical content of more than 70% was seen for archaeal histone dimers that were completely folded as shown for the rHMfB-T54K variant (Figure 2.2). Based on θ222 values, all the rHMfB variants had almost the same α-helical content (+/- 10%) as wild-type rHMfB except those listed in Table 2.3. The rHMfB variants E2D, K13Q, R19Q, R19S, R52K, K53T, K56T and
Figure 2.1. Tricine SDS-PAGE of fractions from a heparin column used to purify the rHMfB-(K13T+R19S+T54K) variant. Aliquots of the E. coli extract loaded onto the column and samples of the initial flow-through and wash are shown in lanes 1, 2 and 3, respectively. Aliquots of fractions of eluted proteins from the column using potassium citrate (20-200 mM) are shown in lanes 4-16. A sample of wt rHMfB was run in lane 17. The gel was stained with Coomassie Blue.
Fig. 2.1.
Figure 2.2. Comparison of the CD spectra of wt rHMfB (•) and the rHMfB-T54K variant (○). The spectra of the protein solutions (200 μg/ml) were measured in 50 mM K₂SO₄, 25 mM MES (pH 6.0) using an Aviv 62A-DS spectropolarimeter. Measurements at 222 nm (θ₂₂₂) provide an estimate of overall α-helical content.
Fig. 2.2.
K56E had an α-helical content between 80-89% of the wild-type rHMfB, while I55M had ~70% α-helical content.

DNA binding studies of rHMfB variants using agarose gel electrophoretic mobility shift assay

Archaeal nucleosomes, formed on DNA molecules longer than ~2 Kb, migrate faster during electrophoresis through agarose gels than histone-free DNA molecules (Sandman et al., 1994, Sandman et al., 1995). This assay was used to identify rHMfB variants that did not form archaeal nucleosomes as well as variants that formed complexes with decreased, increased or the same electrophoretic mobility as wt rHMfB nucleosomes. DNA binding of all the rHMfB variants listed in Table 2.2 was studied using this assay, namely variants that had changes introduced near the N terminus or in α1 helix, L1 or L2, and a few of the phenotypes under histone:DNA ratios that produce maximum gel shifts are shown in Figure 2.3, while the agarose gel shift assay results for all the variants are summarized in Figure 2.4.

At the N terminus, substitutions for E2 and L3 resulted in variants that did not form stable archaeal nucleosomes. In the α1 helix, P7 and the positively-charged residues R10 and K13, when changed, resulted in variants that either did not bind DNA or produced a decreased gel-shift compared to wt rHMfB complexes (Figure 2.4). Residues in L1 that did not produce stable archaeal nucleosomes, when changed, included A17, R19 and V20, while substitutions for R52, K53, T54, I55 and K56 in L2.
Figure 2.3. Agarose gel electrophoretic mobility shift assays of rHMfB variants.

Complexes formed by the wt rHMfA, wt rHMfB and rHMfB-L5V, R10S, E18K, R19S, R25K, R52K and T54K variants with EcoRI-linearized pBR322 DNA molecules at saturating histone:DNA ratios are shown. These protein and DNA mixtures were incubated at 25°C for 15 min and then subjected to electrophoresis through 0.8% (w/v) agarose gels run at 20 V for 16 h. The complexes were visualized by ethidium bromide staining. The control tracks contain histone-free pBR322 DNA (-).
Figure 2.4. Summary of agarose gel electrophoretic mobility shift assays of the rHMfB variants. rHMfB variants that gave no detectable gel shift are indicated by (X). Variants that formed complexes with the same (✓), decreased (▼) or increased (▲) mobility as wt rHMfB-containing complexes with changes introduced at the N terminus, in the α1 and α2 helices and in L1 and L2 regions are listed.
<table>
<thead>
<tr>
<th>N terminus</th>
<th>α1 helix</th>
<th>L1</th>
<th>α2 helix</th>
<th>L2</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2D ▼</td>
<td>P4C ✓</td>
<td>A17V  ✓</td>
<td>R25K ✓</td>
<td>G51K ▲</td>
</tr>
<tr>
<td>E2K ▲</td>
<td>P4S ✓</td>
<td>A17I  ✓</td>
<td></td>
<td>R52K ▼</td>
</tr>
<tr>
<td>L3C ▲</td>
<td>15V ✓</td>
<td>E18D ✓</td>
<td></td>
<td>K53R ▼</td>
</tr>
<tr>
<td>L3I ▼</td>
<td>P7A ✓</td>
<td>E18P ✓</td>
<td></td>
<td>K53T ▲</td>
</tr>
<tr>
<td></td>
<td>P7C ✓</td>
<td>E18K ▲</td>
<td></td>
<td>K53E ▲</td>
</tr>
<tr>
<td></td>
<td>18V ✓</td>
<td>R19K ▼</td>
<td></td>
<td>T54A ▲</td>
</tr>
<tr>
<td></td>
<td>R10S ✓</td>
<td>R19I  ▼</td>
<td></td>
<td>T54K ▲</td>
</tr>
<tr>
<td></td>
<td>K13E ▼</td>
<td>R19Q ▼</td>
<td></td>
<td>T54R ▼</td>
</tr>
<tr>
<td></td>
<td>K13Q ▼</td>
<td>R19S ▼</td>
<td></td>
<td>T54I ▲</td>
</tr>
<tr>
<td></td>
<td>K13T ▼</td>
<td>R19G ▼</td>
<td></td>
<td>I55L ▼</td>
</tr>
<tr>
<td></td>
<td>D14E ▲</td>
<td>V20C ▼</td>
<td></td>
<td>I55T ▼</td>
</tr>
<tr>
<td></td>
<td>A15S ✓</td>
<td>V20I  ✓</td>
<td></td>
<td>I55M ▼</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S21C ✓</td>
<td></td>
<td>K56T ▲</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S21T ▼</td>
<td></td>
<td>K56E ▲</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>K56R ▼</td>
</tr>
</tbody>
</table>

Fig. 2.4.
resulted in variants that did not bind DNA or generated a reduced gel-shift compared to wt rHMfB. On the other hand, residues like E18K in L1 and G51K in L2 produced complexes that migrated faster through agarose gels with the pBR322 DNA than wt rHMfB nucleosomes (Figures 2.3 and 2.4).

rHMfB variants that produced an agarose gel shift demonstrated the ability to form stable complexes with DNA but there were differences in mobility (Figure 2.3). It is possible that the slower-migrating nucleosomes are less compact or that there are fewer interactions between adjacent nucleosomes on the same DNA molecule, while the complexes that migrate faster, like those containing E18K or G51K, are thought to wrap DNA in a tighter configuration and consequently generate more compact nucleosomes or more interactions between adjacent nucleosomes. The reason for these differences is still unclear and, in this regard, wt rHMfA and wt rHMfB also produce complexes with different mobilities in this assay despite having identical residues at all histone fold positions predicted to bind DNA (Decanniere et al., 2000). Due to this uncertainty, we subsequently focused on variants that did not produce an agarose gel shift.

DNA binding studies of rHMfB variants using polyacrylamide gel electrophoretic mobility shift assays with pBR322 DNA

When archaeal nucleosomes are assembled on DNA molecules <1 Kb, the complexes show a gel retardation relative to the histone-free DNA molecules in a polyacrylamide gel shift assay and mononucleosomes are assembled on DNA molecules that are approximately <120 bp (Bailey et al., 1999). Most of the rHMfB variants that
did not form complexes with the linear pBR322 DNA in the agarose gel shift assays, also
did not form complexes with an 89 bp restriction fragment of pBR322 DNA in
polyacrylamide gel shift assays, and complexes that produced the same gel shift or
increased or decreased gel shifts compared to wt rHMfB in the agarose gel mobility shift
assay, generated a polyacrylamide gel shift like wt HMfB as shown with the results of the
L3I variant (Figures 2.5 and 2.6). However, at higher histone: DNA ratios than were
required to generate a polyacrylamide gel shift with wt rHMfB, the R10S, T54K and
K56T variants did form nucleosomes with the 89 bp DNA fragment but the R10S variant
had a slower mobility than wt rHMfB-containing nucleosomes and the T54K variant-contain-
ing complexes migrated as a diffused band, suggesting instability (Figure 2.5).
Based on these observations, additional assays were used to investigate DNA binding and
wrapping under conditions that did not require the complexes formed to survive for
subsequent extended periods of time during electrophoresis.

Polyacrylamide gel shift assays of rHMfB variants with Selex cycle 8 clone 1 DNA

Polyacrylamide mobility gel shift assays were performed with all the rHMfB
variants shown in Table 2.2 using a 110 bp DNA molecule (Selex cycle 8 clone 1),
selected from a synthetic DNA library of $10^{14}$ different molecules by its high affinity for
wt rHMfB (Bailey et al., 2000). Wild-type rHMfB binds with an ~10-fold higher affinity
to Selex clone 1 than to similar-sized fragments from the polylinker regions of pLITMUS
DNA (K. Bailey, 2000). These experiments were designed to determine if a higher
affinity DNA molecule that could more easily accommodate the structure of rHMfB,
could compensate for the loss of DNA contact at 4 positions in a rHMfB variant
Figure 2.5. Polyacrylamide gel-shift retardation assays of complexes formed with wt rHMfB and rHMfB variants. Reactions mixtures contained 0.1 ng of a $^{32}$P-labeled 89 bp restriction fragment of pBR322 DNA and 10, 25, 50, 75, 100, 125 and 150 ng (lanes 1-7 respectively) of the rHMfB variants shown on the left of each panel. Complexes containing wt rHMfB (75ng) are shown in the lane indicated by + and – shows the histone-free pBR322 DNA. The products were visualized by autoradiography following their separation in an 8% (w/v) polyacrylamide gel at 8 V/cm for 2h.
Fig. 2.5.
Figure 2.6. Summary of the polyacrylamide gel retardation assay results for the rHMfB variants with the 89 bp pBR322 DNA. rHMfB variants that did not produce a gel shift are indicated by (✗) and the variants that generated the same (✔) complexes as wt rHMfB are shown. These variants had changed introduced at the N terminus, in the α1 and α2 helices and in L1 and L2 regions.
Fig. 2.6.
tetramer. However, most of the rHMfB variants that did not bind to pBR322 DNA also did not bind to the Selex cycle 8 clone 1 DNA, based on polyacrylamide gel shift assays (Figure 2.7). The exceptions were rHMfB variants K13Q, K13T and K53T that did bind to the Selex cycle 8 clone 1 DNA (Figure 2.8) but the other variants at these positions, namely K13E and K53E, did not bind to the Selex cycle 8 clone 1 DNA. The rHMfB-R10S variant that did bind to the 89 bp pBR322 DNA but resulted in complexes with a slower electrophoretic mobility than the complexes formed by wt rHMfB, and that did not bind to full-length linear pBR322 DNA based on agarose gel shift assays, also did not bind to the Selex cycle 8 clone 1 DNA based on polyacrylamide gel shift assays (Figure 2.8).

Design, construction and DNA binding studies of rHMfB variants with amino acid substitutions in 2 or 3 locations of DNA binding

Not all substitutions in the α1 helix, L1 and L2 regions, at positions predicted to be involved in DNA binding, resulted in variants that did not bind DNA and therefore rHMfB variants were constructed with substitutions in both the L1 and L2 regions (R19S+T54K) and changes in the α1 helix, L1 and L2 regions (K13T+R19S+T54K) to investigate cooperativity of DNA binding (Figures 2.9 and 2.10). As noted, a hydrogen bond is formed between R19 in L1 and T54 in L2a in almost all archaeal histones and a homologous interaction is present at the equivalent positions in most eukaryal histones. This interaction stabilizes the L1 and L2 loops and also limits the extent to which the arginine side chain extends into the DNA helix (Starich et al., 1996, Decanniere et al., 102
Figure 2.7. Summary of the polyacrylamide gel electrophoresis retardation assay results for the rHMfB variants with the 110 bp Selex cycle 8 clone 1 DNA. (☒) indicates rHMfB variants that gave no gel shift and (✔) indicates variants that generated complexes with the same mobility as wt rHMfB. As listed, these variants had amino acid substitutions changed introduced at the N terminus, in the α1 and α2 helices and in the L1 and L2 regions.
<table>
<thead>
<tr>
<th>N terminus</th>
<th>α1 helix</th>
<th>L1</th>
<th>α2 helix</th>
<th>L2</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2D ✓</td>
<td>P4C ✓</td>
<td>A17V ✗</td>
<td>R25K ✓</td>
<td>G51K ✓</td>
</tr>
<tr>
<td>E2K ✓</td>
<td>P4S ✓</td>
<td>A17I ✗</td>
<td>R52K ✓</td>
<td>R52K ✓</td>
</tr>
<tr>
<td>L3C ✗</td>
<td>I5V ✓</td>
<td>E18D ✓</td>
<td>K53R ✓</td>
<td>K53R ✓</td>
</tr>
<tr>
<td>L3I ✓</td>
<td>P7A ✓</td>
<td>E18P ✓</td>
<td>K53T ✓</td>
<td>K53T ✓</td>
</tr>
<tr>
<td></td>
<td>P7C ✗</td>
<td>E18K ✓</td>
<td>T54A ✗</td>
<td>T54A ✗</td>
</tr>
<tr>
<td></td>
<td>18V ✓</td>
<td>R19K ✓</td>
<td>T54K ✗</td>
<td>T54K ✗</td>
</tr>
<tr>
<td></td>
<td>R10S ✗</td>
<td>R19I ✗</td>
<td>T54R ✗</td>
<td>T54R ✗</td>
</tr>
<tr>
<td></td>
<td>K13E ✗</td>
<td>R19Q ✗</td>
<td>T54I ✗</td>
<td>T54I ✗</td>
</tr>
<tr>
<td></td>
<td>K13Q ✓</td>
<td>R19S ✗</td>
<td>I55L ✓</td>
<td>I55L ✓</td>
</tr>
<tr>
<td></td>
<td>K13T ✓</td>
<td>R19G ✗</td>
<td>I55T ✓</td>
<td>I55T ✓</td>
</tr>
<tr>
<td></td>
<td>D14E ✓</td>
<td>V20C ✗</td>
<td>I55M ✗</td>
<td>I55M ✗</td>
</tr>
<tr>
<td></td>
<td>A15S ✓</td>
<td>V20I ✓</td>
<td>K56T ✓</td>
<td>K56T ✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S21C ✓</td>
<td>K56E ✗</td>
<td>K56E ✗</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S21T ✓</td>
<td>K56R ✓</td>
<td>K56R ✓</td>
</tr>
</tbody>
</table>

Fig. 2.7.
Figure 2.8. Polyacrylamide gel-shift electrophoresis retardation assays of complexes formed by wt rHMfB and rHMfB variants with the Selex cycle 8 clone 1 DNA. Reactions mixtures contained 0.1 ng of $^{32}$P-labeled 110 bp Selex cycle 8 clone 1 DNA and 5, 7, 10, 15, 20, 25, 30, 40, and 50 ng (lanes 1-9) of either wt rHMfB or the rHMfB variants that are listed on the left of each panel. Histone-free Selex cycle 8 clone 1 DNA is shown in the - track. The products were visualized by autoradiography, following their separation in an 8% (w/v) polyacrylamide gel.
Fig. 2.8.
Figure 2.9. Illustration of R19-T54a/R19a-T54 dimer interactions. The figure shows the predicted interactions of R19/T54a residues and R19a/T54 in the L1/ L2a and L1a/L2 regions of a rHMFb dimer, respectively. The side chains of R19/R19a are limited to interactions with the phosphodiester DNA backbone by hydrogen bonding with T54a/T54 in L2, respectively. In an archaeal nucleosome containing a (R19S+T54K) tetramer, 8 DNA contacts would be lost (i.e. 4 contacts between each R19 residue and DNA and 4 between each T54 residue and DNA). This figure was generated by RasMol (R. Sayle, Molecular visualization program, RasMol 2.6, www.umass.edu/microbio/rasmol/distrib/html).
Figure 2.10. Illustration of histone-fold residues in rHMfB, predicted to participate in DNA binding. K13/K13a in α1 helix is predicted to interact with the major groove of the DNA and the R19/R19a and T54/T54a residues in the L1 and L2 regions of a rHMfB dimer appear to interact with the minor groove of the DNA, respectively. A (K13T+R19S+T54K) variant tetramer would lack 12 of these DNA contacts. This figure was generated by RasMol (R. Sayle, Molecular visualization program, RasMol 2.6, http://www.umass.edu/microbio/rasmol/distrib/html).
2000, Luger et al., 1997). To maintain the hydrogen bond interaction between the two loops, the charges were retained in the double variant (R19S+T54K); with a positively charged lysine in L2 replacing the arginine in L1 and a hydrophilic serine in L1 replacing the threonine in L2. In the triple variant (K13T+R19S+T54K) a threonine replaced a conserved lysine at position 13, which decreased the positive charge in the α1 helix, in addition to the R19S+T54K changes introduced into L1 and L2. The side chain of K13 in rHMfB was predicted to contact the phosphodiester backbone in the major groove of the DNA as documented for H4-R72 and H2A-R35 in the eukaryal nucleosome (Figure 2.10). The (R19S+T54K) and (K13T+R19S+T54K) variants accumulated as soluble proteins in E.coli and were purified by heparin-sepharose affinity chromatography. Their CD spectra and θ222 values were very similar to that of wt rHMfB indicating that these proteins were fully folded. In fact, the (R19S+T54K) and (K13T+R19S+T54K) variants both had a slightly higher α-helical content than wt rHMfB namely of 105% and 110%, respectively (Figures 2.11 and 2.12). Neither variant showed any DNA binding based on agarose and polyacrylamide gel mobility shift assays performed using full-length linear pBR322 DNA and an 89 bp restriction fragment of pBR322, respectively (Figures 2.13, 2.14, 2.15 and 2.16). Negative results were also obtained when polyacrylamide gel mobility shift assays were performed with the (R19S+T54K) and (K13T+R19S+T54K) variants using the 110 bp Selex cycle 8 clone 1 DNA (Figure 2.17).
Figure 2.11. Comparison of the circular dichroism spectra of wt rHMfB (•) and the rHMfB-(R19S+T54K) variant (○). CD spectra between 200 and 300 nm were measured for these proteins solutions in 50 mM K₂SO₄, 25 mM MES (pH 6.0) using an Aviv 62A-DS spectropolarimeter. The α-helical content of these proteins in solution was estimated by their θ₂₂₂ values.
Fig. 2.11
Figure 2.12. Circular dichroism spectral comparisons of wt rHMfB (•) and the rHMfB-(K13T+R19S+T54K) variant (○). The CD spectra of these proteins solutions was measured in 50 mM K$_2$SO$_4$, 25 mM MES (pH 6.0) between 200-300 nm with an Aviv 62A-DS spectropolarimeter. The α-helical content of these proteins was estimated by their θ$^{222}$ values.
Fig. 2.12.
Figure 2.13. Agarose electrophoretic mobility shift assay using the rHMfB-(R19S+T54K) variant. EcoRI-linearized pBR322 DNA (50 ng) was incubated with 75 ng of wt rHMfA (rA), 250 ng of wt rHMfB (rB) or 50, 150, 250, 300, 350, 400, 500 ng (tracks 1-7) of the (R19S+T54K) variant. The reaction products were subjected to electrophoresis through a 0.8% (w/v) agarose gel at 20 V for 16 h and were detected by ethidium bromide staining.
Fig. 2.13.
Figure 2.14. Agarose gel mobility shift assay of the rHMfB-(K13T+R19S+T54K) variant. EcoRI-linearized pBR322 DNA (50ng) was incubated with 75 ng of wt rHMfA (rA), 250 ng of wt rHMfB (rB) or 50, 100, 150, 250, 300, 350, 400, 500 ng (tracks 1-8) of the (K13T+R19S+T54K) variant. The reaction products were subjected to electrophoresis in a 0.8% (w/v) agarose gel at 20V for 16 h and the complexes were visualized following staining with ethidium bromide.
Fig. 2.14.
Figure 2.15. Polyacrylamide gel electrophoresis retardation assay of complexes formed by the rHMfB-(R19S+T54K) variant with 89 bp pBR322 DNA. Reactions mixtures that contained 0.1 ng of a $^{32}$P-labeled 89 bp restriction fragment of pBR322 DNA and 10, 25, 50, 75, 100, 125, 150 and 200 ng (lanes 1-8) of the (R19S+T54K) variant were visualized by autoradiography, following their separation in an 8% (w/v) polyacrylamide gel that was run at 8 V/cm for 2h. Complexes formed by wt rHMfB (75ng) are shown in the lane indicated by (+), while histone-free DNA is shown in lane (-).
Fig. 2.15.
Figure 2.16. Polyacrylamide gel-shift retardation assay of complexes formed with the rHMfB-(K13T+R19S+T54K) variant and 89 bp pBR322 DNA. Reactions mixtures that contained 0.1 ng of a ^32P-labeled 89 bp restriction fragment of pBR322 DNA and 10, 25, 50, 75, 100, 125, 150 and 200 ng (lanes 1-8) of the (K13T+R19S+T54K) variant were detected by autoradiography, following their separation in an 8% (w/v) polyacrylamide gel at 8 V/cm for 2h. Histone-free DNA is shown in the (-) lane, while complexes containing wt rHMfB (75ng) are shown in the lane indicated by (+).
Fig. 2.16.
Figure 2.17. Polyacrylamide gel electrophoresis retardation assays of complexes formed by wt rHMfB and the rHMfB-(R19S+T54K) and (K13T+R19S+T54K) variants with the 110 bp Selex cycle 8 clone 1 DNA. The DNA (0.1 ng) was incubated with 5, 7, 10, 15, 20, 25, 30, 40, and 50 ng (lanes 1-9) of wt rHMfB or the rHMfB variants indicated to the left of each gel panel and the products generated were visualized by autoradiography, following their separation through an 8% (w/v) polyacrylamide gel run at 8 V/cm for 2h. The (-) track contained the histone-free Selex cycle 8 clone 1 DNA.
Fig. 2.17.
DNA circularization assays

Linear DNA molecules that are too short to be circularized by DNA ligase in the absence of protein-induced bending form monomer circles when incubated with archaeal histones in the presence of DNA ligase (Bailey et al., 1999). This assay for histone-DNA interaction does not depend on the survival of the complexes formed for extended periods of time during electrophoresis and did not require a change of buffer conditions. It was used to assay DNA binding by rHMfB variants that did not form complexes with DNA based on gel shift assays. As shown in Figure 2.18, rHMfB-(K13T) and (T54K) variants did facilitate DNA circularization, although 2-fold and 5-fold higher histone: DNA ratios were needed for circularization than by wt rHMfB, respectively. At very high histone:DNA ratios, DNA ligase is inhibited by the histones (data not shown). Addition of the rHMfB-(R19S) and (R19S+T45K) variants also resulted in monomer circle formation but only at more than 10-fold higher histone: DNA ratios than with wt rHMfB (Figure 2.18). However, no monomer circles were formed in the presence of the rHMfB-(K13T+R19S+T54K) variant, even at 20-fold higher histone: DNA ratios than were needed for DNA circularization by wt rHMfB. A ladder, formed by tandemly ligated linear multimers and circular dimers, was generated even in the presence of very large amounts of the (K13T+R19S+T54K) variant indicating that this protein did not inhibit the activity of DNA ligase, and therefore probably did not bind to DNA (Figure 2.18).

Circular dichroism assays of rHMfB interactions with DNA

The ellipticity of a DNA solution from 260-300 nm was reduced by histone (H2A+H2B) dimer binding (Oohara and Wada 1987) and adding increasing amounts of...
Figure 2.18. Ligase-catalyzed DNA circularization assays in the presence of wt rHMfB and rHMfB variants. Reaction mixtures that contained a $^{32}$P-labeled 88 bp restriction fragment of pLITMUS28 DNA (0.5 ng) and either wt rHMfB or the variant listed on the top of each panel at the amount (ng) shown above each track were incubated with T4 DNA ligase for ~ 12 h at 16°C. Following deproteinization, the products generated were separated by electrophoresis through an 8% (w/v) polyacrylamide gel, run at 8 V/cm for 2 h and visualized by autoradiography. The controls were (A) untreated 88 bp DNA, (B) DNA incubated only with T4 ligase, and (C) DNA and 10 ng of either wt rHMfB or the variant in each gel incubated in the absence of ligase. Tandemly ligated linear multimers and dimer circles (DC) were found in the control track B, whereas with rHMfB addition, the major ligation product was monomer circles (MC) (Bailey et al., 1999).
Fig. 2.18.
wt rHMfB to linear pUC19 DNA, similarly, resulted in a proportional reduction in the ellipticity of the DNA at 275 nm (Figure 2.19). This assay was therefore also used to measure the binding of wt rHMfB and rHMfB variants to DNA in solution. Adding the K13T and T54K variants to the pUC19 DNA resulted in a decrease in \( \theta_{275} \), but higher amounts were needed to obtain the same reduction in ellipticity as observed with wt rHMfB (Figure 2.19). At much higher histone: DNA ratios, the R19S and (R19S + T54K) variants also produced slight reductions in \( \theta_{275} \), consistent with the results of the DNA circularization assays obtained with these variants. In contrast, the triple variant (K13T + R19S + T54K) and a bovine serum albumin (BSA) control showed no decrease in ellipticity of the DNA even at very high histone: DNA ratios, consistent with the ligase-circularization assay (Figure 2.19).

**Structural studies of rHMfB using crosslinking of rHMfB-cysteine variants**

rHMfB variants L3C, P4C, P7C, V20C and S21C were incubated Cu\(^{2+}\)-phenanthroline to generate disulfide bonds as probes for the rHMfB dimer structure. It was assumed that only residues close enough and in the proper orientation to reach each other would be crosslinked to form disulfide bridges within the dimer. The rHMfB-L3C, P4C and P7C variants were crosslinked into dimers whereas the V20C and S21C variants were not crosslinked (Figure 2.20). Increasing the concentration of these histones by 5-fold had no effect on the extent of crosslinking and so it was assumed that dimer formation was a prerequisite for crosslinking and the random collision of two monomers was not sufficient for crosslinking (data not shown). Agarose gel shift with the
Figure 2.19. Circular dichroism assays of rHMfB and rHMfB variants binding to pUC19 DNA. (A) CD spectra of a pUC19 DNA (30 μg/ml) solution and mixtures containing rHMfB and DNA at 0.1, 0.2 and 0.3 histone:DNA mass ratios. CD intensities are shown as the difference in ellipticity between left (ε_L) and right (ε_R) circularly polarized light.

(B) The decrease in the ellipticity of the pUC19 DNA at 275 nm resulting from the addition of increasing amounts of bovine serum albumin (BSA), wt rHMfB or the indicated variants expressed as a fraction of the ellipticity at 275 nm of the histone-free pUC19 DNA.
Fig. 2.17.
Figure 2.20. Copper phenanthroline crosslinking of rHMfB variants. Wild-type rHMfB or the cysteine-substituted variants (20 μg each) indicated above each track were incubated with 7.5 mM o-phenanthroline and 2.5 mM cupric chloride at 25°C for 1h. The reaction products (+) were separated by denaturing tricine SDS-PAGE adjacent to untreated control samples (−). The lanes indicated by (M) contain broad-range SDS-PAGE molecular weight markers (BioRad Laboratories, Richmond, CA). The positions of rHMfB monomers (~7 kDa) and dimers (~14 kDa) are indicated.
Fig. 2.20.
DISCUSSION

As discussed in detail below, results obtained confirm the functional homologies of archaeal and eukaryal histone fold residues that were predicted based on their apparent structural homology. HMfB residues L3 at the N terminus, R10 and K13 in α1, R19 in L1 and K53, T54 and K56 in L2 are central to DNA binding and HMfB-DNA complex stabilization. These residues are almost unusually conserved in archaeal histones and therefore the information obtained here from the rHMfB variants can be extrapolated readily to DNA binding issues for all the archaeal and most eukaryal histone fold residues.

At least one rHMfB variant, generated by a substitution of L3 at the N terminus, R10 or K13 in α1 helix, R19 or V20 in L1, and K53, T54 or K56 in L2 could not form a stable histone: DNA complex based on agarose gel mobility shift assays (Figure 2.21). These positions are equivalent to those shown to contact the DNA in an eukaryal nucleosome (Luger et al., 1997, Luger and Richmond., 1998a). In yeast, mutations at these positions in H3 and H4 suppress the defective phenotype resulting from mutations in the chromatin remodeling complex SWI/SNF, consistent with these residues being involved in stabilization of DNA binding (Kruger et al., 1995, Kurumizaka and Wolffe, 1997, Santisteban et al., 1997, Wechser et al., 1997). To date there are ~30 archaeal histone sequences available that have residues which are almost completely conserved at
these sites except at position 13 which is occupied by either a lysine (as in rHMfB) or an arginine (www.biosci.ohio-state.edu/microbio/Archaealhistones/Alignments/alignments.html).

Archaeal histones have unique N terminal E2L3P4_A6P7 sequences that form a proline tetrad structure within an archaeal dimer. This region is predicted to interact directly with the DNA, near the center of a rHMfB dimer (Decanniere et al., 2000), and in the agarose gel shift assay, the rHMfB-(E2D) variant produced only a slight gel shift, however the E2K variant formed complexes similar to those formed by wt rHMfB (Figure 2.4). This was surprising as the side chain of E2 is normally hydrogen-bonded to R10, and forms part of an ion-pair network with D14, E33 and R37 (Figure 1.8) that positions the R10 side chain appropriately for DNA contact (Decanniere et al., 2000). Changing from the negatively charged glutamate (E2) to a positively charged lysine residue would be expected to result in an unfavorable ionic interaction with R10. But based on θ222 nm measurements, the E2K variant was fully folded and it does bind to DNA. In rHMfA, it is the backbone rather than the side chain of E2 that is involved in hydrogen bonding with R10 (Figure 1.9) (Decanniere et al., 2000), and possibly this is also the case for the rHMfB-(E2K) variant. It is also possible that a compensatory favorable interaction is created between D14 and the lysine residue in the E2K variant that restores some stability to the α1 helix that would otherwise be through unfavorable interactions between E2 and R10 in the E2K variant.

The L3C variant did not produce a gel shift in the agarose EMSA and the conservative L3I variant gave a decreased gel shift when compared to wt rHMfB,
Figure 2.21. Illustration of rHMfB residues identified as being involved in DNA binding. The upper image was generated by RasMol (R. Sayle, Molecular visualization program, RasMol 2.6, http://www.umass.edu/microbio/rasmol/distrib/html) and shows an rHMfB dimer complexed with DNA, as predicted by homology with the eukaryal histone-fold DNA contacts established in the eukaryal nucleosome (Luger et al., 1997). The numbers above the alignment indicate positions in the rHMfB sequence and (W) marks a residue which, when changed, resulted in at least one rHMfB variant that did not generate a complex, based on a gel shift assay. Residues that are conserved in at least 30 available archaeal sequences are listed below the rHMfB sequence (Reeve et al., 1999). Histone-fold residues that contact the DNA in H3 and H4 are circled and changes in residues that suppress the SIN phenotype in yeast are shown by a downward pointing arrow below the wild-type residue (Kruger et al., 1995, Santisteban et al., 1997, Wechser et al., 1997). The H4-T82I substitution is lethal in H4 because of a defect in chromatin assembly. It is suppressed by addition of the A89V substitution (Smith et al., 1996).
suggesting L3 is important but not essential for DNA binding. Surprisingly, the P4C and P4S variants formed stable complexes, like wt rHMfB, based on the agarose and polyacrylamide EMSAs (Figures 2.4, 2.6 and 2.7). Both these residues are conserved in all archaeal histones and form the proline tetrad that is predicted to orient the two α1 helices to allow hydrogen bond formation between R10/ R10a and backbone oxygens (Decanniere et al., 2000). Although the P7A variant did bind DNA, the P7C variant did not bind DNA, probably due to destabilization of the proline tetrad.

Two positively charged residues in α1, R10 and K13, apparently also contribute to DNA binding. The R10S variant did not bind DNA, based on agarose EMSAs, whereas the conservative R10K variant gave a decreased gel shift when compared with wt rHMfB (Figure 2.4). Homologous residues to R10 in H3 (R69) and H4 (R32) face the major groove and make hydrogen bond contacts with the DNA phosphates in the eukaryal nucleosome (Luger et al., 1997, Luger and Richmond, 1998a). The K13E, K13Q and K13T variants gave negative results with both the agarose and polyacrylamide gel EMSAs (Figures 2.4, 2.6 and 2.7) and K13 in HMfB is homologous to H4 (R72) and H2A (R35) that contact the DNA phosphodiester backbone through hydrogen bond interactions (Luger et al., 1998). The conservative K13R variant (provided by K. Sandman), however, forms complexes with electrophoretic mobilities that are indistinguishable from those complexes formed by wt rHMfB. Both arginine and lysine residues are found naturally at position 13 in different archaeal histones (www.biosci.ohio-state.edu/microbio/Archaealhistones/Alignments/alignments.html).
Many substitutions introduced into the L1 and L2 regions resulted in negative EMSA. The A17V and A17I substitutions in L1 generated variants that did bind DNA, based on EMSA. These substitutions would introduce more hydrophobicity and a larger side chain into the L1 region, indirectly affecting DNA binding. A similar V43I replacement at the equivalent position in H4 resulted in introducing greater bulk and this was argued to generate a Sin - phenotype and thereby cause a disruption in the L1 DNA binding site (Kruger et al., 1995, Wescher et al., 1997).

Almost all R19 substitutions in L1 resulted in variants that gave negative EMSA (Figures 2.4, 2.6 and 2.7), although the conservative R19K variant formed complexes with decreased mobility when compared with complexes formed by wt rHMfB. R19 is almost completely conserved in all archaeal histones and is structurally homologous to H3-R83, H4-R45 and H2A-R42, which in the eukaryal nucleosome, makes guanidinium side chain contacts with DNA phosphates in the minor groove. These interactions with the DNA phosphate backbone are limited by hydrogen bonding between the arginine residue and a conserved threonine residue in L2. This prevents arginine:DNA base interactions (Luger et al., 1997). However in H2B, a glycine residue (G50) is located at this position, filled by an arginine residue in HMfB, H2A, H3 and H4, and this glycine is hydrogen-bonded to R77 in L2 of H2A which allow R77 to reach further into the DNA minor groove but it still does not interact with the DNA bases because the DNA straightens and changes trajectory at this site (Luger et al., 1997). T54 is the L2 partner of R19 in L1 in rHMfB and none of the T54 variants constructed, namely T54A, T54I, T54R and T54K, nor the T54V, T54Y, T54C and T54S variants, constructed by K.
Sandman, formed stable complexes based on agarose EMSAs. A structurally homologous H3-(T118I) variant has a Sin' phenotype and forms nucleosomes with increased sensitivity to micrococcal nuclease and DNase I (Kruger et al., 1995, Kurumizaka and Wolffe, 1997). In the eukaryal nucleosome, H3-T118, H4-T80 and H2B-T85 form hydrogen bonds with DNA phosphate oxygen atoms (Luger et al., 1997) and the crystal structure of the rHMfA dimer contains a sulfate anion next to T54, apparently mimicking a DNA phosphate group (Decanniere et al., 2000). As T54 is conserved in all archaeal histones (www.biosci.ohio-state.edu/microbio/Archaealhistones/Alignments/alignments.html), it appears that T54 is critically important for DNA binding. Presumably, it restrains the reach of the arginine (R19), conserved in L1, to the DNA and also directly interacts with the DNA through a hydrogen bond.

Although the V20C variant in L1 also gave negative EMSA results, the V20I variant formed complexes with mobilities like wt rHMfB complexes (Figures 2.4, 2.6 and 2.7). This position is occupied by either a valine or an isoleucine in archaeal histones and, in rHMfB, V20 has been shown to interact with I55a in L2 through hydrophobic contacts (Starich et al., 1996, Decanniere et al., 2000). Similar hydrophobic interactions also occur between the homologous residues H3-F84 in L1 and H4-V81 in L2, between H4-I46 in L1 and H3-I119 in L2, between H2A-V43 in L1 and H2B-I86 in L2 and between H2B-I51 in L1 and H2A-I78 in L2 (Luger et al., 1998). It seems likely that the V20C variant does not interact with the DNA because this substitution destabilizes the L1-L2 whereas the isoleucine residue in the V20I variant is still able to make stabilizing
hydrophobic contacts with I55 in L2a. Substitutions in L2, namely I55L, I55T and I55M all resulted in variants that formed complexes with decreased mobility in the agarose gel shift assay when compared with wt rHMfB, while complexes formed by the I55V variant had the same mobility as wt rHMfB (Figure 2.4). Valine apparently can substitute fully for isoleucine at position 55 in L2 to maintain a stable hydrophobic interaction between the paired loops in a rHMfB dimer, whereas changes to leucine, threonine or methionine all result in a slight destabilization of the L1-L2 interface. Consistent with these results, variants with non-conservative substitutions, V20D and I55C, did not accumulate in E. coli (Table 2.3) presumably because those hydrophilic substitutions disrupted the hydrophobic interaction between L1 and L2, and prevented correct folding.

Several other substitutions in the L2 region at positions 52, 53 and 56 resulted in variants that either gave no gel-shift or produced reduced shifts when compared with wt rHMfB (Figure 2.5). The R52K variant gave a reduced agarose EMSA but apparently bound DNA normally in the polyacrylamide EMSAs (Figures 2.6 and 2.8). The homologous position in H3 is filled by R116 and the R116H variant has a Sin' phenotype, consistent with reduced hydrogen bonding with D85 in the α3 helix of H4 and less stabilization of the L1-L2 DNA binding interface (Kruger et al., 1995, Kurumizaka and Wolffe, 1997, Luger et al., 1997). In an eukaryal nucleosome, H3-R116 and H4-R78 are very close to the DNA phosphodiester backbone but actually face away from the DNA and form a buried salt-bridge with an aspartate in the α3 helix (Luger et al., 1997). Similarly, in rHMfB, R52 forms an intramolecular salt bridge with in α3 and this interaction stabilizes the L1-L2 region of the rHMfB dimer (Starich et al., 1996,
Decanniere et al., 2000). An arginine residue is conserved at position 52 in all archaeal histones currently available (www.biosci.ohio-state.edu/microbio/Archaealhistones/Alignments/alignments.html) and therefore it seems likely that the R52K variant must be able to form a similar stabilizing bond with D59, although the slightly altered interaction with the DNA suggests some destabilization.

Consistent with this misfolding resulting in instability in E. coli, the R52H and R52Q variants did not accumulate in E. coli (Table 2.3) and these variants most likely could not form a salt bridge interaction with D59 and would therefore be misfolded.

Based on negative agarose and polyacrylamide EMSAs, the K53E and K53T variants with substitutions in L2, did not form stable complexes with the DNA (Figures 2.4 and 2.6), whereas the conservative K53R variant (made by K. Sandman) formed complexes with reduced mobility compared to wt rHMfB complexes. As noted, K53 forms an intermonomer salt bridge with E18 that stabilizes the paired loop regions in a HMfB dimer and is also predicted to form salt bridges with DNA phosphates. This prediction is supported by the presence of a chloride anion at this location in the crystal structure of (rHMfA)₂ (Starich et al., 1996, Decanniere et al., 2000). Based on these arguments, the K53E and K53T variants will lack ionic interactions with E18 and therefore have a less stable L1-L2 interface and will have lost salt bridge interactions with the DNA. In constrast, the K53R variant retains DNA binding activity presumably because the positively-charged arginine substitution maintains a salt bridge interaction with E18 and with the DNA. Substitutions at position 56 in L2, resulted in the K56T and K56E variants, that also did not form stable complexes with DNA, based on agarose
EMSAs (Figure 2.4). However, although this residue is conserved in most archael histones, it does not have a clear structural homolog in the eukaryal histones (Sandman et al., 1998, Luger et al., 1997). Therefore, the role of this residue in DNA binding cannot be readily modeled using information from the eukaryal nucleosome. It is possible that K56 contacts the DNA minor groove via hydrogen bond interactions but this can only be confirmed by an archael nucleosome structure.

Most of the rHMfB variants that did not bind to a fragment of the pBR322 DNA, based on polyacrylamide EMSAs, also did not bind to a DNA molecule that was selected by its high HMfB affinity in the Selex procedure. However, the rHMfB-K13Q, K13T and K53T variants did form complexes with the Selex DNA (Figure 2.7), indicating that their defects in DNA binding can be suppressed by a higher affinity DNA molecule. Possibly intrinsic curvature or increased flexibility makes this DNA better able to accommodate to the tetramer structure of these histone variants that have fewer DNA binding sites (Bailey et al., 2000).

As discussed, not all changes at most of the critical positions resulted in rHMfB variants that did not produce agarose gel shifts, except for position 54, and at least one variant was obtained at every position investigated that did form a stable complex based on agarose EMSAs (Figures 2.4, 2.6 and 2.7). This observation argues for the additive roles of residues in DNA binding in which the loss of binding at one histone fold position does not prevent DNA binding at other locations. Furthermore, the DNA binding contacts that remain appear to be sufficient to assemble an archael nucleosome. This was further investigated by constructing the (R19S+T54K) and (K13T+R19S+T54K)
variants with substitutions at two (L1 and L2) and three (α1, L1' and L2) histone fold locations, respectively, which would be predicted to lose 8 and 12 contacts with the DNA in a tetramer-based archaeal nucleosome (Figures 2.9 and 2.10). Individual substitutions at these locations generated variants that did not form stable complexes based on the agarose and polyacrylamide EMSAs, but did bind and wrap DNA based on the θ275 reduction and DNA circularization assays (Figures 2.19 and 2.18). The (R19S+T54K) variant with two substitutions retained some DNA binding and wrapping activity, while the (K13T+R19S+T54K) variant lost all DNA binding activity (Figures 2.18 and 2.19).

The results obtained using copper phenanthroline crosslinking of rHMfB cysteine-substituted variants was recently confirmed by the rHMfB crystal structure. At the N terminus, residues L3, P4 and P7 are sufficiently close to L3a, P4a and P7a to interact within a dimer at 1.75, 3.63 and 7.2 Å, respectively, whereas V20 and S21 in the L1 region are far apart from V20a and S21a, on opposite ends of a dimer, at 34 and 30 Å, respectively (Figure 2.22). The proline residues at positions 4/4a and 7/7a form a proline tetrad through van der Waals interactions and hence the C4/4a and the C7/7a variants were expected to crosslink.
Figure 2.22. Illustration identifying the positions of residues that were changed to cysteines in rHMfB variants. The N terminal L3, P4 and P7 residues are close to their partners in a dimer namely L3a, P4a and P7a, respectively, whereas V20 and S21 in the L1 regions are on the opposite ends of a dimer molecule from V20a and S21a. RasMol (R. Sayle, Molecular visualization program, RasMol 2.6, http://www.umass.edu/microbio/rasmol/distrib/html) was used to generate this figure.
INTRODUCTION

Studies of variants of the archaeal histone, rHMfB, established that residues L3 at the N terminus, R10 and K13 in the α1 helix, R19 in L1 and K53, T54 and K56 in L2 form DNA contacts in an archaeal nucleosome (Chapter 2). At the N terminus in rHMfB, L3 is believed to interact either hydrophobically with the DNA bases or through main-chain interactions with the phosphodiester backbone. The two positively-charged residues at positions 10 and 13 in the α1 helices of rHMfB are thought to form hydrogen bonds with the phosphodiester DNA backbone in the major groove via side-chains, as shown by homologous interactions in H3, H4 and H2A (Luger et al., 1997). In L1 of rHMfB, R19 is thought to interact with a DNA phosphate in the minor groove, while T54 in L2 of the opposite monomer is believed to prevent the guanidinium side chain of R19 from penetrating the DNA bases by forming a hydrogen bond with R19. This interaction is conserved in most of the eukaryal histone-fold pairs and T54 in L2 of rHMfB is also believed to participate in DNA binding by hydrogen bonding to a DNA phosphate. Homologous interactions also occur in the (H3+H4) histone-fold pairs and at one end of
an (H2A+H2B) dimer in the eukaryal nucleosome (Luger et al., 1997). In rHMfB, K53 is believed to be salt-linked to the DNA phosphodiester backbone. Conserved salt bridges from lysines in L2 to DNA phosphates in the minor groove also occur in the eukaryal nucleosome but at equivalent positions to G51 in most of the archaeal histones (Luger et al., 1997). In rHMfB, the role of K56 in DNA binding remains unclear. A homologous interaction is absent in the eukaryal nucleosome. In rHMfB, V20 in L1 of one monomer and I55 in L2 of the opposite monomer interact hydrophobically to stabilize the loop structures in a dimer, and consequently have an indirect effect on DNA binding. This interaction is also found at homologous residues in the L1 and L2 regions of the eukaryal histones (Luger et al., 1997).

The crystal structure of the eukaryal nucleosome shows more DNA phosphate contacts are made from the histone fold L1-L2 interfaces than from the α1α1 helices at the N termini (Luger et al., 1998a). Based on the homology of the archaeal and eukaryal histone folds and the residues involved in DNA binding, we predicted that the L1-L2 sites in the archaeal histones would also contact a larger surface area of the DNA compared to the α1α1 sites. Therefore, adding more DNA contacts to the L1-L2 region would have a better chance of fixing the position of the DNA than at the α1α1 interface. The goal of these experiments was to make archaeal histones that would bind with higher affinity to the DNA and generate more stable nucleosomes. We hypothesized that DNA binding in the L1-L2a and L2-L1a regions of rHMfB dimers could be further enhanced by the addition of more positively-charged residues that could reach from L1-L2 interfaces into the minor groove and contact the DNA phosphodiester backbone. To test
this hypothesis, we added a positively charged lysine residue independently in L1 and L2, and also constructed a variant with an additional lysine introduced into both the L1 and L2 regions.

**MATERIALS AND METHODS**

**Chemicals and kits**

Chemicals were purchased from Sigma Chemical Co. (St. Louis, MO), QIAGEN DNA prep kits from QIAGEN, Inc. (Valencia, CA), radioactively-labeled chemicals from ICN (Costa Mesa, CA), oligonucleotides from Ransom Hill Bioscience (Ramona, CA), *Pfu* DNA polymerase and *E. coli* XL1 Blue supercompetent cells from Stratagene Inc. (La Jolla, CA). All restriction and DNA modifying enzymes were purchased from Gibco-BRL (Life Technologies Inc.; Gaithersburg, MD), New England BioLabs (NEB, Beverly, MA) or Boehringer Mannheim (Indianapolis, IN). Chicken blood topoisomerase I was a gift from K. Sandman.

**Site-specific mutagenesis and purification of rHMfB variants**

The substrate used to construct the (E18K+G51K) variant was the pKS323 construct that previously had a mutation resulting in a G51K substitution. In *hmfB*, the (E18K+G51K) variant was generated using the same mutagenic primers that were used to create the E18K single variant, described earlier (Table 2.1) and the presence of the desired mutations was confirmed by DNA sequencing. The (E18K+G51K) variant was
expressed in E. coli JM105 as previously described and purified as specified for the single variants except that the E. coli extract was loaded onto a heparin-sepharose column in 45 mM potassium citrate, 50 mM Tris-HCl (pH 8.0) and the variant was eluted using a linear 45 to 200 mM potassium citrate gradient in 50 mM Tris-HCl (pH 8.0).

CD analyses of rHMfB variants

Circular dichroism spectra of the rHMfB-E18K, G51K and (E18K+G51K) variants were measured using an AVIV 62A-DS spectropolarimeter (Aviv, Lakewood, NJA) as described in Chapter 2. CD measurements at 222 nm ($\theta_{222}$) were used to calculate % folding of the rHMfB variants, based on the $\theta_{222}$ value of wt rHMfB being equivalent to 100%.

Agarose EMSA with the higher affinity rHMfB variants

Agarose gel mobility shift assays were performed with EcoRI-linearized pBR322 DNA molecules and the rHMfB-E18K, G51K and (E18K+G51K) variants as described in Chapter 2. The rHMfB variants (50-450 ng) were incubated with the EcoRI-linearized pBR322 DNA (50 ng) at room temperature (~20°C) for 15 min and the complexes were run on 0.8% (w/v) agarose gels, as described in Chapter 2. The reaction products were visualized following staining with ethidium bromide.
Polyacrylamide EMSA with the higher affinity rHMfB variants and 89 bp $^{32}$P-pBR322 DNA

A similar protocol was used for the polyacrylamide gel shift assays as described for the rHMfB variants in Chapter 2. Increasing amounts of the rHMfB-E18K, G51K and (E18K+G51K) variants (10-175 ng) were mixed with an 89 bp $^{32}$P-pBR322 DNA molecule (0.1 ng) at 25°C for 30 min as described in Chapter 2. Controls included histone-free DNA (0.1 ng) as well as reactions mixtures containing the DNA (0.1 ng) and wt rHMfB (75 ng). Electrophoresis was performed through 8% (w/v) polyacrylamide (acrylamide/bis-acrylamide, 8:0.13 (w/w) gels run in 1X TBE buffer at 8 V/cm for 1.5 h. The gels were dried and products visualized by autoradiography.

Polyacrylamide EMSA with the higher affinity rHMfB variants and 110 bp $^{32}$P-Selex clone 1 DNA

Polyacrylamide gel retardation assays were performed with the higher affinity $^{32}$P-Selex cycle 8 clone 1 DNA and the rHMfB-E18K, G51K and (E18K+G51K) variants as described in Chapter 2. Increasing amounts of the rHMfB variants (0.1-50 ng) were incubated with 0.1 ng of the Selex DNA and electrophoresis was performed through 8% (w/v) polyacrylamide (acrylamide/bis-acrylamide, 8:0.13 (w/w) gels run in 1X TBE buffer at 8 V/cm for 1.5 h.
CD assay for measurement of DNA binding of the higher affinity rHMfB variants

Circular dichroism experiments were performed with the rHMfB-G51K and (E18K+G51K) variants and the full-length EcoRI-linearized pUC19 DNA (30 μg/ml) in 100 mM KCl, 25 mM K$_3$PO$_4$ (pH 7.0) at 25°C using an AVIV 62A-DS spectropolarimeter (Aviv, Lakewood, NJ), described in Chapter 2. The spectra of the rHMfB variants with DNA as well as the rHMfB variants or DNA alone were recorded between 200 and 300 nm at 1 nm intervals. Increasing amounts of the rHMfB variants (0.1-1.3 protein: DNA mass ratios) were added to pUC19 DNA (30 μg/ml) and the CD spectra of each histone addition was measured. These data were analyzed using Kaleidagraph software. The ellipticities of the histone-DNA complexes at 275 nm were calculated as fractions of the ellipticity of the protein-free pUC19 DNA ($\theta_{275}$).

DNA circularization assays with the higher affinity rHMfB variants

The rHMfB-E18K, G51K and (E18K+G51K) variants (0.5-50 ng) were incubated with an 88 bp pLITMUS28A10 DNA molecule (0.5 ng) in ligase buffer [50 mM Tris-HCl (pH 7.8), 10 mM MgCl$_2$, 10 mM DTT, 1 mM ATP, 50 μg BSA/ml] for 25 min at 25°C as described in Chapter 2. Phage T4 DNA ligase (80 U) (New England BioLabs, Beverly, MA) was added and incubation was continued at 16°C overnight (> 12 hours). The reactions were deproteinized and the products were separated by electrophoresis in 8% native polyacrylamide gels (acrylamide/bis-acrylamide, 8:0.13 w/w) at 8 V/cm for 2 hours. The products were visualized by autoradiography. Ligation products were excised from the wet gels and digested with HindIII or BsiWI. Following
deproteinization of the reactions, the products were electrophoresed through native 8%
polyacrylamide (acrylamide/bis-acrylamide, 8:0.13 w/w) gels as described in Chapter 2.
The gels were dried and the products were visualized by autoradiography (data not
shown).

DNA topology studies of the higher affinity rHMfB variants

Preparation of DNA substrates for topology assays

Negatively supercoiled pUC19 plasmid DNA was purified from *E. coli* DH5α
using a QIAGEN Megaprep DNA kit from QIAGEN, Inc. (Valencia, CA), quantitated by
UV spectroscopy, and 50 μg aliquots were relaxed using chicken blood topoisomerase I
(cbt)(15 μl) in 250 μl reaction mixtures containing 1X TPEN buffer [10 mM Tris-HCl
(pH 8.0), 2 mM K₃PO₄, 1 mM EDTA₄(pH 8.0) and 50 mM NaCl] at 37°C for 30 min. The
amount of cbt needed to relax 1 μg of DNA was determined empirically for each
preparation of cbt. Proteins were removed by phenol-chloroform and by chloroform
extractions, and the resulting DNA was precipitated, resuspended in H₂O and quantitated
by UV spectroscopy.

Topology assays of complexes with the higher affinity rHMfB variants

Topology assays were performed using the protocol by Musgrave *et al.*, (1991)
with modifications provided by K. Sandman. Relaxed pUC19 DNA (500 ng) was
incubated with increasing amounts of rHMfB or the rHMfB variants (200-750 ng) in 1X
TPEN buffer in a 20 μl reaction at 37°C for 10 min. After 10 min, 10 μl of a 3X
topoisomerase mixture [10 μl of 10X TPEN buffer, 12 μl of 1 M Tris-HCl (pH 8.0), 1 μl of 0.5 M EDTA (pH 8.0), 60 μl glycerol and 15 μl of cbt in a total volume of 100 μl] was added and incubation continued at 37°C for 30 min. The amount of cbt needed to remove any plectonemic supercoils in the histone-free DNA was determined for each preparation of cbt. The products were deproteinized by addition of 30 μl of a 2X proteinase K mixture [16 μl of proteinase K (5 mg/ml), 16 μl of 0.5 M EDTA (pH 8.0), 40 μl of 20% SDS in a total volume of 800 μl] and incubation at 56°C for 30 min, followed by phenol:chloroform and chloroform extractions. The resulting DNA was ethanol precipitated, resuspended in 12 μl sample buffer (8% sucrose (w/v), 0.05% bromophenol blue) and an aliquot (6 μl) was subjected to electrophoresis in a 1.5% (w/v) agarose gel (20 x 25 cm) in TBE buffer [90 mM Tris-borate, 2 mM EDTA (pH 8.0)] at 40V for 16-18 h, or in a (11 x 14 cm) agarose gel at 40V for 20-24 h. The remaining DNA was stored at −20°C for 2-D agarose gel electrophoresis analysis. The products were visualized with UV light after staining with ethidium bromide.

2-D agarose gel electrophoresis to distinguish between negatively and positively supercoiled topoisomers

DNA samples, generated as described above, were subjected to agarose gel electrophoresis in the first dimension using a 20 x 25 cm agarose gel in lanes 2, 6, 10 and 14 at 40V for 18 h. The bottom 5 cm of the gel was removed, the gel was rotated 90°, sealed in the tray with 1% agarose in TBE and soaked in TBE buffer containing 15 μg ethidium bromide per L buffer for 30 min. Electrophoresis was then continued in the
Construction, purification and protein folding studies of rHMfB higher affinity variants

The rHMfB-E18K, G51K and (E18K+G51K) variants were constructed to include positively-charged lysine residues in the L1 and L2 regions. The E18K variant accumulated as a folded protein in *E. coli*, despite the fact that E18 apparently forms an intermolecular salt bridge in wt rHMfB with K53 in L2 (Starich *et al.*, 1996, Decanniere *et al.*, 2000). However, other archaeal histones do have glutamine, proline, serine, alanine or phenylalanine residues at position 18, although almost all have a lysine residue at position 53. HHbC from *Halobacterium* NRC-1, however, has an arginine at position 53 (Figure 1.5). The L1-L2 salt bridge interaction in wt rHMfB is therefore not conserved in all archaeal histones and is apparently not essential for rHMfB folding. As in rHMfB, a glycine is present at position 51 in all archaeal histones except those from *M. jannaschii* and HMvA from *M. voltae* which naturally have a lysine at this position (Figure 1.5). The combined rHMfB-(E18K+G51K) variant had both changes in the L1 and L2 regions. These variants were purified from *E. coli* using heparin affinity chromatography as described in Chapter 1 and a sample elution profile is shown in Figure 3.1. These variants bind more strongly to heparin and higher concentrations of potassium citrate (~175-200 mM) are needed to elute these proteins from the heparin columns than wt.
rHMfB (−100-150 mM). Circular dichroism spectra showed that purified preparations of these variants had > 90% of the α-helical content of wt rHMfB (Figures 3.2, 3.3 and 3.4).

Agarose and polyacrylamide EMSAs with the higher affinity variants

As described in Chapter 2, complexes formed by the rHMfB-E18K variant with EcoRI-linearized pBR322 DNA (~ 4 Kb) migrated faster through agarose gels than complexes formed by wt rHMfB, suggesting that the E18K variant generated more compact complexes. However, the complexes formed by the G51K variant migrated still faster than those with the E18K variant, and complexes formed by the (E18K+G51K) variant migrated even faster than those formed by either of the single substitution variants (Figure 3.5). Polyacrylamide EMSAs revealed that only 50 ng of the G51K and the (E18K+G51K) variants were needed to form complexes with 89 bp DNA substrate molecules, whereas 100 ng of the E18K variant and 150 ng of wt rHMfB were required for the same result (Figure 3.6).

Polyacrylamide EMSAs were also performed using the 110 bp Selex cycle 8 clone 1 DNA. Wt rHMfB has an ~ 10-fold higher affinity for this DNA when compared to a 110 bp pLITMUS 28 DNA molecule (K. Bailey, 2000) and the E18K, G51K and (E18K+G51K) variants had 10, 20 and 20-fold higher affinities, respectively, for the Selex DNA than wt rHMfB (Figure 3.7).
Figure 3.1. Tricine SDS-PAGE of fractions from a heparin column used to purify rHMfB-(E18K+G51K). A sample of the *E. coli* extract that was DNase I and heat-treated and loaded onto the heparin-sepharose column was in lane 1. The proteins in the flow-through and column wash were in lanes 2 and 3, respectively, and aliquots of fractions that eluted from the column between ~100-200 mM of the potassium citrate linear gradient were in lanes 4-16. A sample of wt rHMfB was run as a control and size standard in lane 17.
Fig. 3.1.
Figure 3.2. Circular dichroism spectra of wt rHMfB and the rHMfB-E18K variant. CD spectra of these proteins solutions (200 μg/ml) were measured in 50 mM K₂SO₄, 25 mM MES (pH 6.0) between 200 and 300 nm using an Aviv 62A-DS spectropolarimeter. The wt rHMfB spectrum is shown with closed circles, while the E18K variant is shown with open circles. Measurements at 222 nm (θ₂₂₂) provide an estimate of α-helical content.
Fig. 3.2.
Figure 3.3. Circular dichroism spectra of wt rHMfB and the rHMfB-G51K variant. The spectra of these proteins solutions (200 µg/ml) were measured in 50 mM K₂SO₄, 25 mM MES (pH 6.0) between 200 and 300 nm using an Aviv 62A-DS spectropolarimeter. The wt rHMfB spectrum is shown with closed circles, while the G51K variant is shown with open circles. The α-helical content of these proteins is measured at 222 nm (θ₂₂₂).
Figure 3.4. Circular dichroism spectra of wt rHMfB and the rHMfB-(E18K+G51K) variant. An Aviv 62A-DS spectropolarimeter was used to measure the circular dichroism spectra of these protein solutions (200 μg/ml) between 200 and 300 nm in 50 mM K₂SO₄, 25 mM MES (pH 6.0). Closed circles represent the wt rHMfB, while the (E18K+G51K) variant is shown with open circles. Measurements at 222 nm (θ₂₂₂) provide an indication of α-helical content.
Fig. 3.4.
Figure 3.5. Agarose gel EMSA of complexes formed by the rHMfB-E18K, G51K and (E18K+G51K) variants. Complexes formed by wt HMfA and wt rHMfB, and by the E18K and G51K variants with the EcoRI-linearized pBR322 DNA under conditions that produced maximum gel shifts are in lanes 2-5, respectively. Wt rHMfA and rHMfB gave maximum gel shifts at histone:DNA mass ratios of 1.5 and 5, respectively, whereas the rHMfB-E18K and G51K variants gave maximum gel shifts at histone:DNA mass ratios of 6. The complexes formed by the rHMfB-E18K+G51K variant at histone:DNA mass ratios of 1, 3, 4, 5, 6, 7, 8 and 9 are shown in lanes 6-13, respectively. Histone-free pBR322 DNA is shown in lanes 1 and 14.
Fig. 3.5.
Figure 3.6. Polyacrylamide EMSA of complexes formed by the rHMfB-E18K, G51K and E18K+G51K variants. Reaction mixtures contain 0.1 ng of a $^{32}$P-labeled 89 bp restriction fragment of pBR322 DNA and 10, 25, 50, 75, 100 and 125 ng of the rHMfB variants as indicated on the top of each panel. The track marked by 0 contains histone-free DNA, while lane 2 shows complexes formed by 150 ng of wt rHMfB. The products were visualized by autoradiography, following their separation in an 8% (w/v) polyacrylamide gel at 8 V/cm for 2h.
Fig. 3.6.
Figure 3.7. Polyacrylamide EMSAs of complexes formed by wt rHMfB and rHMfB-E18K, G51K and E18K+G51K variants with the 110 bp Selex cycle 8 clone 1 DNA.

Reactions mixtures contained 0.1 ng of $^{32}$P-labeled DNA and 0, 0.1, 0.3, 0.5, 1, 5, 10, 20, 30 and 50 ng (lanes 1-9) of wt rHMfB or the rHMfB variant. The products were visualized by autoradiography after their separation in an 8% (w/v) polyacrylamide gel.
<table>
<thead>
<tr>
<th></th>
<th>wt rHMfB</th>
<th>rHMfB-E18K</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>rHMfB-G51K</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rHMfB-E18K+G51K</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3.7.
Circular dichroism assays of rHMfB higher affinity variant interactions with DNA

The ellipticity of the EcoRI-linearized pUC19 DNA at 275 nm was reduced by binding of the G51K and E18K variants (Figures 3.8 and 3.9). Lower amounts of (E18K+G51K) and G51K:DNA were needed to obtain the same reduction in ellipticity as obtained with wt rHMfB (Figure 3.10). For example, at a histone: DNA mass ratio of ~ 0.8:1 (w/w), the $\theta_{275}$ signal of the DNA was reduced by 50% with wt rHMfB, whereas a 50% reduction in $\theta_{275}$ occurred at a mass ratio of ~ 0.4:1 (w/w) with the G51K and the (E18K+G51K) variants.

DNA circularization assays with the higher affinity rHMfB variants

The abilities of wt rHMfB and the rHMfB-E18K, G51K and (E18K+G51K) variants to facilitate ligase-catalyzed circularization of an 88 bp DNA were compared. Monomer circles were generated in the presence of wt rHMfB at 25 ng and 50 ng of G51K, although fewer circles were produced with the G51K variant. In the presence of the E18K variant, fewer circles were produced than with G51K at 10 and 25 ng amounts and none with greater than 50 ng of E18K. The inhibition of circularization at higher histone concentrations appears to result from an inhibition of DNA ligase. This was also observed with wt rHMfB and the G51K variant in the presence of > 100 ng histone amounts (data not shown). In contrast, regardless of the histone: DNA ratio, monomer circles were never generated in the presence of the rHMfB-(E18K+G51K) variant (Figure 3.11).
Figure 3.8. Circular dichroism assays of rHMfB-G51K variant binding to pUC19 DNA. CD spectra of histone-free EcoRI-linearized pUC19 DNA (30 μg/ml) solution and mixtures of 0.1, 0.2, 0.3, 0.4, 0.6 and 0.8 (G51K variant:DNA) mass ratios were recorded between 200 and 320 nm. The CD intensities shown are the differences in ellipticity between left ($\varepsilon_L$) and right ($\varepsilon_R$) circularly polarized light.
Fig. 3.8.
Figure 3.9. Circular dichroism spectra of the rHMFb-(E18K+G51K) variant binding to pUC19 DNA. The changes in ellipticity of a histone-free pUC19 DNA (30 μg/ml) solution and mixtures of 0.1, 0.2, 0.4, 0.6, 0.8 and 1.0 (histone: DNA) mass ratios of the (E18K+G51K) variant were measured by circular dichroism. CD intensities shown are differences in ellipticity between left (ε_L) and right (ε_R) circularly polarized light.
Fig. 3.9.
Figure 3.10. Decrease in the ellipticity of pUC19 DNA at 275 nm resulting from addition of increasing amounts of wt rHMfB (●), G51K (□) and E18K+G51K (*) variants. The results are expressed as a fraction of the ellipticity at 275 nm of the histone-free pUC19 DNA.
Fig. 3.10.
Figure 3.11. Ligase-catalyzed DNA circularization assays of DNA binding by wt rHMfB and the rHMfB-E18K, G51K and E18K+G51K variants. A $^{32}$P-labeled 88 bp restriction fragment (0.5 ng) was mixed either with wt rHMfB or the variant listed using the amounts (ng) of histone shown above each track, T4 DNA ligase was added and the mixture was incubated at 16°C for >12 h. The reactions products were deproteinized and separated by electrophoresis through an 8% (w/v) polyacrylamide gel run at 8 V/cm for 2h. The products were visualized by autoradiography. Lanes A contain untreated DNA, lanes B contain DNA exposed only to T4 ligase, and lanes C contain DNA exposed only to 10 ng of wt rHMfB or the variant shown. The ladder of tandemly ligated 88-mers, generated in the absence of histone, and dimer circles (DC) are shown in track B. In the presence of 25 ng wt rHMfB, E18K and G51K the predominant ligation product was monomer circles (MC) (Bailey et al., 1999).
Fig. 3.11.
DNA topology studies with the higher affinity rHMfB variants

Archaeal histones constrain DNA in both positive and negative supercoils (Musgrave et al., 1991, Sandman et al., 1994) and topology studies were performed with the rHMfB-E18K, G51K and (E18K+G51K) variants with relaxed pUC19 DNA (Figures 3.12 and 3.13). Under the 2D electrophoresis conditions used, topoisomers are separated in the first dimension by linking number, with increasing linking number resulting in faster mobility. Then after ethidium bromide addition, all molecules become more positively supercoiled and positive and negative topoisomers with the same linking number can be separated by electrophoresis in the second dimension. Slower-migrating negatively supercoiled topoisomers form the left arm of the arch with faster-migrating positively supercoiled topoisomers forming the right arm of the arch. Relaxed DNA molecules are located at the top of the arch. Consistent with previous results (Musgrave et al., 1991), the switch from negative to positive supercoiling of pUC19 DNA occurs at histone:DNA mass ratios of ~0.5:1 (w/w) for wt rHMfB (Figure 3.12). Very similar results were obtained for the rHMfB-E18K and G51K variants with switches from negative to positive supercoiling at histone:DNA mass ratio of 0.6:1. In both cases, the number of positive topoisomers decreased at mass ratios > 0.8:1 and negative topoisomers reappeared (Figures 3.12 and 3.13). The reappearance of negative DNA topoisomers was also observed previously with wt rHMfB at high histone:DNA ratios, and this probably reflects the inability of topoisomerase to access the DNA substrates in the presence of large amounts of bound histones (Musgrave et al., 1991). Binding by the (E18K+G51K) variant, however, resulted predominantly in negative supercoiling with
Figure 3.12. Topology assays of the DNA in complexes formed by wt rHMfB and the rHMfB-E18K variant. Relaxed pUC19 DNA was incubated with increasing amounts of histone (histone:DNA mass ratios (w/w) of 0.4, 0.6, 0.8, 1, 1.1, 1.2, 1.3, 1.4 and 1.5) and the complexes formed were treated with cbt, deproteinized and the resulting DNA molecules were subjected to electrophoresis through 1.5 % (w/v) agarose gels. These complexes are shown in lanes 2-10, respectively. Histone-free relaxed DNA was run in lane 1 (upper gel). Samples of the topoisomers formed by the wt rHMfB and the E18K variant at 0.4, 0.6, 0.8 and 1:1 mass ratios (tracks 1-4) were subjected to 2D electrophoresis (bottom gels). The left arm of an arch results from negatively supercoiled topoisomers and the right arm from positively supercoiled molecules. The band at the apex of each arch results from relaxed DNA molecules.
Fig. 3.12.
Figure 3.13. Topology assays of the DNA in complexes formed by the rHMfB-G51K and (E18K+G51K) variants. Relaxed pUC19 DNA was incubated with increasing amounts of histone (histone:DNA mass ratios (w/w) of 0.4, 0.6, 0.8, 1, 1.1, 1.2, 1.3, 1.4 and 1.5) and the complexes formed were treated with cbt, deproteinized, and the resulting DNA molecules were subjected to electrophoresis through 1.5 % (w/v) agarose gels. These complexes are shown in lanes 2-10, respectively. Histone-free relaxed DNA was run in lane 1 (upper gel). Samples of the topoisomers formed by the G51K and (E18K + G51K) variants at 0.4, 0.6, 0.8 and 1:1 mass ratios (tracks 1-4) were subjected to 2D electrophoresis (bottom gels). The left arm of an arch results from negatively supercoiled topoisomers, and the right arm from positively supercoiled molecules. The band at the apex of each arch results from relaxed DNA molecules.
Fig. 3.13.
very few positive supercoils formed and, furthermore, the number of positive supercoils decreased at higher histone:DNA mass ratios (Figure 3.13).

**DISCUSSION**

The DNA binding studies described in this Chapter demonstrate that the rHMfB-E18K, G51K and (E18K+G51K) variants bind to DNA with a higher affinity than wt rHMfB. Complexes formed by these variants with linear pBR322 DNA migrated faster through agarose gels than complexes formed by wt rHMfB with complex mobilities increasing in the following order: wt rHMfB, E18K, G51K, (E18K+G51K) (Figure 3.5). These variants must therefore produce more compact complexes, consistent with tighter binding or there might have been greater interaction between multiple nucleosomes. Archaeal nucleosomes formed by the rHMfB-E18K variants would contain tetramers with 4 additional positively-charged lysine residues in the L1 DNA binding sites. The K18 residue is adjacent to R19 which, based on the eukaryal nucleosome, most likely contacts the DNA phosphodiester backbone in the minor groove. It seems likely therefore that the K18 residues also bind DNA in the minor groove providing additional hydrogen bond contacts or salt bridge interactions with the DNA. Similarly, the rHMfB-G51K variant was expected to generate a compact nucleosome like the E18K variant because of the addition of K51 next to two positively-charged R52K53 residues. Even though R52 does not contact the DNA directly, it is involved in maintaining the stability of the L1-L2 interface, while K53 forms salt links to the DNA phosphates. Therefore, the
rHMfB-G51K variant tetramer, like the E18K variant, had 8 sites that are important for DNA binding in L2. The G51K variant, however, was modeled after the 4 eukaryal core histones that naturally contain lysines at this position in L2 namely H3-K115, H4-K77, H2A-K74 and H2B-K82. These residues are important in the eukaryal nucleosome because they extend the DNA contacts from the histone folds by approximately half a helical turn (4 bp) on each end of the dimer by salt bridge interactions (Figures 1.7 and 1.8) (Luger et al., 1997). Therefore, the increased compaction observed with the G51K-containing complexes compared to the E18K-containing complexes in the agarose EMSA is most likely due to the additional 12 bp of DNA contacted by a G51K-containing nucleosome. Note that 4 bp of DNA are believed to be overlapping between two such contacts in a tetramer. It is also interesting that all 5 histones from M. jannaschii and HMvA from M. voltae also have natural lysine residues at position 51. In the combined rHMfB-(E18K+G51K) variant, 4 positively-charged residues might be sufficient to anchor the DNA in each L1-L2 region and hence provide 16 DNA interactions in a nucleosome (Figure 3.14). In H4, K44R45 and K77R78K79 are at the equivalent positions in L1 and L2, respectively, and contact the DNA (Luger et al., 1997).

The results of the polyacrylamide gel retardation assays with both the 89 bp restriction fragment of pBR322 and the 110 bp Selex cycle 8 clone 1 DNA show that the rHMfB-E18K, G51K and (E18K+G51K) variants bind with a higher affinity to DNA than wt rHMfB (Figures 3.6 and 3.7). The G51K and the (E18K+G51K) variants both have almost a 3-fold higher affinity for the pBR322 DNA than wt rHMfB, while the E18K variant has a 1.5-fold higher affinity for this DNA than wt rHMfB. Similarly, the G51K
and the (E18K+G51K) variants both have a 20-fold higher affinity for the Selex cycle 8 clone 1 DNA than wt rHMfB, while the E18K variant has a 10-fold higher affinity for this DNA than wt rHMfB. Therefore, the G51K and the combined (E18K+G51K) variants have an ~ 2-fold higher affinity for both the plasmid DNA and the Selex DNA compared to the E18K variant. In solution, the rHMfB-G51K and (E18K+G51K) variants cause a proportional reduction in the ellipticity of the DNA at 275 nm as wt rHMfB but at a 2-fold lower histone amount in circular dichroism assays (Figure 3.10). Hence these results are consistent with higher affinity binding of these variants that was also observed in the gel shift assays with the other DNA fragments (Figures 3.6 and 3.7) and are believed to result from the increased number of bp of DNA contacted by these variants in a nucleosome. Specifically, complexes containing the G51K and (E18K+G51K) variants have almost the same affinity for different DNA fragments, which is ~ 2-fold higher than the affinity of the E18K variant. Therefore, these 4 lysines that contact the additional 12 bp of DNA in the G51K and (E18K+G51K) variant-containing nucleosomes contribute more to generating higher affinity molecules than the 4 lysines in the E18K-containing nucleosomes.

In the ligase-catalyzed DNA circularization assays, however, fewer monomer circles were evident for the G51K variant than wt rHMfB and even fewer circles were generated with the E18K variant (Figure 3.11). On the other hand, the combined (E18K+G51K) variant did not produce any monomer circles at any of the histone concentrations that were tested. One can speculate that the rHMfB-(E18K+G51K) variant forms much tighter complexes with the DNA at the L1-L2 sites that prevents the
Figure 3.14. Illustration of the clusters of positively-charged residues in the L1 and L2 sites in the rHMfB-(E18K+G51K) variant tetramer. In wt rHMfB, R19 in L1 and K53 in L2 contact the DNA. In the (E18K+G51K) variant, 2 additional positively-charged residues are present in each L1-L2 region that are predicted to bind to the DNA. Note that the space-fillings of the residues (not substitutions) are shown. This figure was generated by RasMol (R. Sayle, Molecular visualization program, RasMol 2.6, http://www.umass.edu/microbio/rasmol/distrib/html).
Fig. 3.14.
ends of the DNA from reaching each other because they are less mobile and are consequently not sealed by DNA ligase (Figure 3.15).

Since the orientation of the tetramer on the DNA is still unknown, two models are provided that might explain why ligation is not observed with the (E18K+G51K) variant and the 88 bp pLITMUS DNA in the ligase-induced circularization assays. The crystal structure of the eukaryal nucleosome shows 27-28 bp of DNA being contacted by each histone fold dimer (Luger et al., 1997). If the additional 12 bp of DNA that are contacted by the 4 lysine residues at position 51 are considered in these models, then a total of 66-68 bp of DNA are contacted by this variant (again, assuming a 4 bp overlap in contacts where the two L2 regions are in close proximity in a tetramer). Therefore, 20-22 bp of DNA would not be contacted in a G51K variant-containing nucleosome with this 88 bp pLITMUS DNA fragment. If the nucleosome is symmetric, then one could expect that 10-11 bp of DNA are present at each end of the structure. Otherwise, 20-22 bp of DNA would be divided unevenly between the two ends of the molecule. However, when the E18K substitution is combined with G51K to generate the (E18K+G51K) variant, the 16 DNA contacts from positively charged residues in L1 and L2 regions in a nucleosome might tether the DNA at these sites and, consequently, prevent the ends from reaching each other because they are held in different planar orientations. Alternatively, if the dimer:dimer interface of a tetramer is positioned asymmetrically on the DNA, the ends of the DNA might not reach each other because the opening of the horseshoe-shaped tetramer must be accommodated within the DNA. Tight binding to DNA at positions 18 and 51 in a tetramer would render
Figure 3.15. Why does DNA binding by the rHMfB-E18K+G51K variant prevent ligase-catalyzed DNA circularization? The rHMfB tetramer models were generated by RasMol (R. Sayle, Molecular visualization program, RasMol 2.6, http://www.umass.edu/microbio/rasmol/distrib/html) and a hypothetical DNA fragment was drawn around the protein core. Wild-type side chains are highlighted in this figure. The upper image illustrates an E18K+G51K tetramer with the dimer:dimer interface at the center of the DNA, whereas the figure on the bottom has the dimer asymmetrically positioned on the DNA. The location of DNA contacts made at positions 18 and 51 are indicated. The ends of the DNA molecule might be prevented from coming into contact (and therefore available for ligation) by being held in a very tight configuration by the 8 additional lysine-DNA binding sites predicted to be present in such archaeal nucleosomes. The ends of the DNA molecule are oriented in different planes, in the upper illustration, or are too far apart from each other because the opening of the horseshoe-shaped tetramer must be accommodated within the DNA fragment as shown in the image on the bottom.
Fig. 3.15.
the ends of the DNA inflexible and, therefore, would be unable to reach each other. Hence, the ends would not be expected to be sealed by DNA ligase.

Topology studies with wt rHMfB and relaxed circular template DNA show a switch from negative to positive supercoiling at histone:DNA mass ratios greater than 0.6:1 (w/w) (Sandman et al., 1994), and it has been suggested that this change results from a rotation in the orientation of one histone dimer relative to the second dimer within the histone tetramer core of an archaeal nucleosome (Reeve et al., 1997). The dimer:dimer interface does not involve L1/L2 interfaces and residues that directly contact DNA, and therefore as expected, topology experiments with the E18K and G51K variants revealed that they also underwent a switch from negative to positive supercoiling with increasing histone:DNA mass ratios (Figures 3.12 and 3.13). As also observed with wt rHMfB, negative supercoiling was observed again at higher histone:DNA mass ratios (> 0.8:1 (w/w)), which reflects a shortcoming of the assay rather than a physiological effect, namely topoisomerase I cannot access the DNA in histone-saturated complexes (Musgrave et al., 1991). Consistent with this idea is the fact that the (E18K+G51K) combined variant produced mostly negative supercoils at all the histone:DNA mass ratios tested (Figure 3.13). Very tight binding of DNA affects the ability of the tetramer interface to move within a nucleosome and, consequently, very few positive supercoils are generated. In fact, the number of positive supercoils decreases with increasing amounts of the (E18K+G51K) variant, suggesting that the two dimers have difficulty reorienting in tightly constrained structures. These results provide more evidence of greater compaction in nucleosomes containing the rHMfB-(E18K+G51K) variant.
While these higher affinity rHMfB variants are interesting in confirming structural prediction *in vitro*, they might not be useful *in vivo*. They might block gene expression as they could sequester DNA sites that are normally accessed by transcriptional machinery. The "sites exposure model" suggests that DNA is more often transiently exposed at the ends and the dyad of the DNA, than at other sites, by uncoiling of the DNA. This allows transcription activators to gain access to specific DNA sites (Polach and Widom, 1995 and 1996). The higher affinity rHMfB variants would be predicted to hold the DNA in a tighter configuration at these sites and consequently prevent the DNA from unwinding, which would prevent transcription from proceeding.

It is certainly interesting, however, that nature selected for a lysine at position 51 in the histones from the hyperthermophiles, *M. jannaschii* and *M. voltae*. It is possible that these histones were an intermediate step in the evolution of the 4 eukaryal core histones. Consistent with this hypothesis is the fact that one of the histones from *M. jannaschii*, MJ1647, and HMvA from *M. voltae* contain C terminal extensions to their histone folds that are also present in the eukaryl core histones but absent in all other known archaeal histones (Li *et al.*, 2000).
CHAPTER 4
COMPARATIVE STUDIES OF DNA AFFINITIES OF rHMfA, rHMfB AND rHMfB VARIANTS

INTRODUCTION

*M. fervidus* cells contain two histones, HMfA and HMfB, that exhibit growth-phase dependent differences in synthesis, *in vivo*, suggesting different roles for these proteins in genome compaction in *M. fervidus*. HMfA comprises as much as 80% of the total HMf population in exponentially growing cells, whereas HMfB abundance increases to ~50% of the HMf population in stationary phase cells (Sandman et al., 1994).

Although all of the DNA binding residues, identified in Chapters 2 and 3, are conserved in HMfA, rHMfB forms complexes with DNA that migrate faster during agarose electrophoresis than complexes formed by rHMfA. It therefore appears that rHMfB-containing complexes are more compact than rHMfA-containing complexes but gel shifts are observed at a lower histone:DNA ratio with rHMfA, and the maximum gel shift occurs at a lower histone:DNA ratio with rHMfA than with rHMfB (Sandman et al., 1994). Furthermore, topology studies have established that HMfA complexes change from negative to positive supercoiling, *in vitro*, at a lower histone:DNA ratio than rHMfB-containing complexes, but rHMfB binding introduces more positive supercoils.
than rHMfA binding. When these complexes were visualized by EM, rHMfB introduced more nucleosomes into a closed circular plasmid DNA than rHMfA at the same histone:DNA ratio (Sandman et al., 1994). Therefore, it is clear that HMfA and HMfB bind and compact DNA differently and these results are consistent with HMfB playing a greater role in archiving a primarily inactive genome in stationary phase and HMfA maintaining a reduced amount of genome compaction, appropriate for increased activity during exponential growth (Sandman et al., 1994). They also indicate that HMfA generally has a lower affinity for DNA than HMfB.

Previous studies as described in Chapters 2 and 3 have also shown that rHMfB variants differ from wt rHMfB in DNA binding. rHMfB variants with substitutions at residues K13, R19 and T54 have reduced affinities for DNA or exhibit no DNA binding based on EMSAs. Some variants had apparently lost their DNA binding ability in these assays. However, further investigation using ligase circularization and CD assays revealed that they still retained their DNA binding ability but that the complexes formed apparently did not survive gel electrophoresis. The data obtained with HMfA and HMfB and with HMfB variants has indicated both qualitative and quantitative differences in DNA binding that depend on both the protein and DNA molecules. We have investigated the use of surface plasmon resonance to pursue these observations qualitatively, and to obtain affinity constants for archaeal histones.

The rHMfB-E18K, G51K and (E18K+G51K) variants, on the other hand, formed complexes with the DNA that migrated faster than wt rHMfB in the agarose gel shift assays, suggesting that more compact nucleosomes are generated by these variants
These variants also produced polyacrylamide gel shifts at lower histone:DNA ratios than wt rHMfB and also reduced the ellipticity of the DNA, at 275 nm, at a lower histone:DNA ratios than wt rHMfB in CD assays. These results demonstrate that these variants bind to DNA with a higher affinity and generate more compact nucleosomes than wt rHMfB.

The next step, therefore, was to compare the relative affinities of these complexes and to calculate the affinity constants of these binding interactions. We developed an assay using surface plasmon resonance to measure archaeal histone-DNA binding under equilibrium conditions. This method used a stationary DNA molecule, which was attached to a sensor surface on one end, and archaeal histones were passed over the sensor surface in the mobile phase. An isothermal titration calorimetry assay was also developed to measure the affinity of archaeal histone-DNA binding, in solution, under equilibrium conditions. This technique has previously been used to measure non-covalent interactions between biomolecules such as protein:DNA complexes by measuring the heat effects of their interaction. Since most interactions involve the uptake or release of heat energy, microcalorimetry measures these heat effects to probe the magnitude of the interaction and larger heat effects indicate greater interaction between molecules (Cooper, 2000).

There has been a lot of controversy in the literature about which method is most accurate and/or valid in studying binding interactions between macromolecules. Studies with the lac operon are one such example. The binding of cAMP-ligated CRP to the promoter region of the lac operon in the activation of transcription has been extensively
studied using various techniques, and the binding affinities of the complexes vary based on length of the DNA and the assay conditions employed. For a 216 bp fragment of the lac operon, an equilibrium dissociation constant \( K_D = 4.3 \times 10^{-10} \) M was obtained (Vossen et al., 1997), while a \( K_D = 1.2 \times 10^{-11} \) M was obtained for a DNA molecule of 203 bp (Fried and Crothers, 1984) using EMSA and filter binding assays. With shorter DNA fragments of ~40 bp, however, these complexes gave \( K_D \) values = \( 1.7 \times 10^{-7} \) M in EMSAs (Lui-Johnson et al., 1986) and \( K_D = 3.1 \times 10^{-9} \) M in filter binding assays (Ebright et al., 1989). Isothermal titration calorimetry assays were also performed on these complexes and this method provided a \( K_D \) of \( 2.5 \times 10^{-6} \) M for binding of CRP(cAMP)\(_2\) to the 40 bp lac operon DNA (Shi et al., 1999). Hence, it is clear that the affinity of these binding interactions is dependent on the conditions used to study them. The study by Shi et al., (1999) argues that isothermal titration calorimetry is the most direct method of measuring binding interactions under equilibrium conditions compared to EMSAs and filter binding assays, which are hampered by the need for monitoring a rapid binding reaction with a few data points. Similarly, studies of Lac repressor binding to operator DNA sequences also showed differences in affinity based on experimental conditions; a \( K_D \) of \( 4.2 \times 10^{-9} \) M was calculated from EMSA data while a \( K_D \) of \( 2 \times 10^{-10} \) M was obtained from surface plasmon resonance experiments. Again these authors suggest that surface plasmon resonance experiments study binding under equilibrium conditions and in real time, while gel mobility shift assays do not (Bondeson et al., 1993). Therefore, we opted to compare histone-DNA interactions under a variety of conditions and with different DNA sequences.
MATERIALS AND METHODS

Reagents and kits

Chemicals were purchased from Sigma Chemical Co. (St. Louis, MO) unless otherwise specified. QIAGEN DNA prep kits were purchased from QIAGEN, Inc. (Valencia, CA) and radioactively-labeled chemicals from ICN (Costa Mesa, CA). Oligonucleotides were purchased from Ransom Hill.Bioscience (Ramona, CA). All restriction and DNA modifying enzymes were purchased from Gibco-BRL (Life Technologies Inc.; Gaithersburg, MD), New England BioLabs (NEB, Beverly, MA) or Boehringer Mannheim (Indianapolis, IN). The Biacore sensor CM5 chips, the amine coupling kit containing N-hydroxysuccinimide (NHS), N-ethyl-N'-(3-diethylaminopropyl)-carbodiimide (EDC) and ethanolamine-HCl, and HBS-P buffer were purchased from Biacore, Inc. (Piscataway, NJ). Streptavidin (immunopure quality) was purchased from Pierce Chemical Co. (Rockford, IL).

Construction and purification of DNA substrates for Biacore experiments

Selex clone 20 DNA was PCR amplified from a pGEM T-easy construct described by Bailey et al., (2000). A 93 bp fragment of this DNA was generated for use in the Biacore experiments by PCR amplification using a 5' biotinylated primer (20RBIO; 5' biotin-CACGAATTCGAGCTCTCTGCGGCCT) and primer 20F (5' GCTTGGATCCTGTCGGCACAGTTGA). The PCR contained 2 ng of template DNA, 10 pmol of primers 20RBIO and 20F, 200 μM dNTPs, 300 mM MgCl₂, 50 mM KCl, 20
mM Tris-HCl (pH 8.4) and 5 U Taq DNA polymerase (BRL) in a total volume of 100 μl. PCR amplification was performed using 30 cycles of 95°C for 1 min, followed by annealing at 75°C for 1 min, and extension at 72°C for 1 min. The 93 bp amplification product (Table 4.1) was purified using QIAGEN PCR purification kits and was quantitated by UV spectroscopy.

Purification of archaeal histones for Biacore experiments

Wild-type rHMfA, wt rHMfB, and the rHMfB-K13T, R19S, (R19S+T54K), (K13T+R19S+T54K), G51K and (E18K+G51K) variants were purified as described in Chapters 2 and 3, dialyzed against HBS-P buffer [150 mM NaCl, 10 mM HEPES (pH 7.5) and 0.005 % (w/v) polyoxyethylenesorbitan (P20) surfactant], quantitated by amino acid composition determinations and stored at -70°C. Concentrations are expressed in terms of histone tetramers.

Development of the surface plasmon resonance (Biacore) assay

Immobilization of the DNA on a sensor CM5 chip

Surface plasmon resonance experiments were performed using Biacore in collaboration with Dr. Z. Kelman (Univeristy of Maryland, Rockville). The Biacore 2000 instrument (Biacore Inc., Piscataway, NJ) was programmed to activate the sensor CM5 chip and bind streptavidin, with the resulting changes in the refractive index of the sensor chip being continuously monitored and recorded in a sensorogram. A constant flow of 5 μl/min of HBS-P buffer was maintained over the sensor chip CM5 surface. The sensor
Name

<table>
<thead>
<tr>
<th>DNA sequence used in experiments</th>
<th>Sequences</th>
<th>Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sclcx clone 1 (110 bp)</td>
<td>CAAGTCGATACCCGACCTTCCTGCTCCTACGCTTAATCTAAAGCCCGGTTTACTCGGAGTTTAGGTGGAAGCTGAGCATTGTCATCCTGACTGACG</td>
<td>Polyacrylamide EMSAs</td>
</tr>
<tr>
<td>Sclcx clone 20 (80 bp)</td>
<td>ATCAACCGTACCTGCGGCGCTTCTTAATTCAATTTAAAAGCCCGGTTTACTCGGAGTTTAGGTGGAAGCTGAGCATTGTCATCCTGACTGACG</td>
<td>ITC experiments</td>
</tr>
<tr>
<td>Sclcx clone 20 (92 bp)</td>
<td>GCTTCGATCCCGCCGACAGTTGAGCGATACAAAACGCGAGGAGGCTGATTATATAGATAATCAAGGGCCACAGGAGCTGAAATTAGT</td>
<td>Biacore experiments</td>
</tr>
<tr>
<td>Sclcx clone 20 (110 bp)*</td>
<td>CAAGTCGATACCCGACCTTCCTGCTCCTACGCTTAATCTAAAGCCCGGTTTACTCGGAGTTTAGGTGGAAGCTGAGCATTGTCATCCTGACTGACG</td>
<td>Polyacrylamide EMSAs</td>
</tr>
<tr>
<td>pLITMUS 28 (94 bp)</td>
<td>CGAATGGCATGCGGCGCAACGCGCTTAGACCTGAGAATCTGGAGATATGCACTACCCACAGGCCCTGGTTTAGTGAAGCTGACGCCGTTTAGTACAC</td>
<td>Biacore experiments</td>
</tr>
<tr>
<td>pLITMUS 28 (-110 bp)*</td>
<td>CGAATGGCATGCGGCGCAACGCGCTTAGACCTGAGAATCTGGAGATATGCACTACCCACAGGCCCTGGTTTAGTGAAGCTGACGCCGTTTAGTACAC</td>
<td>Polyacrylamide EMSAs</td>
</tr>
<tr>
<td>pBR322 (80 bp)</td>
<td>SphI and SstI digest</td>
<td>Polyacrylamide EMSAs</td>
</tr>
<tr>
<td>ZmaRI-linearized pUC19 (2.7 Kb)</td>
<td>used in CD assays</td>
<td>Polyacrylamide EMSAs</td>
</tr>
</tbody>
</table>

* indicates experiments that were performed by K. Bailey and C. Ulz (unpublished results)
^ DNA sequences (5' to 3') are listed under each name and the experiment it was used in is also listed

Table 4.1. DNA sequences used in different binding assays with archaeal histones
chip was activated by injection of 30 μl of 0.2 M EDC, 0.05 M NHS. 30 μl of Streptavidin (200 μg/ml in 10 mM sodium acetate pH 5.0) was then passed over the surface and any active esters that remained were quenched by passage of 30 μl of 1 M ethanolamine (pH 8.5) (Figure 4.1). Uncrosslinked streptavidin molecules were washed from the sensor surface by 10 μl of SDS solution [0.5% (w/v)]. The biotinylated 93 bp Selex DNA (100 μl of 5 μg/ml in HBS-P buffer) was injected and allowed to flow over the chip. Loosely bound DNA molecules were removed by a 10 μl wash of 0.5% (w/v) SDS, and the density of DNA on the sensor surface was determined (Figure 4.2).

Comparative binding of archaeal histones to the immobilized DNA

The relative affinities of the archaeal histones for the Biacore-chip bound DNA were determined using sequential injections of the different proteins with a constant flow rate of 5 μl/min maintained throughout the experiment. After initial equilibration of the sensor chip with HBS-P buffer, 10 μl solutions of 25 μg/ml wt rHMfA, wt rHMfB, and of the K13T, R19S, (R19S+T54K), (K13T+R19S+T54K), G51K and (E18K+G51K) variants of rHMfB were injected sequentially over periods of 125 s. The histone solution was then replaced with HBS-P buffer, and the histone dissociation from the DNA was monitored. The sensor chip surface was regenerated by washing with 20 μl of 0.5% (w/v) SDS between sequential histone additions. Responses were measured in resonance units (RU), where 1000 RU corresponded to a density of ~1 ng/mm² of proteins on the sensor surface (Bondenson et al., 1993). The data obtained were analyzed by Biaevaluation version 3.0 software.
Figure 4.1. Sensorgram of streptavidin crosslinking to the sensor surface. A flow rate of 5 μl/min was maintained throughout the run. (A) Activation of the carboxylated dextran matrix by EDC + NHS. (B) Crosslinking of streptavidin to the activated matrix. (C) Inactivation of the unreacted groups by ethanolamine-HCl. Response is measured in resonance units (RU). The start and end of each reagent injection are indicated by vertical arrows. Differences in RU between the beginning and end of each addition reflect differences in the refractive index of the buffer and the injected sample.
Fig. 4.1.
Figure 4.2. Sensorgram showing the immobilization of the 5' biotinylated 93 bp Selex clone 20 DNA to a streptavidin-activated sensor surface. A flow rate of 5 µl/min was maintained throughout the run and response was measured in resonance units (RU). (A) 0.5% SDS wash to remove unreacted ethanolamine-HCl. (B) Injection of biotinylated DNA onto activated matrix. (C) 0.5% SDS wash to remove loosely bound DNA. (D) The amount of DNA immobilized on the sensor surface. Arrows show the start and end of each injection.
Equilibrium affinity assays of archaeal histone-DNA interactions using Biacore

Solutions with increasing concentrations (0.5-5 μM) of wt rHMfA, wt rHMfB or the G51K variant (0.1-5 μM) were injected sequentially and allowed to flow over the sensor chip CM5 at a constant rate of 5 μl/min. After ~125 s, each histone injection was replaced by HBS-P buffer and the dissociation of the histones from the DNA was monitored. The length of time of the histone injection (~125 s) was chosen to allow equilibrium binding to the DNA to be detected. The sensor chip surface was regenerated between the addition of different concentrations of each histone by a 20 μl wash using 0.5% (w/v) SDS. Interactions of the HBS-P buffer with the sensor chip were also monitored and this background curve was subtracted from the experimental data. In control experiments, the same concentrations of the different archaeal histones were passed over a sensor chip CM5 that was activated with streptavidin but did not contain the biotinylated DNA. The data of the control experiments were subtracted from the experimental data using Biaevaluation version 3.0 software. Equilibrium affinity constants were calculated by fitting the data to a steady state model namely: A + B ⇌ AB, where A denotes the concentration of the analyte (archaeal histones expressed in terms of tetramers) and B is the density of the ligand (DNA on the sensor surface). A plot of R eq vs. concentration was used to calculate the equilibrium association and dissociation constants (K A and K D) using the following equation:

\[
R_{eq} = \frac{(K_A \cdot C \cdot R_{max})}{[K_A \cdot C \cdot (n +1)]}
\]
where \( R_{eq} \) is the steady state response level

\[ C \] is the molar concentration of analyt (histone) in solution at any given time

\( R_{\text{max}} \) is response level at saturation

\( n \) is a steric interference factor

\[ K_D = \frac{1}{K_A} \]

The affinity constants are reported in terms of dissociation constants \((K_D)\), which are defined as the concentration of protein required to bind to 50% of the DNA.

**Preparation of DNA for isothermal titration calorimetry (ITC) experiments**

Selex clone 20 DNA was purified from KS1418 strain \((E. coli\#DH5\alpha\) containing pKS577). This plasmid contains 36 tandem copies of an 80 bp fragment of the Selex clone 20 DNA, cloned between EcoRV sites (K. Sandman, unpublished data) (Table 4.1). Approximately 33 mg of pKS577 DNA was purified from cells harvested from a 10L fermentor culture of \(E. coli\) KS1418 using 4 QIAGEN Giga-prep columns by following the manufacturer’s instructions. This DNA was dissolved in 48 ml of BRL REACT 2 buffer \([50 \text{ mM NaCl, 10 mM MgCl}_2, 50 \text{ mM Tris-Cl} (\text{pH 8.0})]\) containing 16,000 U of EcoRV (BRL) and the mixture was incubated at 37°C for 16 h. Completion of the restriction digestion was confirmed by electrophoresis, using an 8% polyacrylamide gel, and the 80 bp Selex clone 20 DNA was separated from the vector DNA by vector precipitation with polyethylene glycol (PEG) 8000 (Sigma Chemical Co., St. Louis, MO), as described by Tan et al., (2000). The reaction was placed on ice and solid NaCl was
added to the reaction mixture to generate a final concentration of 0.5 M, and 9 ml of a
40% (w/v) solution of PEG 8000 was then added in a drop-wise manner, with stirring, to
yield a final concentration of ~ 6% (w/v) PEG 8000. The mixture started to become
turbid due to DNA precipitation after the addition of ~ 7 ml of the PEG 8000 solution and
incubation was continued on ice for 2 h, without further stirring. The precipitated DNA
was removed by centrifugation at 10,000 x g for 30 min and the 80 bp DNA remaining in
the solution was then precipitated by adding two volumes of 95% (w/v) ethanol to the
supernatant and incubation overnight at ~20°C (~ 16h). The precipitated DNA was
pelleted by centrifugation at 10,000 x g for 30 min at 4°C and the pellet was washed and
resuspended in 300 mM NaCl, 0.1. mM EDTA and 20 mM Tris-HCl (pH 7.6). The DNA
was quantitated using UV spectroscopy. Further purification of the DNA was performed
using anion-exchange chromatography with DEAE-cellulose (Sigma Chemical Co., St.
Louis, MO). A matrix bed volume of ~ 55 ml of DEAE-cellulose was packed in a 60 ml
syringe barrel. The pH of the resin was adjusted to ~ 8.0 using 50 mM Tris-HCl (pH 8.0)
and the DNA was loaded onto the column in 50 mM Tris-HCl (pH 8.0). The column was
washed with ~125 ml of 50 mM Tris-HCl (pH 8.0) and the DNA was eluted using ~60 ml
of 0.5 M NaCl, 50 mM Tris-HCl pH (8.0). The DNA was concentrated in Millipore
Ultra-free filters (MWCO 5000), precipitated, redissolved in H2O and quantitated by UV
spectroscopy. Approximately 15 mg of the 80 bp Selex clone 20 DNA was obtained
using this procedure.
Purification of archaeal histones for ITC experiments

Wild-type rHMfB and the G51K variant were purified as described previously in Chapters 2 and 3, dialyzed against HBS-P buffer, quantitated by amino acid composition analysis, and stored at \(-70^\circ C\). Protein concentrations are expressed in terms of histone tetramers.

Development of ITC assays to evaluate archaeal histone-DNA binding

ITC experiments were performed in collaboration with Dr. F. Schwarz at the University of Maryland (Rockville) with wt rHMfB and the rHMfB-G51K variant using a Microcal VP titration calorimeter (Microcal Inc.) and following a procedure described by Wang et al., (2000). The instrument contained a reference vessel and a sample vessel in an adiabatic enclosure. A syringe was used to add ligand solutions into the sample compartment. HBS-P buffer was placed in the reference vessel. Sequential 5 µl additions of 50 µM 80 bp Selex clone 20 DNA were made to the sample compartment, with stirring, that contained 1.5 ml solutions of 5 µM wt rHMfB or G51K at 25°C. Two or three additional titrations of the DNA into the protein solutions were made after saturation in binding was reached to obtain the heats of dilution, and these values were subtracted from the experimental data. A non-linear, least squares minimization software program, [Origin 5.0 (Microcal Inc.)], was used to fit the incremental heat of the \(i\)th titration \([\Delta Q(i)]\) of the total heat \((Q_i)\) to the concentration of the total titrant, \([X_i]\), according to the following equations:
Q_i = nC_i\Delta H_0^\circ V (1 + X_0/nC_i + 1/nK_d C_i) - ((1 + X_0/nC_i + 1/nK_d C_i)^2 - 4X_0/nC_i)^{1/2})/2

and

\Delta Q(i) = Q(i) + dV_i/2V (Q(i) + Q(i - 1)) - Q(i - 1)

where

C_i is the total concentration of DNA in the sample compartment

V is the volume of the sample compartment

n is the stoichiometry of binding

K_d is the binding constant or equilibrium affinity association constant (K_a)

RESULTS

Relative affinities of archaeal histones for Selex clone 20 DNA

For kinetic experiments, only a low amount of immobilized DNA should be present on the matrix to prevent mass transport limitations of the protein in solution and to allow the reaction to reach equilibrium within a reasonable time limit (Bondeson et al., 1993). To meet this requirement, a density of 100-200 RU of the biotinylated 93 bp Selex clone 20 DNA was immobilized on the sensor surface (Figure 4.2), and the binding of wt rHMfA, wt rHMfB and the rHMfB- K13T, R19S, (R19S+T54K), (K13T+R19S+T54K), G51K and (E18K+G51K) variants was followed by monitoring surface plasmon resonance. Figures 4.3 and 4.4 show superimposed sensorgrams for these different reactions. The steady state levels demonstrate that wt rHMfB has a higher affinity for the Selex clone 20 DNA than wt rHMfA (Figure 4.3) and the rHMfB single variants K13T
Figure 4.3. Superimposed sensorgrams of archaeal histone interactions with the immobilized, biotinylated 93 bp Selex clone 20 DNA. Solutions of wt rHMfA, wt rHMfB and rHMfB-K13T, R19S, (R19S+T54K) and (K13T+R19S+T54K) variants were passed over the immobilized DNA on the sensor surface after equilibration of the surface with HBS-P buffer. After 250 s, the injection of the archaeal histone solution was replaced by the injection of HBS-P buffer, and the histone dissociation was followed. The sensor surface was regenerated by injection of 0.5% SDS between additions of different archaeal histones. The flow rate was 5 µl/min throughout the experiment and binding was measured in resonance units (RU). The data were analyzed using Biaevaluation 3.0 software.
and R19S have a much lower affinity than wt rHMfB. Surprisingly, the rHMfB-(R19S+T54K) variant bound with a slightly higher affinity than either of the single substitutions (K13T) or (R19S) but exhibited a much slower rate of complex formation (Figure 4.3). Consistent with the lack of DNA binding obtained using all the assays (Chapters 2 and 3), the rHMfB-(K13T+R19S+T54K) variant showed almost no interaction with the sensor chip-bound DNA (Figure 4.3). Also consistent with previous results, the rHMfB-G51K and (E18K+G51K) variants had much higher affinities for the immobilized Selex clone 20 DNA than wt rHMfB (Figure 4.4).

Affinity analysis of archaeal histone-DNA binding

Solutions with increasing concentrations of wt rHMfA, wt rHMfB and rHMfB-G51K were passed over the surface of sensor chips that had either the 93 bp fragment of Selex clone 20 DNA (Figures 4.5 and 4.6) or a 93 bp fragment from the pLITMUS28 polylinker region (Table 4.1 and Figures 4.7 and 4.8) immobilized on the sensor surface. Based on the sensorgrams obtained, the equilibrium affinity dissociation (K_D) constants of wt rHMfA binding to the pLITMUS DNA and Selex DNA were calculated to be 2.3 +/- 0.5 x 10^{-6} M and 3.2 +/- 0.7 x 10^{-6} M, respectively (Table 4.2). For wt rHMfB binding to the pLITMUS DNA, the K_D was 1.5 +/- 0.3 x 10^{-6} M, which is very similar to that obtained with wt rHMfA, whereas rHMfB had a much higher affinity for the Selex DNA with a K_D = 4.4 +/- 0.8 x 10^{-8} M (Table 4.2). The rHMfB-G51K variant generated a K_D of 2.6 +/- 0.1 x 10^{-7} M for binding to the pLITMUS DNA and a K_D = 1.9 +/- 1.1 x 10^{-8} M for binding to the Selex clone 20 DNA (Table 4.2).
Figure 4.4. Superimposed sensorgrams of DNA binding by wt rHMfB and the rHMfB-G51K and (E18K+G51K) variants. The experimental conditions were the same as in Figure 4.3, but the sensorgrams showing the binding of other rHMfB variants to DNA are not labeled for clarity.
Fig. 4.4.
Figure 4.5. Affinity analysis of wt rHMfA binding to the 93 bp Selex cycle 8 clone 20 DNA. The biotinylated DNA was immobilized on the sensor surface at a density of 206 RU and solutions with increasing concentrations of rHMfA (0.5-5 μM tetramers) were passed over the sensor surface for 2 min. Subsequently, histone:DNA dissociation was followed by injecting HBS-P buffer. Biaevaluation version 3.0 software was used to calculate $K_A$ ($M^{-1}$) and $K_D$ ($M$) equilibrium affinity constants.
Fig. 4.5.
Figure 4.6. Affinity analysis of wt rHMfB binding to the 93 bp Selex clone 20 DNA.

The biotinylated DNA was immobilized on the sensor surface at a density of 206 RU and increasing concentrations of rHMfB (0.5-5 μM tetramers) were passed over the sensor surface for 2 min. Subsequently, histone:DNA dissociation was followed by injecting HBS- P buffer. Biaevaluation version 3.0 software was used to calculate $K_a (M^{-1})$ and $K_D (M)$ equilibrium affinity constants.
Fig. 4.6.
Figure 4.7. Affinity analysis of wt rHMfA binding to a 93 bp fragment of pLITMUS 28 polylinker. The biotinylated DNA was immobilized on the sensor chip surface at a density of 176 RU and increasing concentrations of rHMfA (0.5-5 μM tetramers) were passed over the sensor surface for 2 min. Subsequently, histone:DNA dissociation was followed by injecting HBS-P buffer. Biaevaluation version 3.0 software was used to calculate $K_a$ (M$^{-1}$) and $K_D$ (M) equilibrium affinity constants.
Figure 4.8. Affinity analysis of wt rHMfB binding to a 93 bp fragment of pLITMUS 28 polylinker. The biotinylated DNA was immobilized on the sensor chip surface at a density of 176 RU and solutions with increasing concentrations (0.5-5 μM tetramers) of rHMfB were passed over the sensor surface for 2 min. Subsequently, histone:DNA dissociation was followed by injecting HBS- P buffer. Biacorevaluation version 3.0 software was used to calculate $K_a$ ($M^{-1}$) and $K_D$ (M) equilibrium affinity constants.
<table>
<thead>
<tr>
<th>Protein</th>
<th>Non-selected DNA*</th>
<th>K_d values (M)*</th>
<th>Selex DNA b</th>
<th>Polyacrylamide EMSA*</th>
<th>Biacore</th>
<th>Polyacrylamide EMSA*</th>
<th>Biacore</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wt rHMfA</td>
<td>1.2 x 10^{-8}</td>
<td>2.3 +/- 0.5 x 10^{-8}</td>
<td>3.2 +/- 0.7 x 10^{-8}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wt rHMfB</td>
<td>2.4 x 10^{-5}</td>
<td>1.5 +/- 0.3 x 10^{-5}</td>
<td>5.4 x 10^{-4}</td>
<td>4.4 +/- 0.8 x 10^{-3}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rHMfB-G51K</td>
<td>N/d</td>
<td>2.6 +/- 0.1 x 10^{-7}</td>
<td>N/d</td>
<td>1.9 +/- 1.1 x 10^{-8}</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* DNA molecules used in these experiments are listed in Table 4.1

^ ITC experiments were also used to calculate K_d for wt rHMfB = 6.0 +/- 0.6 x 10^{-8} and K_d = 3.8 +/- 1.0 x 10^{-8} for the G51K variant with an 80 bp fragment of Selex clone 20 DNA

A 110 bp fragment of pLITMUS 28 DNA was used for polyacrylamide EMSAs and a 93 bp pLITMUS DNA was used for Biacore experiments

b A 110 bp Selex clone 20 was used for polyacrylamide EMSAs and a 93 bp Selex clone 20 DNA was used for Biacore experiments

* These experiments were performed by K. Bailey and C. Utz (unpublished results)

Table 4.2. Affinity dissociation constants of archaeal histones binding to different DNA molecules
Isothermal titration calorimetry analysis of archaeal histone-DNA binding

A preliminary study investigated the quantitation of wt rHMfB and the G51K variant binding to the 80 bp Selex clone 20 DNA by isothermal titration calorimetry (ITC) (Table 4.1 and Figures 4.9 and 4.10). As shown, DNA binding was exothermic and, therefore, energetically favorable with the free energy of binding ($\Delta G^o$) = -41.2 kJ/mol for wt rHMfB and -42.4 kJ/mol for the G51K variant. During the saturation phase, two modes of interaction were apparent in these curves (Figures 4.9 and 4.10). Wild-type rHMfB showed a weaker interaction with the DNA between 70 and 100 min, followed by a stronger interaction between 100 and 110 min. The G51K variant, on the other hand, shows a weaker binding to DNA between 20 and 32 min and a stronger binding between 32 and 40 min. However, the basis of this phenomenon is uncertain. Possibly an HMf tetramer binds first to the 80 bp DNA and this is followed by binding of another dimer. Alternatively, there could be a change from negative to positive supercoiling due to a re-orientation of the dimer:dimer interface within a complex. Visual comparison of the binding interaction curves of wt rHMfB and the G51K variant to the Selex clone 20 DNA showed that G51K achieved saturation in DNA binding within 40 min, while wt rHMfB required ~110 min to reach saturation. An average of the two phases in these curves was fitted to the Origin one-sites binding model and used to calculate equilibrium affinity dissociation ($K_D$) constants of 6.0 +/- 0.8 x 10^{-8} M for wt rHMfB and 3.8 +/- 1.0 x 10^{-8} M for the G51K variant (Table 4.2). This model assumes that an HMf tetramer binds to an 80 bp DNA molecule.
Figure 4.9. Titration of the energy released by sequential injection of the 80 bp Selex clone 20 DNA into a solution of wild-type rHMcB. In the upper panel each peak represents an injection of DNA and the energy released following each sequential addition of 5 µl of 50 µM DNA into a 1.5 ml solution of rHMcB (5 µM tetramers) in HBS-P buffer (150 mM NaCl, 10 mM Hepes pH 7.5, 0.005% surfactant P20) was recorded. The lower panel shows the binding isotherm generated from the titration shown in the upper panel. An equilibrium association constant ($K_a$) was obtained from a least-squares fit of the data to the Origin one-site binding model.
Fig. 4.9.
Figure 4.10. The isothermal calorimetric scan of titration of 80 bp Selex clone 20 DNA into a solution containing the rHMfB-G51K variant. The conditions for this experiment were the same as those described in Figure 4.9.
Fig. 4.10.
DISCUSSION

Several assays have been used to calculate DNA binding affinities of archaean histones including EMSAs, CD assays, Biacore and ITC experiments. While CD, Biacore and ITC assays are performed under equilibrium conditions, EMSAs are not. Also, CD and ITC experiments are performed completely in solution while one end of the DNA molecule is attached to the sensor surface in the Biacore experiments, but the protein remains in solution. Gel shift assays, on the other hand, require more stable complexes to be formed that can survive for extended periods of time during electrophoresis. Therefore, different DNA fragments and a wide range of assays were used to study the binding affinities of these non-sequence specific DNA binding proteins. It is believed that the assays performed in solution and in the Biacore experiments would more likely represent conditions in vivo than EMSAs.

The relative affinities of the different archaean histones for the Selex clone 20 DNA, generated by surface plasmon resonance experiments in this Chapter, are fairly consistent with the results from the EMSA, CD and ligation assays described in Chapters 2 and 3. Wt rHMfB has a higher affinity for this DNA than wt rHMfA (Figure 4.3), with equilibrium affinity dissociation constants (K_D) of 4.4 +/- 0.8 x 10^-8 M for wt rHMfB and an ~350-fold lower dissociation constant for rHMfA (K_D = 3.2 +/- 0.7 x 10^-5 M) (Table 4.2). This was expected as rHMfB was the ligand used in the SELEX protocol that selected this DNA molecule from a large population of random synthetic DNA sequences (Bailey et al., 2000). Based on the ITC experiments, an equilibrium affinity dissociation constant (K_D) of 6.0 +/- 0.6 x 10^-8 M was obtained for wt rHMfB binding to the Selex
clone 20 DNA, which is consistent with the Biacore-based results (Table 4.2). However, wt rHMfA and wt rHMfB bound to the non-selected 93 bp fragment of pLITMUS28 DNA with similar affinities ($K_D = 2.3 \pm 0.5 \times 10^{-6}$ M for wt rHMfA and $K_D = 1.5 \pm 0.3 \times 10^{-6}$ M for wt rHMfB) (Table 4.2), demonstrating that the strength of rHMfA and wt rHMfB binding is dependent on the DNA sequence. Similarly, polyacrylamide EMSAs showed that wt rHMfB bound with an ~500-fold higher affinity to a 110 bp Selex clone 20 DNA than to a 110 bp fragment of pLITMUS 28 DNA and the two proteins bound with approximately the same affinity to pLITMUS 28 DNA (Table 4.2) (K. Bailey and C. Utz, 2000).

Based on the surface plasmon resonance experiments, the rHMfB-K13T and R19S variants had similar and much lower relative affinities for the Selex clone 20 DNA fragment than wt rHMfB (Figure 4.3). These results differ somewhat from the CD assays which indicated that K13T bound with a higher affinity than R19S although these assays employed different DNAs (Selex clone 20 for Biacore assays vs. pUC19 for CD assays), and whereas the CD measurements are performed under solution conditions, one end of the DNA molecule is tethered to the chip surface in the Biacore assays. The CD and Biacore assays of the (R19S+T54K) variant were also consistent in that both indicated lower DNA affinity relative to wt rHMfB, but differed in that the (R19S+T54K) variant apparently bound to DNA better than single-substitution variants, when assayed by Biacore, but not as well when assayed by CD. It is also worth noting that the R19S variant gave almost no DNA binding in the CD assay and no polyacrylamide EMSA, suggesting that not all the complexes formed have the same overall structure. As in all
previous assays, the variant with substitutions at all three predicted DNA binding sites, (K13T+R19S+T54K), did not bind to the DNA immobilized on the sensor chip in the Biacore assay.

Consistent with the results of the EMSAs and CD assays (Chapter 3), the rHMfB-G51K and (E18K+G51K) variants also had much higher affinities for the Selex DNA than wt rHMfB, based on Biacore and ITC experiments (Figures 4.4 and 4.10). Equilibrium affinity analysis yielded affinity constants for G51K that were ~2-fold and 12-fold higher for the Selex clone 20 DNA ($K_D = 1.9 +/- 1.1 \times 10^{-8} \text{M}$) and the reference 93 bp pLITMUS 28 DNA ($K_D = 2.6 +/- 0.1 \times 10^{-7} \text{M}$) than wt rHMfB, respectively (Table 4.2). The ITC experiments similarly generated an equilibrium dissociation constant ($K_D$) of $3.8 +/- 1.0 \times 10^{-8} \text{M}$ for G51K binding to the Selex clone 20 DNA, which is ~2-fold lower than the $K_D$ obtained using surface plasmon resonance, and is ~1.6-fold higher than the $K_D$ obtained for wt rHMfB binding to the Selex clone 20 DNA by ITC analysis (Table 4.2). This slight difference in binding constants might be due to differences in experimental conditions of the two assays, described earlier. Based on CD assays, which were also performed under equilibrium conditions, the rHMfB-G51K variant had an ~2-fold higher affinity for the full-length pUC19 DNA than did wt rHMfB (Chapter 3), whereas the Biacore experiments indicated a 12-fold higher affinity for the G51K variant with the 93 bp pLITMUS DNA than wt rHMfB (Table 4.2). On the other hand, EMSAs indicated that the G51K variant had a 3-fold higher affinity for an 89 bp fragment of pBR322 DNA but a 20-fold higher affinity for the 110 bp Selex clone 1 DNA than wt rHMfB (Chapter 3). Again, it should be emphasized that the EMSA experiments were
run under non-equilibrium conditions. Taken together, these affinity measurements confirm that the rHMfB-G51K variant does bind with a higher affinity to DNA than wt rHMfB. As mentioned earlier, the rHMfB-G51K variant has an additional positively charged lysine residue in L2 that is believed to extend the DNA binding region by 4 bp on each end of a dimer and, therefore, might anchor the DNA at the L1-L2a and L2-L1a histone-fold sites to generate more stable nucleosomes.
CHAPTER 5

GENERAL DISCUSSION

Summary

NMR and high resolution X-ray crystal structures of archaeal and eukaryal histones have shown that they have a common histone fold, comprised of a long central $\alpha2$ helix, separated from the two shorter $\alpha1$ and $\alpha3$ helices by L1 and L2 regions, respectively. In a dimer, the two $\alpha2$ helices are anti-parallel and cross each other near the center of the dimer and, consequently, L1 is in contact with L2a while L2 contacts L1a ("a" denotes the other partner in a dimer, described earlier). The N termini of the $\alpha1$ helices also interact on one end of the dimer, while the $\alpha3$ helices are not in close proximity (Starich et al., 1996, Decanniere et al., 2000, Arents and Moudrianakis, 1995, Luger et al., 1997). Archaeal histones are most similar in primary sequence to the histone-fold regions of H4 and archaeal histones form nucleosomes that share many features with the (H3+H4) tetramer, and it seems likely that the archaeal nucleosome is the ancestor of the histone tetramer-containing structure at the center of the eukaryal nucleosome (Sandman et al., 1998, Sandman and Reeve, 2000).

Archaeal nucleosomes contain a tetrameric histone core that protects ~60 bp of DNA from micrococcal nuclease digestion (Pereira et al., 1997, Reeve et al., 1997),
while ~ 73 bp of DNA are protected from micrococcal nuclease digestion by (H3+H4) tetramers (Hayes et al., 1991). Both the archaeal and eukaryal tetramers wrap DNA in an overwound conformation of ~ 10.2 bp/turn in negative and positive toroidal supercoils (Pereira and Reeve, 1999, Musgrave et al., 1991, Hamiche and Richard-Foy, 1998, Alilat et al., 1999) and although archaeal and eukaryal tetramers can bind to virtually any DNA sequence, they recognize and preferentially respond to certain positioning signals. It is also the (H3+H4) tetramer that initiates the positioning and assembly of the eukaryal nucleosome (Pereira and Reeve, 1999, Sandman and Reeve, 1999, Dong and Van Holde, 1991, Hayes et al., 1991).

The goal of this work was to identify and quantitate DNA binding residues in the archetype archaeal histone, rHMfB. We have shown that DNA binding is additive at 6 sites per rHMfB dimer and have isolated higher affinity variants of rHMfB, demonstrating that natural archaeal histones have not been optimized for high-affinity DNA binding. We have also compared and documented the differences in the DNA binding affinities of the two wild-type histones in *M. fervidus* that may relate to their different functions in vivo.

**Histone fold-DNA contacts in rHMfB (Chapter 2)**

Site-directed mutagenesis studies with rHMfB revealed that residues at the N terminus, on the surface of the α1 helix, and in the L1 and L2 loops are involved in DNA binding. Specifically, L3, R10, K13, R19, K53, T54 and K56 make direct contact with the DNA and V20 and I55 indirectly affect DNA binding by stabilizing the integrity of
L1-L2 binding regions. These residues are located at sites identified in the crystal structure of the eukaryal nucleosome as being responsible for DNA binding (Luger et al., 1997). Consistent with these results, mutational analyses of the yeast histones have shown that residues at these locations, when altered, suppress the defects in chromatin remodeling complexes (SIN phenotype), and this is thought to reflect the fact that these histones assemble only partially stabilized nucleosomes (Kruger et al., 1995, Santisteban et al., 1997, Wechser et al., 1997, Smith et al., 1996). Apparently, each of the DNA contacts independently contributes to DNA binding but, overall, DNA binding to form an archaeal nucleosome is additive. The loss of DNA binding at one location apparently does not decrease DNA interactions at other sites. Removal of three DNA contacts, namely a K13T substitution in the α1 helix, R19S in L1, and T54K in L2 was required to generate a rHMfB variant that showed no DNA binding in any of the assays employed.

In the eukaryal histones, N and C terminal extensions and tails that flank the histone fold also interact with the DNA. These tails contain the lysine residues that are the targets for regulatory acetylation and provide a supporting framework for the histone-fold regions of these proteins (Luger et al., 1997, Luger et al., 1998b), but they are not required for nucleosomal stability. They can be removed by trypsin treatments without deleterious effects on nucleosome positioning and integrity (Pruss et al., 1995). Apparently, therefore, archaeal histones that do not have such tails do not lack DNA binding contacts that are essential for nucleosome positioning and stability.
Higher affinity variants of rHMfB and their interactions with DNA (Chapter 3)

The rHMfB variants E18K, G51K and (E18K+G51K) were constructed to extend the DNA binding interface in the L1 and/or L2 regions, by adding positively-charged lysine residues at positions occupied by lysines in eukaryal H4 histones (Luger et al., 1997). The rHMfB-G51K variant was modeled after the 4 eukaryal histones that have natural lysines at this position which interact with the phosphodiester backbone through salt bridges and extend the DNA binding region by half a helical turn on each end of the dimer. Agarose EMSAs revealed that the G51K variant formed more compact structures with the linear pBR322 DNA than E18K-containing complexes that also had 4 lysine substitutions per tetramer, while the (E18K+G51K) variant formed the most compact structures in these assays. Again, this confirms that DNA binding is additive and one can improve DNA binding by introducing positively charged amino acids into L1 and L2 that interact with the phosphodiester DNA backbone. Polyacrylamide EMSAs revealed that the (E18K+G51K) and G51K variants bound with approximately the same affinity (~3-fold higher than wt rHMfB) to an 89 bp restriction fragment of pBR322 DNA and with ~20-fold greater affinity to the 110 bp Selex cycle 8 clone 20 DNA, and both variants had higher affinities for these DNAs than the E18K variant. Similarly, CD assays of G51K and (E18K+G51K) binding revealed higher affinities for DNA than wt rHMfB. However, in DNA circularization experiments, the G51K and E18K variants generated fewer circles than rHMfB, and no circles were generated in the presence of the rHMfB-(E18K+G51K) variant. The very tight binding of this variant to DNA apparently prevented the juxtapositioning of the ends of the DNA molecule needed for ligation.
(Figure 3.15). However, the DNA topology experiments showed that despite their higher affinity for DNA, complexes formed by these variants did exhibit the change from negative to positive supercoiling, although topoisomerase I access to the DNA was apparently sterically blocked at lower histone:DNA ratios than observed with wt HMfB. Also, consistent with tighter DNA binding, the (E18K+G51K) variant produced more negative than positive supercoils, suggesting that the dimer:dimer interface is somewhat restricted in its mobility within these tightly constrained complexes and consequently cannot reorient itself and generate positive supercoils (Figures 3.12 and 3.13).

Affinity studies of DNA binding of wt rHMfA, wt rHMfB and its variants (Chapter 4)

The results of the affinity studies with wt rHMfA and rHMfB using Biacore assays have shown that rHMfB has an ~350-fold higher affinity for Selex clone 20 DNA than rHMfA (Table 4.2). This was expected as wt rHMfB was used to select this DNA molecule from a random synthetic DNA library of $10^{14}$ DNA sequences (Bailey et al., 2000). On the other hand, rHMfA and rHMfB bound with similar affinities to a non-selected fragment of pLITMUS 28 DNA of the same size as the Selex DNA. Similar trends were also observed in polyacrylamide EMSAs where rHMfB bound with an ~500-fold higher affinity to the Selex clone 20 DNA than rHMfA, while both proteins had similar affinities for the pLITMUS 28 DNA (Table 4.2) (K.Bailey and C. Utz, personal communication). Therefore, one can conclude that rHMfA and rHMfB bind with varying affinities to different DNA molecules. The Selex DNA sequences contain repetitive
dinucleotide motifs that are believed to position archaeal nucleosomes \textit{in vivo} and are found in the genomes of the histone-containing Euryarchaeota but not in the Crenarchaeota and Bacteria that lack histones (Bailey \textit{et al.}, 2000). One could argue that \textit{wt HMfA} and \textit{wt HMfB} play different roles in regulating the balance between genome compaction and gene regulation in \textit{M. fervidus} by binding with varying affinities to different DNA sequences and preferentially associating with DNA sequences that are better able to accommodate to HMf structure and that serve as nucleosomal positioning elements. It would be useful to measure the affinities of HMfA, HMfB and HMfAB for DNA sequences isolated from genomic Selex experiments. This would provide a better understanding of the balance between gene regulation and DNA compaction \textit{in vivo}.

The results of surface plasmon resonance experiments with \textit{rHMfB-K13T} and R19S variants showed much lower affinities for Selex clone 20 DNA than \textit{wt rHMfB}, which is consistent with the results of CD assays also performed under equilibrium conditions (Figures 4.3 and 2.19). However, in the CD assays, K13T had a slightly higher affinity than R19S for pUC19 DNA while R19S had a higher affinity for Selex DNA than K13T in Biacore assays. These differences in affinity can be explained by differences in conditions employed in these assays. The (R19S+T54K) variant also bound with a lower affinity to Selex clone 20 DNA than \textit{wt rHMfB} in the Biacore and CD assays, but the (R19S+T54K) variant bound with a slightly higher affinity to DNA than R19S in the Biacore assay than in the CD assay. Again, these differences could be attributed to differences in experimental conditions in these assays. The \textit{rHMfB-(K13T+R19S+T54K)} variant showed no DNA binding in the Biacore experiments which
is consistent with the results of all other assays (Figure 4.3). Hence, this variant is believed to have lost all 3 DNA binding sites.

Consistent with the results of the EMSA and CD assays, the results of the Biacore experiments show much higher affinities for rHMfB-G51K and (E18K+G51K) than wt rHMfB (Figure 4.4). The $K_D$ values were $\sim$2-fold higher for the G51K variant binding to Selex clone 20 DNA than wt rHMfB, and $\sim$12-fold higher for the non-selected pLITMUS DNA fragment than wt rHMfB (Table 4.2). Isothermal titration calorimetry experiments also showed an $\sim$1.6-fold higher $K_D$ for the G51K variant than for wt rHMfB binding to the Selex clone 20 DNA (Table 4.2). CD assays, which were also performed in solution and under equilibrium conditions, provided a 2-fold higher $K_D$ for the G51K variant binding to the pUC19 DNA than for wt rHMfB, while EMSAs generated $\sim$ 3-fold and 20-fold higher $K_D$ values for G51K binding to a fragment of pBR322 DNA and Selex clone 20 DNA, respectively (Chapter 3). Thus, complexes containing the G51K variant are consequently more stable than wt rHMfB-containing complexes due to the addition of a lysine residue in each L2 region that is believed to bind the phosphodiester backbone in the minor groove and extend the DNA binding region by 4 bp at each end of the dimer.

**Overall conclusions**

The results of these experiments provide additional support to the argument that archaeal nucleosomes are structurally very similar to (H3+H4) tetrasome (Sandman and Reeve, 2000). Most of the DNA binding residues, identified in rHMfB from *M. fervidus*, are conserved in all archaeal histones and hence the basis of DNA binding by archaeal...
histones appears to have been universally conserved. This conservation is also clearly apparent in eukaryal histones H3 and H4, which are most similar to the archaeal histones and are also the most conserved of all the eukaryal histones. It seems likely that a histone-fold based system of DNA packaging existed before the separation of the Eukarya and Archaea, and that eukaryal histones have since evolved and gained N and C terminal extensions to package much larger genomes and to interact with very complex chromatin remodeling and transcriptional machineries (Sandman et al., 1998).

It seems noteworthy that in the hydrogen hypothesis, proposed by Martin and Müller (1998), Eukarya evolved from a symbiotic interaction between a hydrogen-producing α-proteobacterium and a hydrogen-consuming, presumably histone-containing methanogen. When the environment became aerobic, the methanogen became a completely dependent intracellular symbiont and the eukaryal nucleus is hypothesized to have arisen from this methanogen, while cellular metabolism appears to have been primarily derived from the bacterium. Consistent with this hypothesis, contemporary gene storage and expression systems are similar in Archaea and Eukarya, whereas cellular metabolism is more similar in Eukarya and Bacteria. The need for DNA content increased with eukaryal evolution and, therefore, a highly efficient system of DNA packaging and organization would have been necessary. The archaeal-based system of DNA wrapping appears to be the most efficient of all the prokaryotic DNA packaging systems in terms of the number of DNA bp packaged per dalton proteins and may therefore have out-competed the other bacterial histone-like DNA packaging proteins (Sandman et al., 1998).
Future experiments

In this work, reverse genetics and biochemical assays have been used to identify residues that bind DNA. Several presumably side-chains interactions have been identified but it is more difficult to identify histone fold main-chain interactions with the DNA and, for this, a high-resolution crystal structure of the archaeal nucleosome is needed. Experiments are currently underway, in collaboration with Dr. U. Heinneman (Max Delbrüch Institute for Molecular Medicine, Berlin) to generate such a structure using an 80 bp region of the Selex cycle 8 clone 20 DNA and wt rHMfB. A high resolution X-ray crystal structure of archaeal nucleosomes formed by the higher affinity rHMfB-G51K variant would also confirm that the addition of this positively charged residue extends the L1-L2 region of DNA binding to the next helical turn. Translational positioning experiments with these variants would demonstrate that additional bp of DNA are contacted by these variants than wt rHMfB to generate more stable nucleosomes.

The ligase-induced DNA circularization assay should be repeated with DNA molecules of different lengths to determine if the rHMfB-(E18K+G51K) higher affinity variant that did not facilitate monomer circles production with the 88 bp would generate monomer circles with DNA fragments that have ends that can contact each other. This would be consistent with lack of DNA circularization being due to tighter DNA binding. It would also be interesting to undertake 'in vitro evolution' Selex procedures using the E18K, G51K and (E18K+G51K) variants to determine if DNA sequences, selected by these variants, would have an even higher affinity for archaeal histones than those
selected by wt rHMfB. Such experiments could also be performed using fragments of
archaeal genomic DNA to select natural DNA sequences with very high histone
affinities, although the Selex experiments undertaken with HMfB and M.
thermoautotrophicum DNA did not isolate any very high affinity sequences (Bailey et al.,
2000). It may in fact be deleterious to have very high affinity sites within an archaeal
genome because this might prevent the DNA from unwinding for processes like DNA
replication and transcription.

An issue that remains to be addressed is the possibility that interaction by residues
in one of the two monomers in an archaeal histone dimer may be sufficient for stable
DNA binding. Recently, histone homodimers (rHMfAA, rHMfBB) and heterodimers
(rHMfAB, rHMfBA) have been constructed in which the two archaeal histone monomers
are covalently fused in tandem by an 8 amino acid linker (F. Marc, personal
communication). Mutations can now be introduced into one monomer of these fused
‘dimers’ to see if the other partner in a dimer can rescue the mutation. It is also still not
clear whether the HMf histones bind to DNA as dimers and that subsequently associate
with a second dimer to form a tetramer on the DNA, or whether pre-assembled tetramers
bind to the DNA. Currently experiments are underway to identify which residues are
necessary for tetramer formation and whether such residues are necessary for archaeal
nucleosome assembly (K. Sandman, personal communication). Studies are also in
progress to determine if complexes can be formed with DNA fragments as short as 30 bp,
which are predicted to be the minimum DNA fragment length required to accommodate
all archaeal histone dimer contacts based on studies with (H2A+H2B) dimers and the
nucleosome crystal structure (Oohara and Wada, 1987; Lüger et al., 1997). Based on previous EMSA studies (Bailey et al., 1999), 60 bp is the minimum length of DNA needed to form a stable complex, but as shown in Chapters 2-4, other assays can detect archaeal histone-DNA binding that are not detectable by EMSAs.

Many of the experiments reported here reveal subtle differences in DNA binding of rHMfA and rHMfB even though all the residues identified in rHMfB as required for DNA binding are identical in both histones. The crystal structures of the rHMfA and rHMfB dimers similarly are superimposable except for a slight deviation in the backbone atoms of the C termini of the α3 helices which result in a slightly different positioning of the α2-α3 interfaces (Decanniere et al., 2000). The α3 domain-swapping experiments between rHMfA and rHMfB should be undertaken to determine if this α2-α3 interaction plays an indirect role in DNA binding. DNA binding studies with the covalently-linked rHMfBA and rHMfAB heterodimers are presently in progress (F. Marc, personal communication) which should overcome the problem of rHMfA and rHMfB spontaneously disassembling and reassembling in solution into homodimers and heterodimers. The results of these experiments might also be of value in determining the potential of the rHMfA/B heterodimers in genome organization and compaction in *M. fervidus*. This activity has previously precluded determining the activity of the rHMfA/B heterodimer.

The precise path that the DNA takes in an archaeal nucleosome is still unknown and therefore archaeal histone-DNA crosslinking studies would be useful to map the path of the DNA in an archaeal nucleosome. During my studies, I initiated experiments to
crosslink the rHMfB-S21C variant to both a natural 80 bp DNA derived from the 7S rRNA encoding gene of *M. fervidus* and to the 110 bp Selex clone 20 DNA. The photoactivatable heterobifunctional crosslinking agent 4-azidophenacylbromide (APB) was used which has a bromoacetyl group that attaches to cysteines, and a phenyl azide moiety that can be crosslinked either to DNA on an adjacent protein if the nitrene group is sufficiently close enough for van der waals contacts with the DNA or protein (Lee et al., 1997). However, I was unable to attach the 4-APB to rHMfB-S21C under a variety of conditions tested. Similar experiments could alternatively be performed by incorporating a nucleolytic chelator-metal complex (1,10-phenanthroline-copper or Fe2+-EDTA) into the protein and identifying crosslinking sites by hydroxyl radical cleavage, as was done with an H4-cysteine variant and the 5S rRNA gene sequence (Pendecast et al., 1992, Flaus et al., 1996). Alternatively, photoactivatable probes could be introduced at specific sites in the DNA that could then be crosslinked to rHMfB and assembled into archaenal nucleosomes (Pruss et al., 1996).

Nucleosomes repress eukaryal transcription initiation and block DNA access by eukaryal transcription activators (reviewed in Steger and Workman, 1996). *In vitro* transcription assays are currently being pursued in the presence of archaenal nucleosomes and the initial results indicated that rHMfB-containing archaenal nucleosomes reduce or completely eliminate *in vitro* transcription by *M. thermoautotrophicum* RNA polymerase in the presence of native TATA-binding protein (TBP) and transcription factor TFIIB. The inhibitory effects of rHMfB on transcription appear to be directly proportional to the amount of rHMfB added and, as expected, addition of the rHMfB-(K13T+R19S+T54K)

246
variant that lacks DNA binding ability had no deleterious effects on transcription (M. Xie, personal communication). Transcription studies can also be conducted in the presence of the higher affinity archaeal histones like rHMfB-E18K, G51K or the (E18K+G51K) variant to confirm that they bind more tightly to DNA, and it is hypothesized that these variants would reduce or completely inhibit transcription at a lower histone:DNA ratio than wt rHMfB as they might prevent unwinding of the DNA to allow these processes to occur.

Other histone fold proteins: potential for histone interactions and DNA binding

Non-histone, histone-fold containing proteins have evolved within Eukarya and have been found to be associated with multi-subunit complexes in transcriptional machinery and nucleosomal remodeling complexes. Although it is not clear what roles the histone-folds play in these proteins, several residues that are involved in DNA binding and in stabilizing the dimer in the archaeal and eukaryal histones are conserved in these proteins. It is hypothesized that conservation of the histone fold dimerization interface and DNA binding residues would aid in nucleosomal remodeling for transcription by allowing other histone-fold proteins to replace histones (Luger et al., 1998a). Therefore, it might be useful to identify these conservations in sequence and structure in histone fold proteins in the Eukarya and relate them to residues in the predecessors of the eukaryal histones, namely the HMf family of archaeal histones.

The Drosophila chromatin accessibility complex (CHRAC) is a multi-protein complex that contains the ATPase ISWI and functions to remodel nucleosomes for gene
regulation. Recently, two of the smallest CHRAC subunits, CHRAC-14 and CHRAC-16 were identified and characterized (Corona et al., 2000). These proteins contain histone folds with short N and C terminal extensions and most closely resemble the mammalian transcription factors NF-YB (CBF-A) and NF-YC (CBF-C) and the yeast HAP3 and HAP5, respectively. Primary sequence comparisons suggest homology between CHRAC-14 and H2B, and CHRAC-16 and H2A. Moreover, CHRAC-14 and CHRAC-16 form heterodimers like (H2B+H2A) complexes, in vivo and in vitro, via their histone folds. CHRAC-14 and CHRAC-16 might aid in nucleosome sliding by disrupting local interactions between histones and DNA since they contain several conserved residues required for DNA binding. Specifically, these residues include CHRAC-16 homologs of R29 and R35 in the α1 helix, V43 in L1 and K74 in L2 of H2A and CHRAC-14 homologs of K43 in the α1 helix, I51 in L1 and T85 in L2 of H2B. Structural modeling of the putative (CHRAC-14 + CHRAC-16) heterodimer with the (H2B+H2A) heterodimer confirms that (CHRAC-14 + CHRAC-16) could bind to DNA non-specifically. Hence, it is thought that CHRAC-14 and CHRAC-16 might function to remodel nucleosomes either directly by binding to nucleosomal DNA and displacing the (H2B+H2A) complex or indirectly by dimerizing with these histone molecules that have been removed from the DNA (Corona et al., 2000).

Similarly, NF-Y or CBF (a CCAAT-binding complex in mammalian cells), designated HAP in yeast, could interact with histones. The CCAAT box is a ubiquitous element found at ~25% of eukaryal promoters. NF-Y contains 3 subunits namely NF-YB (CBFA or HAP3), NF-YC (CBFC or HAP5) and NF-YA (CBFB or HAP2). NF-YB and
NF-YC contain histone folds and resemble the eukaryal (H2A+H2B) heterodimer, while NF-YA does not bear resemblance to any known DNA-binding motif (Mantovani et al., 1999). All 3 NF-Y subunits are needed for DNA binding but specificity for CCAAT boxes is achieved via the non-histone-like protein, NF-YA. Interestingly, only histone fold domains of NF-YB and NF-YC complexes are needed to bind to (H3+H4) tetramers but these proteins did not bind to (H2A+H2B) dimers. Contacts between the (H3+H4) tetramer and (H2A+H2B) dimer are made through a 4-helix bundle between H2B and H4 and these are conserved in NF-YB, suggesting strong evolutionary relatedness of these two complexes (Caretti et al., 1999).

Sequence alignments between histones and NF-YB and NF-YC subunits suggest conservation in DNA binding residues. Specifically, the proline cap at the N terminus of the α1 helix as well as arginine and lysine residues in the α1 helix are conserved in both these monomers. NF-YB has a lysine in L1 at the equivalent position of the rHMfB R19 residue, thought to contact the DNA minor groove phosphodiester backbone via hydrogen bonding. A conserved isoleucine, involved in β-sheet interactions in L1, is also present in NF-YB and NF-YC at positions corresponding to V20 in rHMfB. In the L2 region, a conserved lysine is found in both NF-YB and NF-YC, which is homologous to lysines at this position that are salt-linked to the DNA in all 4 eukaryal histones and which extend the DNA contacts to the next helical turn. Also, a homologous lysine-threonine interaction in L1-L2 (arginine-threonine in all archaeal dimers and most eukaryal dimers) is maintained in a (NF-YB + NF-YC) heterodimer on one end of the molecule. Further, an intramolecular buried salt bridge, needed for the maintenance of
loop stability, corresponding to the R52-D59 interaction in rHMfB nucleosomes, is conserved in NF-YB and NF-YC. Thus, the NF-Y histone fold complex like CHRAC is well suited to interact with both nucleosomal DNA and histones. It could remove histones from nucleosomal DNA and free CCAAT promoters for transcriptional activation. NF-Y could also exploit the histone fold to dimerize with TAFs and Dr1-DRAPl that might activate transcription at TATA boxes near promoters (Caretti et al., 1999).

Activator-dependent eukaryal transcription requires the TATA box-binding protein TBP and several other TFIID components including TBP-associated factors (TAFs) and some non-TAF co-activator proteins. In Drosophila, the crystal structure of the dTAF42 and dTAF62 reveal that (dTAF42 + dTAF62) form a heterotetramer, reminiscent of the (H3+H4) tetramer formed at the center of the eukaryal nucleosome. The root-mean-squared (rms) deviations between the α-C atoms of the dTAF42 versus H3 and dTAF62 versus H4 are both 1.6Å, which is similar to the values obtained in comparing the eukaryal core histones dimers with each other. Like eukaryal histones, dTAF42 and dTAF62 do not have the potential to homodimerize. It is proposed that the 2 dTAF42 dimers form a heterotetramer via a 4-helix bundle. TFIID is also proposed to contain a histone-like octameric structure comprised of two dimers of the H2B-like dTAF28 bound to the (dTAF42 + dTAF62) tetramer. There does not appear to be an H2A homolog in the dTAFs (Xie et al., 1996).

dTAF42 and dTAF62 homologs have also been identified in humans that also share structural similarity with histones. Previous studies indicated that an octameric histone-like TAF complex was present within TFIID, comprised of the central (hTAF31
+ hTAF80) heterotetramer that resembles (H3+H4) eukaryal nucleosome tetramer and
two homodimers of the H2B-like hTAF20 (Xie et al., 1996). Recently hTAF135 was
shown to heterodimerize with hTAF20, homologous to the (H2A+H2B) interaction in the
eukaryal nucleosome. Further complications to an octameric TFIID model arise from two
other proteins, hTAF28 and hTAF18, which also heterodimerize via their histone folds in
TFIID (Gangloff et al., 2000). Additional studies are needed to elucidate the structural
interactions between these 3 types of histone fold dimers.

Sequence alignments of the dTAFs with histones reveals some conservation in the
dNA binding residues. Positively charged amino acids occur in the α1 helix of dTAF62
at positions equivalent to R10 in rHMfB and at homologous sites to K13 in rHMfB in
dTAF42 and dTAF28. However, the L1-L2 arginine-threonine interaction, conserved in
most histones, does not appear to be conserved in these TAFs. A lysine residue is found
in dTAF42 at the homologous position to K53 in rHMfB, and hence it is believed to be
capable of making a salt link to DNA, while all 3 TAFs have positively-charged amino
acids at these positions in the 4 eukaryal core histones. The intramolecular salt bridge
R52-D59, conserved in the archaeal and eukaryal histones, is also found in dTAF42 and
dTAF62 while the D22-R52 intramolecular salt bridge, found in rHMfB and conserved in
the archaeal histones, is also conserved in dTAF62. Although there are several
similarities in DNA binding residues between TAFs and the histones, the structural
assembly of the TAF complex, the abundance of acidic residues in the α1 helices and the
lack of arginines interacting with DNA in the minor groove suggest a different mode of
DNA interaction. It is possible that other proteins play an accessory role in allowing
DNA binding or the TAFs are merely capable of dimerizing with histones (Luger et al., 1998). It would be interesting to see if an octameric TAF complex displays DNA binding, when assembled in vitro.

The histone fold, which is conserved in eukaryal and archaeal histones and in other proteins involved in nucleosome remodeling and transcription control, most likely, serves as a dimerization interface or a DNA-binding surface for interaction of these proteins (Luger et al., 1998). In this manner, larger or alternative complexes might be assembled with different histone-fold monomers and DNA. Other histone fold proteins might replace histones in chromatin to allow transitions in DNA organization for processes like transcription to occur.
LIST OF REFERENCES


253


CHRAC-16, are developmentally regulated subunits of chromatin accessibility complex (CHRAC). EMBO J. 19, 3049-3059.


Craig, N. and Nash, H. 1984. E. coli integration host factor binds to specific sites in DNA. Cell. 39: 707-716.


Fuller, R.S., Funnell, B.E. and Kornberg, A. 1984. The dnaA protein complex with the E. coli chromosomal replication origin (oriC) and other DNA sites. Cell. 38: 889-900.


Kawarabayasi, Y. et al. 1998. Complete sequence and gene organization of the genome of a hyperthermophilic archaeabacterium, Pyrococcus horikoshii OT3. DNA Research. 5: 55-76.


260


