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. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

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.
BIOCHEMICAL AND STRUCTURAL CHARACTERIZATION OF THE ARCHAEAL HISTONE HMF FROM THE HYPERTHERMOPHILIC METHANOGEN METHANOOTHERMUS FERVIDUS

DISSERTATION

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

By

Rowan Andrew Grayling, B.S., M.S.

*****

The Ohio State University
1995

Dissertation Committee:

C. J. Daniels
J. A. Krzycki
J. N. Reeve
F. R. Tabita

Approved by

Adviser
Department of Microbiology
"The role of the infinitely small in nature is infinitely large"

Louis Pasteur
This work is dedicated to Life on Earth,
which many admire, but few truly appreciate.
ACKNOWLEDGEMENTS

I thank especially my adviser, Dr. John Reeve, for his guidance, support and friendship over the past five years. I am also indebted to Dr. Kathleen Sandman for her expert advice, friendship and support, several successful collaborations, and a seemingly endless supply of recombinant \textit{Escherichia coli} clones and recombinant proteins. Encouragement and advice from Dr. Wayne Becktel and my committee members, Dr. Joseph Krzycki, Dr. Charles Daniels, and Dr. F. Robert Tabita has been greatly appreciated. I thank Don Ordaz in the OSU Fermentation Facility, Dr. John Lowbridge in the OSU Biochemical Instrumentation Center, and Bill Swoager in the Department of Microbiology for their advice and technical assistance. Many thanks also to the members of the Reeve lab - past and present, international and domestic - for their friendship and support. Finally, my sincere thanks to my close friend Holly Hillyer, for her support, encouragement, and friendship over the last few years, and also to my parents, for their lifelong support and encouragement, without which none of this would have been possible.
VITA

November 25, 1966..............................Born – Auckland, New Zealand

1988..................................................B.S. in Cellular and Molecular Biology,
University of Auckland, Auckland, New Zealand

1990..................................................M.S. in Cellular and Molecular Biology,
University of Auckland, Auckland, New Zealand

PUBLICATIONS


FIELDS OF STUDY

Major field: Microbiology
   Studies in the Biochemical and Biophysical Characterization of Thermostable Proteins
   Dr. J. N. Reeve, Adviser
TABLE OF CONTENTS

DEDICATION ..................................................................................................................... ii
ACKNOWLEDGEMENTS .................................................................................................. iii
VITA .................................................................................................................................. iv
TABLE OF CONTENTS ...................................................................................................... vi
LIST OF FIGURES ........................................................................................................... xi
LIST OF TABLES ............................................................................................................. xv
LIST OF ABBREVIATIONS ............................................................................................. xvi
CHAPTER I: GENERAL INTRODUCTION ........................................................................ 1
  DNA binding proteins and genome compaction ....................................................... 1
  Histones and nucleosome structure in eukaryotes ................................................. 2
  Prokaryotic DNA binding proteins ......................................................................... 8
  Bacterial DNA binding proteins ............................................................................ 8
  Archaeal DNA binding proteins .......................................................................... 10
  Curved DNA ............................................................................................................ 26
  Circular dichroism spectroscopy ......................................................................... 30
  Thermophiles, and protein and DNA thermal stability .................................... 32
  Goals of this study ................................................................................................. 36
CHAPTER II: PURIFICATION AND BIOCHEMICAL CHARACTERIZATION OF NATIVE AND RECOMBINANT HMF PROTEINS ................................................................. 37
  Introduction ............................................................................................................. 37
  Materials and methods ......................................................................................... 39
Electrophoresis techniques ................................................................. 82
CD spectroscopy ................................................................................... 83
Data analysis .............................................................................................. 85

Results ........................................................................................................ 86

Secondary structures of rHMfA and rHMfB and analysis of a rHMfA*HfB heterodimer containing mixture ........................................ 86
Salt titration of rHMfA and rHMfB ........................................................... 93
Cooperativity of ion uptake by rHMfA and rHMfB at pH 2 ................. 96
Thermal stability of rHMfA and rHMfB in KCl ...................................... 99
Anion dependence of rHMfA and rHMfB salt-based thermostabilization ........................................................... 102
van't Hoff analysis of rHMfA and rHMfB thermal denaturations .......... 107
Reversibility of rHMfA and rHMfB thermal denaturation and rates of refolding .............................................................. 110
Thermal stability of rHFoB ................................................................. 113

Discussion .............................................................................................. 118

CHAPTER IV: THE SEQUENCE SPECIFICITY OF HMF-DNA INTERACTIONS ......................................................... 135

Introduction .............................................................................................. 135

Materials and Methods ........................................................................... 136
Chemicals and reagents ........................................................................... 136
Bacterial strains and plasmids ................................................................ 137
Molecular biological methods ................................................................. 137
Primers used .............................................................................................. 138
Preparation of plasmid DNAs ................................................................. 139
Micrococcal nuclease digestion ................................................................. 139
Electrophoresis techniques .................................................................. 140
CHAPTER V: CHARACTERIZATION OF HMF-DNA COMPLEXES

Introduction .................................................. 193

Materials and Methods ........................... 197

Chemicals and reagents ....................... 197

Plasmid DNA preparation and standard molecular biological methods ....................... 197

Electrophoresis techniques ................... 198

Electrophoretic mobility shift assay (EMSA) studies ........................................... 199

Protein-DNA crosslinking ...................... 199
In vivo crosslinking and MNase digestion .................................................................200

14C-labelling of rHMfA and rHMfB ........................................................................202

Size-exclusion HPLC (SE-HPLC) ...........................................................................203

Image analysis and curve fitting ...........................................................................203

Results ....................................................................................................................204

Crosslinking of HMf-DNA complexes ..................................................................204

MNase digestion of *Mt. fervidus* nucleoprotein .................................................211

Direct investigation of HMf-DNA complexes by size-exclusion HPLC ................217

Properties of the rHMfA and rHMfB DNA binding reactions ..............................225

The most likely structure(s) of HMf-DNA complexes ........................................235

Discussion ...............................................................................................................238

CHAPTER VI: GENERAL DISCUSSION ..................................................................242

Summary .................................................................................................................242

Concerns and future experiments .........................................................................246

The HMf family of proteins as model systems to study the histone-DNA interaction and protein thermodynamics ..................................................248

Evolutionary considerations ..................................................................................249

LIST OF REFERENCES ...........................................................................................251
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Structure of the nucleosome core particle</td>
<td>3</td>
</tr>
<tr>
<td>2.</td>
<td>Dendrogram showing the grouping of amino acid sequences of prokaryotic and eukaryotic DNA binding proteins</td>
<td>11</td>
</tr>
<tr>
<td>3.</td>
<td>Alignment of the primary sequences of members of the HMf family of archaeal histones with consensus sequences for the structured central regions of the eukaryal histones</td>
<td>17</td>
</tr>
<tr>
<td>4.</td>
<td>Model for nucleosome-like structure (NLS) formation by HMf</td>
<td>23</td>
</tr>
<tr>
<td>5.</td>
<td>Models indicating the rotational degrees of freedom of nucleotide base pairs, and the formation of local DNA curvature</td>
<td>27</td>
</tr>
<tr>
<td>6.</td>
<td>Optimized fermentation and protein purification conditions</td>
<td>48</td>
</tr>
<tr>
<td>7.</td>
<td>Growth phase dependent synthesis of HMfA and HMfB</td>
<td>54</td>
</tr>
<tr>
<td>8.</td>
<td>Design of control experiments used to demonstrate the non-selectivity of the protein purification method used in the growth curve analysis</td>
<td>56</td>
</tr>
<tr>
<td>9.</td>
<td>Validation of heat-treatment/DNA cellulose purification method as a means to establish HMfA/HMfB ratios in native HMf preparations</td>
<td>58</td>
</tr>
<tr>
<td>10.</td>
<td>Specificity and concentration dependence of HMf crosslinking</td>
<td>61</td>
</tr>
<tr>
<td>11.</td>
<td>Tricine-SDS PAGE of the products of HCHO crosslinking</td>
<td>64</td>
</tr>
<tr>
<td>Page</td>
<td>Section</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td>Comparisons of non-denaturing gel systems and resolution of native HMf preparations</td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>Two dimensional PAGE of native HMf and a heat-treated equimolar mixture of rHMfA and rHMfB dimers</td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td>Analysis of rHMfA expression in <em>E. coli</em></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>C-source code for the 9-point quadratic-cubic smoothing routine described in the text</td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td>CD spectra of mixtures of (rHMfA)₂ and (rHMfB)₂ dimers and non-denaturing PAGE analysis of rHMfA•rHMfB heterodimer formation</td>
<td></td>
</tr>
<tr>
<td>17.</td>
<td>Dependence of CD spectra on rHMfA and rHMfB concentration and salt titrations of rHMfA and rHMfB at pH 2, and pH 7.5</td>
<td></td>
</tr>
<tr>
<td>18.</td>
<td>Ion uptake calculations for the folding reactions of rHMfA and rHMfB at pH 2</td>
<td></td>
</tr>
<tr>
<td>19.</td>
<td>Dependence of the thermal stability of HMf proteins on salt concentration</td>
<td></td>
</tr>
<tr>
<td>20.</td>
<td>Anion-dependence of rHMfA and rHMfB thermostability</td>
<td></td>
</tr>
<tr>
<td>21.</td>
<td>Thermal denaturation of HMf proteins in buffer containing 2.75 M gdm-cl</td>
<td></td>
</tr>
<tr>
<td>22.</td>
<td>Dependence of buffer on temperature and Van't Hoff analysis of rHMfA and rHMfB thermal unfolding</td>
<td></td>
</tr>
<tr>
<td>23.</td>
<td>Reversibility of rHMfA and rHMfB thermal denaturation</td>
<td></td>
</tr>
<tr>
<td>24.</td>
<td>Kinetics of rHMfA and rHMfB refolding</td>
<td></td>
</tr>
<tr>
<td>25.</td>
<td>Thermal denaturation of rHFoB, and comparison to rHMfA and rHMfB</td>
<td></td>
</tr>
<tr>
<td>26.</td>
<td>Reversibility of rHFoB thermal denaturation</td>
<td></td>
</tr>
<tr>
<td>27.</td>
<td>Kinetics of rHFoB refolding</td>
<td></td>
</tr>
<tr>
<td>28.</td>
<td>Predicted differences in the stabilities of HMfA and HMfB</td>
<td></td>
</tr>
<tr>
<td>29.</td>
<td>Alignment of the amino acid sequences of HFoB and HMfB</td>
<td></td>
</tr>
</tbody>
</table>
30. Calculation of DNA bending parameters by AUGUR .................. 145
31. MNase digestion of HMf-DNA complexes assembled in vitro ......................................................... 148
32. Approaches used to assess the sequence specificity of rHMfA-DNA binding ........................................... 151
33. Southern blot analysis of pUC19 restriction digests, using 30 and 60 bp probes .................................................. 153
34. Southern blot analysis of pRG101 restriction digests ............... 155
35. Comparison of DNA protection from MNase digestion by rHMfA and rHMfB, and analysis of the population of ~60 bp fragments used for cloning .................................................... 158
36. Size distribution and predicted intrinsic curvatures of the cloned ~60 bp fragments .................................................. 166
37. Summary of the statistical cloning results .............................................. 168
38. Alignment of unique, non-overlapping cloned ~60 bp DNA sequences ................................................................. 171
39. Construction of pBEND2-60 derivatives .............................................. 175
40. Sequences of the XbaI inserts in the five pBEND2-60 plasmids ................................................................. 177
41. Stereo images of bends predicted for the pBEND2-60 inserts ........................................................................ 179
42. The pBEND2 polylinker region identifying the restriction sites used for the analysis of the pBEND2-60 inserts .......... 182
43. Gel mobility analysis of the five pBEND2-60 inserts ..................... 184
44. Comparative electrophoretic analysis of the pBEND2-60 inserts ........................................................................ 187
45. Organization of the DNA around the core histone octamer and the (H3•H4)2 tetramer ........................................... 194
46. Crosslinking of HMf and DNA under low-salt conditions ........................................................................ 205
47. Salt-dependence of rHMfA- and rHMfB-DNA crosslinking ........................................................................ 208

xiii
48. MNase digestion of fixed and unfixed *Mt. fervidus* nucleoprotein ........................................ 212

49. Contrast-enhanced, negative image of a timecourse of MNase digestion of unfixed *Mt. fervidus* nucleoprotein .......... 215

50. MNase digestion of unfixed and fixed deproteinized *Mt. fervidus* genomic DNA .................................................. 218

51. Size-exclusion HPLC resolution of crosslinked rHMfA-DNA complexes .......................................................... 221

52. Identification of the material in the crosslinked rHMfA-DNA complexes resolved by size-exclusion HPLC .......... 223

53. Agarose gel EMSA of rHMfA and rHMfB using *SmaI*-digested pRG101 DNA .................................................. 227

54. DNA binding curves for preparations of rHMfA and rHMfB ........................................................................... 229

55. EMSA of DNA binding by rHMfA and rHMfB using 8% T, 1.3% C polyacrylamide gels polymerized and run in 1X TBE buffer ........................................................................... 233
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Similarities of HMf proteins with the eukaryal core histones</td>
<td>20</td>
</tr>
<tr>
<td>2.</td>
<td>Physical properties of rHMfA and rHMfB</td>
<td>92</td>
</tr>
<tr>
<td>3.</td>
<td>Melting temperatures of rHMfA, rHMfB, and rHFoB in different buffers</td>
<td>123</td>
</tr>
<tr>
<td>4.</td>
<td>Sequences and analysis of cloned pRG101 DNA fragments, protected from MNase by rHMfA binding</td>
<td>162</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>% C</td>
<td>percent $N,N'$-methylene bisacrylamide (crosslinker %), relative to the total monomer concentration, % T</td>
<td></td>
</tr>
<tr>
<td>% T</td>
<td>percent acrylamide plus $N,N'$-methylene bisacrylamide monomer (total monomer %)</td>
<td></td>
</tr>
<tr>
<td>AUT</td>
<td>acetic acid-urea-Triton X-100</td>
<td></td>
</tr>
<tr>
<td>bp</td>
<td>base pair(s)</td>
<td></td>
</tr>
<tr>
<td>CBB</td>
<td>Coomassie Brilliant Blue R-250</td>
<td></td>
</tr>
<tr>
<td>CD</td>
<td>circular dichroism</td>
<td></td>
</tr>
<tr>
<td>cDPG</td>
<td>cyclic-2,3-diphosphoglycerate</td>
<td></td>
</tr>
<tr>
<td>cpm</td>
<td>counts per minute</td>
<td></td>
</tr>
<tr>
<td>CPMA</td>
<td>circular permutation mobility assay</td>
<td></td>
</tr>
<tr>
<td>DMSI</td>
<td>dimethyl suberimidate</td>
<td></td>
</tr>
<tr>
<td>ddH$_2$O</td>
<td>double-distilled water</td>
<td></td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxynucleotide triphosphate</td>
<td></td>
</tr>
<tr>
<td>dpm</td>
<td>disintegrations per minute</td>
<td></td>
</tr>
<tr>
<td>EDTA</td>
<td>disodium ethylenediaminetetraacetic acid</td>
<td></td>
</tr>
<tr>
<td>EM</td>
<td>electron microscopy</td>
<td></td>
</tr>
<tr>
<td>EMSA</td>
<td>electrophoretic mobility shift assay</td>
<td></td>
</tr>
<tr>
<td>EtBr</td>
<td>ethidium bromide</td>
<td></td>
</tr>
<tr>
<td>gdm-cl</td>
<td>guanidine hydrochloride</td>
<td></td>
</tr>
<tr>
<td>HCHO</td>
<td>formaldehyde</td>
<td></td>
</tr>
<tr>
<td>HPLC</td>
<td>high performance liquid chromatography</td>
<td></td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl $\beta$-D-thiogalactopyranoside</td>
<td></td>
</tr>
<tr>
<td>KPi</td>
<td>potassium phosphate buffer</td>
<td></td>
</tr>
<tr>
<td>MES</td>
<td>2-(N-morpholino)ethanesulfonic acid</td>
<td></td>
</tr>
<tr>
<td>MNase</td>
<td>micrococcal nuclease</td>
<td></td>
</tr>
<tr>
<td>MOPS</td>
<td>3-(N-morpholino)propanesulfonic acid</td>
<td></td>
</tr>
<tr>
<td>MW</td>
<td>molecular weight</td>
<td></td>
</tr>
</tbody>
</table>
NLS  nucleosome like structure(s)
OD   optical density
PAGE polyacrylamide gel electrophoresis
PCR  polymerase chain reaction
PIVES piperazine-\(N,N'-\)bis(2-ethanesulfonic acid)
PMSF  phenylmethylsulfonylfluoride
rpm revolutions per minute
SDM  site-directed mutagenesis
SDS  sodium dodecyl sulfate
SEC  size exclusion chromatography
\(t_{1/2}\)  half life or half time of protein refolding
\(T_m\)  thermal transition (melting temperature) for protein or DNA
denaturation
Tdr thymidine
TEA triethanolamine
TFA trifluoroacetic acid
TMP trimethoprim
tricine \(N\)-[tris(hydroxymethyl)methyl]glycine
Tris tris(hydroxymethyl)aminomethane
U  unit(s) of enzyme activity
UV  ultraviolet light
CHAPTER 1
GENERAL INTRODUCTION

DNA binding proteins and genome compaction

All cells face a genome packaging problem. For example, an exponentially growing *Escherichia coli* cell with a volume of ~1 μm$^3$ must accommodate two to four chromosomes, each of which is a circular 4.7 Mbp DNA molecule ~1000 times longer than the cell itself (Kohara *et al.*, 1987; Krawiec and Riley, 1990). DNA packaging must overcome electrostatic repulsion resulting from the negative charges on the DNA phosphate backbone, and must also be sufficiently dynamic to allow (rapid) gene expression from within a compacted structure. Eukaryal cells, by definition, have a nucleus that is separated from the cytoplasm by a nuclear membrane, in which the genomic DNA is packaged by histones into well-defined nucleosomes that are further compacted into chromatin fibers and chromosomes (Kornberg and Lorch, 1992; Wolffe, 1992). In contrast, the basis of DNA packaging and structures in compacted prokaryotic genomes, in both the *Archaea* and *Bacteria*, are not understood.

Topoisomerase-generated local supercoiling could account for a significant amount of genome compaction in both eukaryotic and prokaryotic cells, with polyamines, Mg$^{2+}$, and other cations providing charge neutralization, and supercoiling-based packaging may be a major component of DNA compaction in prokaryotes with very small genomes (Worcel and Burgi, 1972; Griffith, 1976; Woolley, 1986; Pettijohn, 1988). Prokaryotes, like eukaryotes, do however also contain abundant DNA binding proteins that appear to be involved in genome compaction (Drlica and Rouvière-Yaniv, 1987;
Schmid, 1990). These prokaryotic DNA binding proteins share with the eukaryal histones the properties of small size, net positive charge, abundance, and binding site preferences determined by DNA shape rather than sequence, and consequently have been termed histone-like. However, apart from the HMf family of archaeal proteins, which can legitimately be described as archaeal histones (this work), all of the prokaryotic DNA binding proteins studied to date are structurally unrelated to the eukaryal histones (Grayling et al., 1994).

**Histones and nucleosome structure in eukaryotes**

The eukaryal histones provide the paradigm for ‘structural’ DNA binding proteins. These abundant, highly conserved, basic proteins are found in all eukaryal cells, and bind to DNA in a relatively non-sequence-specific manner. Histones H2A, H2B, H3, and H4 wrap nuclear chromosomal DNA into nucleosome ‘core’ particles which, with the addition of ‘linker’ histones (H1, H5, or their variants) form the structures termed nucleosomes. Arrays of nucleosomes are then further compacted, in the presence of additional non-histone chromosomal proteins, to form the highly condensed structure known as chromatin (Wolffe, 1992).

Nucleosome core particles contain an octamer of histones surrounded by ~146 bp of DNA wrapped in 1.75 negative toroidal supercoils. This core histone octamer is a tripartite structure, composed of an (H3•H4)2 tetramer flanked on either side by a (H2A•H2B) dimer (Figure 1) (Arents et al., 1991; Kornberg and Lorch, 1992; Wolffe, 1992). Nucleosome core particles can be isolated intact by micrococcal nuclease (MNase) digestion of histone H1-depleted chromatin, and can be reconstituted in vitro using purified histones and DNA (Noll and Kornberg, 1977; Thomas and Butler, 1978; Wolffe, 1992). Core particles appear characteristically as regular ‘beads-on-a-string’
Figure 1.

Structure of the nucleosome core particle. A. View down the superhelical axis showing the organization of the histone pairs in the absence of histone H1. Cylinders represent α-helices and the N- and C-termini of the monomers are identified. B. Side view showing the wedge-shaped structure of the protein core. A portion of the DNA is shown in the background which is predicted to interact with the ‘β-bridge’ and ‘paired ends of α-helices’ motifs, as indicated. The second (H2A•H2B) dimer, which would lie behind the plane of the page in panel A and below the structure drawn in panel B, is not shown for clarity. Both (H2A•H2B) dimers and the (H3•H4)2 tetramer ‘core’ are required for complete wrapping of the DNA and protection of the full ~146 bp from MNase digestion. These drawings were adapted from Pruss et al. (1995), Wolffe (1994a), and Wolffe (1995), from original graphics obtained from D. Pruss.
Figure 1.
structures when visualized by electron microscopy (EM; Olins and Olins, 1974; van Holde, 1988; Wolffe, 1992). The three-dimensional structure of the core histone octamer of the nucleosome has been determined to a resolution of 3.1 Å (Arents et al., 1991). It has the overall shape of a short cylinder in which each histone monomer interacts with its partner to form a head-to-tail dimer, and these dimers are arranged within the octamer to form a left-handed protein supercoil (Figure 1; Moudrianakis and Arents, 1993). The four histone monomers, H2A, H2B, H3, and H4, that form the octamer each contain a common tertiary structural motif, termed the histone fold, which consists of a long central α-helix separated on either end from a short α-helix by a loop and β-strand segment (Figure 1; Arents et al., 1991). Histone dimerization results in the pairing of the monomer loop segments within each dimer to form two parallel β-bridge structures per dimer (Figure 1). The amino termini of the N-terminal short α-helices are also paired within each dimer, and together with the β-bridge structures, form positively charged moieties that are positioned on or near the octamer surface. These result in a positively charged track which is predicted to direct the path of the negatively charged DNA around the surface of the nucleosome core (Arents and Moudrianakis, 1993). In addition to the histone fold motifs, the core histones also have unstructured N- and C-terminal ‘tails’ that extend beyond the nucleosome (Wolffe, 1992). These tails provide sites for the post-translational histone modifications (acetylation, phosphorylation, methylation, ADP-ribosylation; Wolffe, 1992; 1994a, b) that may regulate gene expression (Lewin, 1994; Wolffe, 1994a), however, removal of these tails has no detectable influence on the extent or organization of the DNA wrapped around the core histone octamer (Pruss et al., 1995). These tails are not present in the HMf family of proteins (see below).

The wrapping of DNA around the core histone octamer in a left-handed superhelical configuration results in an overwinding of the DNA helix, reducing the
average helical periodicity of the wound DNA from ~10.6 to ~10.2 bp/turn (Wolffe, 1992). The net effect of this supercoiling and overwinding is the storage of ~1 negative supercoil per nucleosome (Wolffe, 1992). Processes such as transcription and replication that require DNA strand separation occur much more readily in negatively supercoiled than relaxed DNA and this storage of negative supercoils by nucleosomes is therefore thought to provide an energetically favorable environment for strand separation in processes such as transcription initiation that otherwise might be thermodynamically difficult (Travers, 1990; Drlica, 1990).

The (H3•H4)2 tetramer has a central role in nucleosome formation and is a much more stable structure that the core histone octamer, both when bound to DNA and when free in solution. In the assembly of nucleosomes, (H2A•H2B) dimers bind to DNA-bound (H3•H4)2 tetramers (Camerini-Otero et al., 1976; Jorcano and Ruiz-Carrillo, 1979; Aragay et al., 1988) that are responsible for both the path of the DNA around the histone octamer, and for the position of the nucleosome on a specific DNA sequence (Dong and van Holde, 1991; Hayes et al., 1991). MNase digestion of H1-depleted nucleosomes (nucleosome core particles) results in protected ~146 bp DNA fragments, whereas (H3•H4)2 tetramers formed in vitro protect ~70 bp from MNase digestion, although in vitro the tetramer binds and folds the same ~146 bp sequence as a fully formed octamer. The extremities of these sequences are, however, more accessible to MNase digestion which therefore results in the smaller ~70 bp fragments generated in these protection assays (Dong and van Holde, 1991). The central role of the (H3•H4)2 tetramer is consistent with phylogenetic studies of histone amino acid sequences, which indicate that histones H3 and H4 are much more evolutionarily conserved than histones H2A and H2B (Thatcher and Gorovsky, 1994).
Nucleosomes are not static units involved solely in DNA compaction, but are dynamic structures that associate, dissociate, and migrate, and that can both activate and repress transcription (Lewin, 1994; Wolffe, 1994a, b). Nucleosomes are positioned on a DNA molecule both by its sequence-dependent, inherent shape and by interactions with other pre-bound, sequence-specific DNA binding proteins, such as transcription factors (Thoma, 1992; Wolffe, 1994a). The histone (H2A•H2B) dimers of positioned nucleosomes have been implicated in regulating both transcription repression and activation, and in determining the extent of chromatin folding (Hansen and Wolffe, 1994; Wolffe, 1994a, b). The diversity of specialized, developmentally regulated H2A and H2B variants found in some organisms during embryogenesis is consistent with important roles in transcriptional regulation and chromatin condensation (Wolffe, 1992; Wolffe, 1995).

Nucleosomes and their component histones may interact with the eukaryal transcriptional apparatus in several ways. Competition between the histone octamers and the RNA polymerase complex or transcriptional activator proteins for DNA binding sites in promoter regions may inhibit transcription. Alternatively, histone binding may activate transcription by juxtaposing otherwise distantly separated activator binding sites. Activator interactions with their cognate binding sites might, in principle, also be facilitated by the display of these sites on the nucleosome surface (Wolffe, 1994a). Once initiated, transcription appears to proceed through nucleosomes, although arguments have been made for the selective and transient displacement of the (H2A•H2B) dimers, or of the entire histone octamer, as a consequence of RNA polymerase-induced changes in local DNA topology (Brooks and Jackson, 1994; Lewin, 1994; Studitsky et al., 1994).
Prokaryotic DNA binding proteins

Although many histone-like proteins have been characterized from prokaryotes, homologs of the eukaryal histones have been documented in only a few archaeal species (Grayling et al., 1994). The DNA packaging problem may not be as severe in prokaryotes as in eukaryotes, which have genome sizes 5 to 10,000 times larger than prokaryotes (Lewin, 1990). The compaction of prokaryotic genomes may therefore be met largely through topoisomerase-generated supercoiling, and by positively charged ions and small polycations such as polyamines. Prokaryotes must, however, respond rapidly to environmental changes, and the involvement of histone-like DNA binding proteins in rapid local and global regulation of gene expression, in addition to their roles in DNA compaction, has been studied extensively (Drlica and Rouvière-Yaniv, 1987; Schmid, 1990; Oberto et al., 1994; Ussery et al., 1994).

Bacterial DNA binding proteins

Six different, relatively abundant, histone-like DNA binding proteins (HU, H-NS, HLPI, H, IHF, and FIS) have been isolated and characterized from E. coli, of which HU is the most evolutionarily conserved and best studied (Drlica and Rouvière-Yaniv, 1987; Schmid, 1990). E. coli HU is a member of a widely distributed family of DNA binding proteins which have been characterized from evolutionarily diverse Bacteria, from eukaryal organelles, and from the archaeon Thermoplasma acidophilum. This family includes the sequence-specific DNA binding proteins integration host factor (IHF) from E. coli and transcription factor TF1 from the Bacillus subtilis bacteriophage SP01 (Drlica and Rouvière-Yaniv, 1987; Pettijohn, 1988; Schmid, 1990; Oberto et al., 1994).
E. coli HU was isolated originally as a factor that enhanced the transcription of bacteriophage lambda genes in vitro. It is a small (18 kDa), abundant (30,000 copies per cell), intrinsically heat-resistant nucleoid-associated protein that in E. coli and Salmonella typhimurium is a heterodimer of two closely related monomers, HU-α and HU-β (Drlica and Rouvière-Yaniv, 1987; Oberto et al., 1994; and references therein). In other bacteria, HU molecules are homodimers of identical HU monomers (Drlica and Rouvière-Yaniv, 1987). Although HU-deficient E. coli strains are viable, they grow more slowly and with perturbed cell divisions that generate anucleate and filamentous cells. They are also unable to support the replication of some phages and plasmids, exhibit unusual cold sensitivity and show an increased sensitivity to γ-radiation (Oberto et al., 1994; Wada et al., 1988; Huisman et al., 1989; Ogura et al., 1990; Boubrik and Rouvière-Yaniv, 1995). HU binds to DNA in a sequence-independent manner, and bends, compacts and constrains DNA molecules into negative toroidal supercoils in vitro, generating nucleosome-like structures (NLS) observable by EM that contain 8 to 10 HU dimers per 275 bp of DNA (Rouvière-Yaniv et al., 1979; Broyles and Pettijohn, 1986). If these NLS are present in vivo, they must be extremely labile towards dissociation, as there is very little evidence for such a protein-DNA complex in lysed E. coli cells examined by EM (Griffith, 1976; Pettijohn, 1988; Grayling et al., 1994). Recently HU has been recognized as being functionally related to the eukaryal high mobility group (HMG) proteins that bind to four-way junction DNAs (cruciform or Holliday structures) and HMG1, synthesized in E. coli, has been shown to complement some HU functions in vivo (Bianchi, 1994). HU binding appears to direct DNA bending that can juxtapose non-adjacent sites on the E. coli chromosome and, by doing so, facilitate events such as site-specific recombination and transcription activation.
The three-dimensional structure of the HU dimer from *Bacillus stearothermophilus*, also termed DNA Binding Protein II, has been determined to a resolution of 3 Å by X-ray crystallography (Tanaka *et al.*, 1984). Each monomer contains three α-helices and an antiparallel 2-stranded β-sheet that projects as an ‘arm’ from the monomer. The two arms of a dimer are predicted to bind to the DNA molecule through contacts with the sugar-phosphate backbone, forming a DNA-protein complex that is structurally very different from the eukaryal nucleosome (Tanaka *et al.*, 1984; Arents *et al.*, 1991).

**Archaeal DNA binding proteins**

Several small, abundant, positively charged DNA binding proteins that lack sequence specificity have been isolated and characterized from a wide variety of archaeal species (Grayling *et al.*, 1994). Based on alignments of their amino acid sequences, these archaeal histone-like proteins form four groups, namely HTa from *Thermoplasma acidophilum* (which appears to be related to the bacterial HU proteins), the MCI family from the *Methanosarcinaceae*, a group of closely related proteins from *Sulfolobus* species, and the HMf family of proteins from the *Methanobacterales* and *Thermococcales* (Figure 2). An open reading frame sequenced from *Methanococcus voltae* (Agha-Amiri and Klein, 1993) suggests that the HMf family may also have distant relatives in the *Methanococcaceae* (Figure 2).

**HTa**

HTa from the archaeal thermophile *Thermoplasma acidophilum* has a molecular mass of 9.9 kDa and binds strongly to DNA under physiological conditions (50 mM K⁺; Searcy, 1975; Searcy, 1976). EM studies of nucleoprotein complexes released by
Dendrogram showing the grouping of amino acid sequences of prokaryotic and eukaryotic DNA binding proteins, generated from the output of the PILEUP program in the GCG software package (Devereux et al., 1984). The HU-like sequences selected were from representatives of very divergent Bacteria. This is not a phylogenetic reconstruction, and therefore does not indicate evolutionary distances, but the horizontal branch lengths are proportional to the similarity between sequences or sequence clusters, and clustering orders within sequence groups are correctly represented. Amino acid sequences were taken from publications or GENBANK, as described in Grayling et al. (1994). Sequence names in bold type correspond to archael proteins. The sequences of the HMf protein family are shown in Figure 3, and the consensus sequences for the histones H2A, H2B, H3, and H4 were used (Wells and McBride, 1989).
Figure 2.
osmotic lysis of *T. acidophilum* cells documented the presence of discrete NLS, but these were not arranged in regular arrays (Searcy and Stein, 1980; Bohrmann *et al*., 1990). The protein component of these NLS was presumed to be HTa, as visually similar NLS could be reconstituted *in vitro* using purified HTa, and HTa has been localized by immunogold labelling to the nucleoid in thin sections of *T. acidophilum* cells (Searcy and Stein, 1980; Bohrmann *et al*., 1990). Based on staphylococcal nuclease protection studies and chemical crosslinking, it has been calculated that ~40 bp of DNA are wrapped around an HTa tetramer to form a NLS *in vitro*, although this would require extreme bending of the DNA helix (Searcy and Stein, 1980; Searcy, 1986).

The primary sequence of HTa (DeLange *et al*., 1981) has very limited, but recognizable homology to the sequences of the eukaryal core histones, however phylogenetic analyses indicate that HTa is a distantly-related member of the HU family of bacterial histone-like proteins (Figure 2; Searcy, 1986; Grayling *et al*., 1994).

**The MC1 protein family**

MC1 is the most abundant chromosomal protein present in several *Methanosarcina* species including the thermophile *Methanosarcina* CHTI-55 (Chartier *et al*., 1985; Chartier *et al*., 1988), and several variants of this protein are also present in *Methanothrix soehngenii* (Chartier *et al*., 1989). MC1 has a molecular mass of ~11 kDa and contains a large number of basic and acidic residues. It has a primary sequence with no recognizable homology to the sequences of other eukaryal, bacterial or archaeal DNA binding proteins (Laine *et al*., 1986), and CD spectroscopy indicates that MC1, from *M. barkeri*, has almost no α-helical content (Imbert *et al*., 1990). Immunogold labeling has localized MC1 to the nucleoid regions of *M. barkeri* cells and MC1 binding bends and causes topological changes in DNA molecules *in vitro*, but MC1 generated NLS have not
been observed and MCI binding does not protect DNA molecules from micrococcal nuclease digestion (Chartier et al., 1988; Imbert et al., 1988; Imbert et al., 1990; Laine et al., 1991). Crosslinking studies have identified a tryptophan containing region within the MCI protein from *Methanosarcina* strain CHTI-55 as a site of direct interaction with DNA (Katouzian-Safadi et al., 1991). This protein has been shown, by DNaseI footprinting and gel retardation assays, to bind to DNA sequences with some specificity, but the actual binding site was defined only as a 20 to 30 bp AT-rich sequence (Teyssier et al., 1994).

**DNA binding proteins from *Sulfolobus* species**

Green et al. (1983) isolated two DNA binding proteins, HSa and NHSa, from *S. acidocaldarius* strain 98-3 which were termed 'histone-like' because they protected DNA against staphylococcal nuclease digestion and thermal denaturation. Subsequently, Grote et al. (1986) isolated several small DNA binding proteins from *S. acidocaldarius* DSM1616 which were grouped into three molecular size classes (7 kDa, 8 kDa, and 10 kDa) and were therefore designated Sac 7a through 7e, Sac 8a and 8b, and Sac 10a and 10b, respectively. Microsequencing revealed that Sac 7a, 7b, and 7d were almost identical proteins that had different levels of post-translational methylation of lysine residues (Kimura et al., 1984; Choli et al., 1988a). EM investigations of the complexes formed by binding the Sac 10b protein to DNA *in vitro* revealed helically interwound fibers, in which the DNA molecule was not significantly compacted (Lurz et al., 1986). Similar EM studies with Sso 7d, the homolog of Sac 7d from *S. solfataricus*, demonstrated that highly condensed DNA-protein aggregates were formed which, at high protein to DNA ratios, were surrounded by protein-free DNA loops (Choli et al., 1988b). In neither case were NLS observed. The three-dimensional structure of Sso 7d,
determined in solution by NMR, consists of a triple-stranded antiparallel β-sheet overlaid by an orthogonal double-stranded β-sheet (Baumann et al., 1994). In DNA binding, the triple-stranded β-sheet is proposed to interact with the DNA major groove, with the double stranded β-sheet interacting simultaneously with the minor groove (Baumann et al., 1995). This structure has no similarity to the α-helical histone fold motif (Arents et al., 1991) but is similar to the structure of Src homology-3 (SH3) domains that participate in eukaryal signal transduction (Baumann et al., 1994). It is not clear how the DNA binding activity of Sso 7d correlates with the SH3 structural homology, since SH3 domains are usually found in kinases or phospholipases that do not bind to DNA.

Four acid-soluble, small, basic proteins have been purified from S. acidocaldarius DSM 639 nucleoprotein complexes (Reddy and Suryanarayana, 1988). Binding by three of these, HSNP-A, HSNP-C and HSNP-C' increases the heat denaturation temperature of DNA molecules in vitro, and these proteins have been proposed to prevent DNA strand separation in vivo at the 70 °C optimal growth temperature of S. acidocaldarius. The fourth protein, DBNP-B, does not protect DNA molecules in vitro from heat denaturation, and DBNP-B was found to be exclusively located in the cytoplasm (Bohrmann et al., 1994). CD spectroscopy and predictions based on their primary sequences indicate that these four proteins have different secondary structures and that HSNP-C', DBNP-B, and HSNP-C may correspond to proteins Sac7e, Sac10b, and Sac8a, respectively (Reddy and Suryanarayana, 1989). EM analyses of immunogold labelled thin sections of S. acidocaldarius cells have localized HSNP-A and HSNP-C' to the nucleoid (Bohrmann et al., 1994).

Three ribonucleases, designated p1, p2, and p3, were recently isolated from S. solfataricus and were subjected to complete amino acid sequencing (Fusi et al., 1995). The primary sequences of p2 and p3 were found to be very similar to the sequences of
Sac 7a through 7e, differing from the Sac proteins only at their extreme C-termini. These proteins may all have similar functions \textit{in vivo} that clearly may not be limited to DNA binding and compaction (Fusi \textit{et al.}, 1995).

The HMf family

The HMf family of proteins are the only known \textit{bona fide} prokaryotic homologs of the eukaryal nucleosome core histones. HMf is the archetype of this family, and was isolated from the hyperthermophilic methanogen \textit{Methanothermus fervidus} [optimal growth temperature (\textit{t}_{\text{opt}}) of 83 °C], hence the name (histone from \textit{Mt. fervidus}). Size exclusion chromatography showed that HMf existed as dimers in solution, and reverse-phase HPLC indicated that HMf preparations contained a mixture of HMfA and HMfB monomers. These two small (MW \textasciitilde 7.5 kDa), basic (pI \textasciitilde 9-10) and abundant polypeptides (previously designated HMf1 and HMf2, respectively) have 85% identical amino acid sequences (Figure 3), and have been localized \textit{in vivo} to the \textit{Mt. fervidus} nucleoid (Sandman \textit{et al.}, 1990; Bohrmann \textit{et al.}, 1994; Sandman \textit{et al.}, 1994b). The DNA binding activities of HMfA and HMfB are very resistant to heat inactivation and DNA binding by these proteins substantially increases the melting temperatures of double stranded DNA molecules \textit{in vitro} (Krzycki \textit{et al.}, 1990). The genes, \textit{hmfA} and \textit{hmfB}, that encode HMfA and HMfB, respectively, have been cloned, sequenced, and expressed in \textit{E. coli} to produce recombinant (r)HMfA and rHMfB (Sandman \textit{et al.}, 1990; Tabassum \textit{et al.}, 1992). Similar genes that encode closely related proteins from other \textit{Euryarchaeota} have also been cloned and sequenced, namely HMtA and HMtB from \textit{Methanobacterium thermoautotrophicum} strain ΔH [\textit{t}_{\text{opt}} of 65 °C; (Tabassum \textit{et al.}, 1992)], HFoA1, HFoA2, and HFoB from \textit{Methanobacterium formicicum} [\textit{t}_{\text{opt}} of 37 °C; (Darcy \textit{et al.}, 1995)], and HPyA1 and HPyA2 from \textit{Pyrococcus} strain GB-3a [\textit{t}_{\text{opt}} of 90 °C; (Sandman
Figure 3.

Alignment of the primary sequences of members of the HMf family of archaeal histones (Sandman et al., 1990; Tabassum et al., 1992; Sandman et al., 1994a; Darcy et al., 1995) with consensus sequences for the structured central regions of the eukaryal histones that form the histone fold motif (Wells and McBride, 1989; Arents et al., 1991; see text). The numbers of N-and C-terminal residues that are not included in the eukaryal sequences are indicated in square brackets. Position 1 is defined as the first methionine residue of the HMf proteins in the alignment, and dashes indicate gaps inserted to optimize the alignment. The numbers in parentheses indicate the optimal growth temperatures of the archaeon from which each HMf-like protein was obtained. The boxed regions indicate conserved sites containing similar or identical residues. The three regions known to form α-helical structures are underlined by coils, separated by lines that indicate the unstructured regions of the polypeptides. There is a conserved hydrophobic repeat in all of the proteins that should result in an amphipathic character for Helix II, and the dashed box in this helix identifies a position that is not conserved but does form part of the conserved hydrophobic repeat.
<table>
<thead>
<tr>
<th>Helix I</th>
<th>Helix II</th>
<th>Helix III</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPyA2 (90)</td>
<td>M A E L P I A P V D R L Y R K A G - - - - - A Q R V S E</td>
<td>HPyA2 (90)</td>
</tr>
<tr>
<td>HmFA (83)</td>
<td>M G E L P I A P V D R L Y R K A G - - - - - A E R V S D</td>
<td>HmFA (83)</td>
</tr>
<tr>
<td>HmFB (83)</td>
<td>M A E L P I A P V D R L Y R K A G - - - - - A E R V S D</td>
<td>HmFB (83)</td>
</tr>
<tr>
<td>HmCA (65)</td>
<td>M A E L P I A P V D R L Y R K A G - - - - - A E R V S D</td>
<td>HmCA (65)</td>
</tr>
<tr>
<td>HmCB (65)</td>
<td>M A E L P I A P V D R L Y R K A G - - - - - A E R V S D</td>
<td>HmCB (65)</td>
</tr>
<tr>
<td>H2A (22 aa)</td>
<td>L Q F P P V G R V H R L L R L K Q V H - - - - - - -</td>
<td>H2A (22 aa)</td>
</tr>
<tr>
<td>H2B (33 aa)</td>
<td>K E S Y S I Y Y Y Y K V L K Q V H - - - - - - -</td>
<td>H2B (33 aa)</td>
</tr>
<tr>
<td>H4 (26 aa)</td>
<td>Q G I T K P A I R R L A R R G - - - - - V K R I S G</td>
<td>H4 (26 aa)</td>
</tr>
</tbody>
</table>

Helix I:
- **HPyA1 (90)**
- **HPyA2 (90)**
- **HmFA (83)**
- **HmFB (83)**
- **HmCA (65)**
- **HmCB (65)**
- **HfoA1 (37)**
- **HfoA2 (37)**
- **HfoB (37)**
- **H2A (22 aa)**
- **H2B (33 aa)**
- **H3 (59 aa)**
- **H4 (26 aa)**

Helix II:
- **HPyA1 (90)**
- **HPyA2 (90)**
- **HmFA (83)**
- **HmFB (83)**
- **HmCA (65)**
- **HmCB (65)**
- **HfoA1 (37)**
- **HfoA2 (37)**
- **HfoB (37)**
- **H2A (22 aa)**
- **H2B (33 aa)**
- **H3 (59 aa)**
- **H4 (26 aa)**

Helix III:
- **HPyA1 (90)**
- **HPyA2 (90)**
- **HmFA (83)**
- **HmFB (83)**
- **HmCA (65)**
- **HmCB (65)**
- **HfoA1 (37)**
- **HfoA2 (37)**
- **HfoB (37)**
- **H2A (22 aa)**
- **H2B (33 aa)**
- **H3 (59 aa)**
- **H4 (26 aa)**

Fig. 3.
et al., 1994a)]. Related genes have also been sequenced from *Methanococcus voltae* \( t_{\text{opt}} \) of 37 °C; (Agha-Amiri and Klein, 1993) and a *Thermococcus* strain, ANI \( t_{\text{opt}} \) of −75°C; (D. Musgrave, personal communication), and a recent report suggests that HMf-related histones are also present in *Methanopyrus kandleri* \( t_{\text{opt}} \) of 98 °C; (Kozyavkin et al., 1994).

These different members of the HMf family have very similar amino acid sequences and also, on average, are ~45% similar to the consensus amino acid sequences established for the eukaryal core histones (Figure 3; Table 1). Primary sequence analyses demonstrate that the HMf family of proteins form a distinct group separate from other prokaryotic DNA binding proteins and are most closely related to the eukaryal core histones (Figure 2). Remarkably, the consensus sequence for each eukaryal core histone is, in fact, more similar to the HMf protein sequences than to the consensus sequences of the other core histones (Table 1). It appears, therefore, that the archaeal HMf proteins and the eukaryal histones evolved from a common ancestor that existed before the divergence of the *Archaea* and the *Eukarya* (Sandman, 1994a). Phylogenetic arguments would suggest that this ancestral protein was probably most similar to the contemporary HPy proteins in *Pyrococcus* species (Figure 2).

By aligning the primary sequences of the HMf family of proteins, and by computer modeling using the *PEPTIDESTRUCTURE* program (Devereux et al., 1984), ~60% of the secondary structures of HMfA and HMfB were predicted to be α-helical. This prediction was supported by CD spectroscopy data (Grayling et al., 1994).

Recently, the three-dimensional solution structure of rHMfB was determined by NMR spectroscopy (Starich et al., 1995), and this data was used to refine the secondary structural predictions for the entire HMf protein family (Figure 3). The extensive homology of these proteins with the eukaryal core histones is not only at the primary
<table>
<thead>
<tr>
<th></th>
<th>HMfB</th>
<th>HPyA1</th>
<th>HPyA2</th>
<th>H2A</th>
<th>H2B</th>
<th>H3</th>
<th>H4</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMfA</td>
<td>91</td>
<td>76</td>
<td>75</td>
<td>46</td>
<td>43</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>HMfB</td>
<td>76</td>
<td>74</td>
<td>46</td>
<td>43</td>
<td>41</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>HPyA1</td>
<td>91</td>
<td>47</td>
<td>39</td>
<td>42</td>
<td>42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPyA2</td>
<td>45</td>
<td>42</td>
<td>39</td>
<td>42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H2A</td>
<td>34</td>
<td>37</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H2B</td>
<td>33</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>36</td>
<td></td>
</tr>
</tbody>
</table>

Numbers are the percentage similarities between the amino acid sequences of the listed polypeptides, and were calculated from the alignment shown in Figure 1. Amino acid similarities between sequences were determined using the following equivalencies: A≡G, E≡D, N≡Q, S≡T, R≡K≡H, L≡I≡V≡M, F≡Y≡W. Bold type indicates similarities between the archaeal and eukaryal histone sequences.
sequence level, but is also evident from the lengths, spacings and positioning of the \( \alpha \)-helical regions in the HMf proteins being almost directly superimposable on the \( \alpha \)-helical elements that form the histone fold motifs in the eukaryal histones (Figure 3; Arents et al., 1991; Grayling et al., 1994). The non-\( \alpha \)-helical regions located between the helices are also structurally conserved, and the NMR data demonstrated that rHMfB has a three-dimensional structure that incorporates the histone fold and \( \beta \)-bridge motifs established for the eukaryal histones (Arents et al., 1991; Starich et al., 1995).

The HMf family of proteins bind and compact DNA, as documented by electrophoretic gel mobility shift assays (EMSAs) and visualized directly by EM (Sandman et al., 1990; Tabassum et al., 1992; Darcy et al., 1995). Increasing the protein to DNA ratio results in progressively more compact structures. Sharp kinks are introduced into linear or circular DNA molecules at low ratios, and 'beads on a string' NLS are formed at higher protein to DNA ratios. Formation of these NLS effectively decreases the length of protein-bound DNA molecules relative to protein-free DNA molecules (Howard et al., 1992) and therefore these complexes migrate faster during agarose gel electrophoresis than do the protein-free DNA molecules, resulting in a 'gel acceleration' mobility shift phenomenon (Sandman et al., 1990).

DNA bound by HMf has an \( \sim 11 \) bp helical periodicity and the DNA molecule wrapped around a NLS is constrained in \( \sim 1.5 \) positive toroidal supercoils (Musgrave et al., 1991). Eukaryal nucleosomes, in contrast, constrain the DNA molecule into \( \sim 1.7 \) negative toroidal supercoils and decrease the helical periodicity to an average of \( \sim 10.2 \) bp per helical turn. Musgrave et al. (1991) calculated that each NLS must contain between 90 and 150 bp of DNA, and proposed that at the low protein to DNA ratios that result in kinks, HMf binding causes a net negative supercoiling but, at higher ratios, protein-protein interactions occur between adjacent DNA-bound HMf molecules that result in
DNA wrapping, NLS formation and net positive supercoiling. These ideas were formulated into a model for the DNA binding and wrapping by HMf that results in NLS formation (Figure 4).

Although HMfA and HMfB have very similar primary and secondary structures, they do have significantly different DNA binding properties. DNA binding by homogeneous preparations of rHMfA and of rHMfB, synthesized in E. coli, has been assayed by EMSAs and topoisomer formation, at increasing protein to DNA ratios (Sandman et al., 1994b). In both cases, DNA binding increases the electrophoretic mobility of linear DNA fragments relative to protein-free DNA molecules, as expected, but the maximum increase occurs at lower protein to DNA ratios for rHMfA than for rHMfB. The complexes formed by rHMfB at saturating protein to DNA ratios do, however, have a greater electrophoretic mobility, indicating that they are more compact structures than those formed by rHMfA. Topological analyses support this conclusion. At high protein to DNA ratios, rHMfB introduces many more positive topoisomers (>+15) into circular pUC18 DNA molecules than does rHMfA (+2 to +6) (Sandman et al., 1994b). The transition from net negative to net positive supercoiling, predicted by the model for NLS assembly (Figure 4), was also documented by the topology assays and occurred at protein to DNA mass ratios of ~0.4 to 0.6. The introduction of a NLS must increase the negative superhelicity in the regions of a topologically closed circular DNA molecule that are not wrapped in the NLS. The apparent cooperativity of rHMfB binding to such DNA molecules suggests that rHMfB must bind preferentially to negatively supercoiled DNA, relative to rHMfA (Sandman et al., 1994b).

Native HMf appears to bind preferentially to DNA sequences that are intrinsically curved. Howard et al. (1992) demonstrated that NLS formation occurred four times more frequently at a highly curved Crithidia fasciculata DNA fragment, cloned in the
Figure 4.

Model for nucleosome-like structure (NLS) formation by HMf. The two polypeptide monomers are represented by differently stippled spheres. At low HMf/DNA molar ratios (<50 monomers per pUC18 DNA molecule), protein binding causes a slight unwinding of the DNA helix and kinks in the DNA molecule (Sandman et al., 1990; Musgrave et al., 1991), and protein-bound relaxed closed circular DNA molecules are constrained with a net negative superhelicity. At higher HMf/DNA molar ratios, protein-protein interactions occur that result in NLS formation and DNA compaction. Since positive superhelicity is introduced into a circular DNA molecule by this wrapping, that is much greater in magnitude than the negative superhelicity introduced by the kinking, such a molecule is constrained with net positive superhelicity (Sandman et al., 1994b). This model indicates that only HMf dimers pre-bound to DNA interact to form the tetramer core of a NLS, however, a single pre-bound dimer could also recruit a second dimer from solution.
• HMf/DNA < 50
• Protein binding 'kinks' DNA
• Net negative superhelicity results from protein-induced helical underwinding of the duplex

Protein-Protein Interaction

• HMf/DNA > 50
• NLS formation and DNA compaction
• Net positive superhelicity results from toroidal wrapping

Figure 4.
plasmid pBR325, than at other non-curved regions of the same plasmid. The intergenic regions of methanogen genomes frequently contain oligo-(dA) tracts (Brown et al., 1989), which, when appropriately phased, should form highly curved regions (Crothers et al., 1990) that may therefore be preferential sites for NLS assembly. Binding of HMf to template DNAs has been shown to inhibit transcription in vitro in a transcription system derived from Methanococcus thermolithotrophicus (Thomm et al., 1992), and therefore regulation of the assembly and disassembly of NLS at preferential sites in intergenic regions could directly regulate transcription initiation in vivo in Mt. fervidus.

The DNA binding activities of the HMf family of proteins are all somewhat heat resistant, and the differences in their heat resistances reflect the growth temperature of the archaeon from which the protein was isolated. The overall heat resistance of these proteins is not surprising considering their small sizes and simple \( \alpha \)-helical structures, however, that there are differences in their heat resistances is surprising, given the extreme conservation in their primary sequences (Figure 3). For example, the DNA binding activities of HMf (from Mt. fervidus, \( t_{opt} \) of 83 °C) and HMt (from Mb. thermoautotrophicum, \( t_{opt} \) of 65 °C) have half lives in excess of 5 h at 95 °C, whereas the activity of HFo (from Mb. formicicum, \( t_{opt} \) of 37 °C) has a half life of \(~2\) h at 95 °C (Darcy et al., 1995), even though there are only 7 non-conservative amino acid differences in the primary sequences of the HFo and HMf proteins. These changes must result in subtle changes in hydrophobic interactions and intrahelical salt bridges that affect the stability of \( \alpha \)-helices within the tertiary structures of these proteins. Extrinsic factors, such as the high (~300 mM) intracellular K\(_3\)DPG concentrations present in Mt. fervidus, probably also contribute to the increased stability of the HMfA and HMfB proteins in vivo. Stroup and Reeve (1992) showed that HMf preparations bind to DNA in such a high salt environment, demonstrating that the HMf proteins are not only stable but also
functional under these conditions. The very high salt concentrations in *Mt. fervidus* cells would be expected to prevent DNA strand separation *in vivo* (Bowater *et al.*, 1994), but proteins, such as HMf, that wrap DNA in positive toroidal supercoils should increase the negative superhelicity of the protein-free regions of the topologically closed genome. HMf wrapping might therefore actually be required *in vivo* for processes such as transcription and replication, which require DNA strand separation elsewhere in the *Mt. fervidus* genome.

**Curved DNA**

Almost all DNA sequences possess some curvature, resulting either from the local base sequence (*intrinsic* curvature), or as a response to interactions with proteins or other ligands (Trifonov, 1991; Yang *et al.*, 1995). Intrinsically curved DNA was discovered in kinetoplast minicircle DNAs from *Leishmania tarentolae*, which were found to contain (dA)$_n$ tracts in phase with the helical repeat (Marini *et al.*, 1982; Diekmann and Wang, 1985; Wu and Crothers, 1984). Statistical analyses of DNA sequences also indicated a periodicity of some dinucleotides with the helical repeat in eukaryal DNA, that was suggested to reflect intrinsic DNA curvature (Trifonov and Sussman, 1980). A very useful practical observation associated with these discoveries was that curved DNA molecules migrate through polyacrylamide gels more slowly than non-curved DNA molecules of the same size (Hagerman, 1992), and this property has been exploited to develop assays to detect and quantify both intrinsic and protein-induced DNA bending (Crothers *et al.*, 1991).

Intrinsic DNA curvature arises from differences in the roll, tilt, and twist angles, together termed the *wedge angles*, that exist between each of the 16 possible dinucleotide base pairs (Trifonov, 1991) (Figure 5A). These differences result in different deflections
Figure 5.

Models indicating the rotational degrees of freedom of nucleotide base pairs, and the formation of local DNA curvature. A. Model of a planar base pair, showing possible tilt, roll, and twist axes, and the corresponding angles which have been measured for each dinucleotide bp (Bolshoy et al., 1991; see text). B. Formation of local DNA curvature. Basepairs that give rise to significant wedge angles (arrows) introduce a net curvature into an otherwise linear DNA molecule, when they are positioned in phase with the helical period. These drawings are modified from Trifonov (1985).
A.

Twist

Roll ($\theta_R$)

Tilt ($\theta_T$)

B.

Figure 5.
of the DNA helical axis towards both the major groove (roll) and the phosphate backbone (tilt). If the appropriate dinucleotides are positioned in phase with the helical repeat (every 10-11 bp), the combined helical deflections can result in local DNA curvature (Figure 5B). Intrinsic curvature as a result of phased (dA)$_2$ dinucleotides and larger phased (dA)$_n$ tracts has been extensively studied. The largest curvatures in natural and synthetic DNAs result from (dA)$_{4-6}$ tracts repeated in phase with the helical period (Crothers et al., 1990; Haran et al., 1994), however several natural DNA fragments have been isolated that are highly curved, but contain no (dA)$_n$ tracts nor (dA)$_2$ dinucleotides (Trifonov, 1991). Using a large number of synthetic and natural DNA fragments, the entire set of 10 wedge angles appropriate to the 16 possible dinucleotide pairs have been measured and the largest helical axis deflections (5 to 8°) found to be associated with the dinucleotides AG, AA, CG, GA, and GC (Bolshoy et al., 1991).

Curved DNA has been documented to be important in the initiation of DNA replication, in transcription, and in site-specific recombination (Travers, 1990; Harrington, 1992; Kahn and Crothers, 1993; Travers and Schwabe, 1993; Segall et al., 1994; Goosen and van de Putte, 1995). Proteins that bend DNA convert some of the energy of the bending that they mediate into local unwinding of the double helix to facilitate these macromolecular events, all of which require DNA strand separation (Travers, 1990). From a thermodynamic viewpoint, proteins that bend DNA should prefer to bind preferentially to DNA sequences that already have the appropriate intrinsic curvature (Travers, 1990; Harrington, 1992; Kahn and Crothers, 1993), however, in reality, intrinsic flexibility, the ability of a particular DNA sequence to be bent, is probably as important as intrinsic curvature for protein binding (Hagerman, 1988; Travers, 1990). That the functional role of many DNA binding proteins is to bend DNA has been demonstrated by experiments which replaced natural protein binding sites in
promoter regions with synthetic, intrinsically curved DNA sequences. In these experiments, normal levels of transcription initiation were observed in vivo (Collis et al., 1989; Bracco et al., 1989). Also, in E. coli, IHF which is required to form the 'intasome' structure in phage lambda integration, can be functionally replaced by HU protein, or by the unrelated eukaryal proteins HMG1 and HMG2, all of which bend DNA in a manner similar to IHF (Segall et al., 1994).

Intrinsically curved DNA participates in the localization of eukaryal nucleosomes (Thoma, 1992), and many nucleosome-localizing sequences have been shown to be both intrinsically flexible and intrinsically curved (Trifonov, 1991; Shrader and Crothers, 1990; Thoma, 1992). In addition, the curvature introduced by specific dinucleotides also apparently directs the rotational positioning of the DNA around the nucleosome. AT-rich sequences appear to be preferentially located where the minor groove faces towards the histone octamer surface, whereas GC-rich sequences are located in the opposite orientation, facing away from the octamer surface, and this situation is likely to accommodate bending that results from AT and GC wedge angles, respectively (Shrader and Crothers, 1990; Trifonov, 1991; Travers, 1994). By analogy, similar dinucleotide 'positioning' signals are expected to exist in DNA sequences that direct the positioning of the HMf-based NLS.

Circular dichroism spectroscopy.

Most biological molecules are chiral, and therefore are optically active and rotate the plane of linearly polarized light. Linearly polarized light can be considered as the sum of left- and right-handed circularly polarized components of equal amplitude, where the electric vector of the propagating electromagnetic wave rotates in a left- or right-handed sense. When linearly polarized light interacts with an optically active molecule, unequal
absorption of the left and right-handed components occurs, giving rise to two separate extinction coefficients ($\varepsilon_L$ and $\varepsilon_R$) at each wavelength. The circular dichroism (CD) of a molecule is defined as the difference between the extinction coefficients for left- and right-handed circularly polarized light, at a particular wavelength (Alder et al., 1973; Johnson, 1988). Linearly polarized light that is differentially absorbed in such a way by a chiral sample will become elliptically polarized, such that the electric vector rotates in an elliptical fashion. The arctangent of the ratio of the minor to major axes of the ellipse traced out by the electric vector is defined as the ellipticity of the absorption at each wavelength. Although CD spectrometers do not directly measure this effect, and instead measure $(\varepsilon_L - \varepsilon_R)$ or $|\text{Absorbance}(L) - \text{Absorbance}(R)|$ values, for historic reasons, CD values are often expressed in terms of the ellipticity, $\theta$, measured in degrees.

The dominant CD effects in proteins result from absorption of light by amide bonds that leads to two electronic transitions (excitations), an $n\pi^*$ transition at ~220 nm, and a $\pi\pi^*$ transition near 190 nm (Woody, 1995). The number and magnitude of these transitions are, however, sensitive to the local protein structure, and therefore the CD of a protein is extremely sensitive to its secondary structure. In particular, CD spectroscopy provides a very sensitive, albeit empirical, estimate of the $\alpha$-helix, $\beta$-sheet, $\beta$-turn, and random coil components of a protein's secondary structure (Freifelder, 1982; Johnson, 1988).

A very useful application of CD spectroscopy is to monitor changes in the secondary structures of a protein. CD spectroscopy can be used to detect and measure conformational changes in proteins in solution that result from changes in salt, pH, temperature, ligand binding, and quaternary structure (Johnson, 1988). Frequently, CD measurements of a protein in solution with different salts, at different pH values, and in the presence of denaturants such as urea of guanidine, are used to determine protein
stability and to infer electrostatic and folding properties (Pace, 1986; Johnson, 1988; Fink et al, 1994). Measurement of CD as a function of temperature can, in addition, be used to define the thermal stability and thermodynamic properties of many proteins, also over wide ranges of solution conditions (Johnson, 1988). For these applications, CD of proteins is usually measured at 220 or 222 nm because the magnitude of the CD over the wavelength range of the nπ* amide electronic transition (215 to 235 nm, but which is centered at ~220 nm), varies linearly with α-helical content (Schellman and Becktel, 1983). Almost all proteins contain at least some α-helical regions and therefore changes in CD at 220 or 222 nm are commonly used to measure protein unfolding and/or refolding in conformational studies (Schellman and Becktel, 1983; Johnson, 1988).

**Thermophiles, and protein and DNA thermal stability**

The existence of thermophiles, organisms that live at high temperatures, was first demonstrated by Brock and Freeze (1969) by the isolation of the thermophilic bacterium *Thermus aquaticus*. Subsequently, many thermophilic and hyperthermophilic microorganisms have been isolated, and with the three-domain taxonomy system developed by Woese et al. (1990), many of these have been identified as Archaea. Speculations on the origin of life and the Earth’s pre-biotic environment (Woese, 1987; Pace, 1991), have led to the suggestion that the ancestor of all modern life was likely to have been a thermophilic microorganism.

**Protein thermostability**

Life at high temperature clearly requires cellular constituents and proteins that must be stable and biologically active at high temperature (Gottschal and Prins, 1991; Hensel et al., 1991). Determining how such proteins resist heat denaturation has
enormous potential for the manipulations of proteins for use in industrial biotechnologies but, so far, the basis of protein thermal stability remains poorly understood (Herbert, 1992; Flam, 1994; Cavagnero, 1995). Comparisons of related enzymes from thermophiles and mesophiles have demonstrated that very few differences may exist in the amino acid sequences of thermostable and non-thermostable variants (Adams, 1993), but to date there is no consistent definition of the properties that correlate with increased protein thermostability.

Proteins at high temperatures face two problems that can lead to loss of biological function; covalent modifications of side chains and/or the polypeptide backbone, and conformational instability, leading to unfolding of secondary and tertiary structures. At high temperature, the rates of Asn and Gln deamidation, hydrolysis of peptide bonds at Asp residues, destruction of disulfide bonds, and oxidation of Cys and Met residues are increased (Hensel et al., 1991; Volkin et al., 1995). The decreased numbers of Asn and sulfur-containing residues in two glyceraldehyde-3-phosphate dehydrogenases and in a glutamate dehydrogenase from hyperthermophilic *Archaea* has been suggested to increase the stability of these enzymes by reduction of covalent modification, relative to similar enzymes from mesophiles (Hensel et al., 1991; DiRuggiero et al., 1993). In addition, potentially labile residues that were present in these proteins were found to be buried within solvent-inaccessible regions that were protected by the proteins' conformational stabilities at high temperature (Hensel et al., 1991).

Most studies of protein thermostability focus on conformational stability, which is determined by both factors inherent in protein primary sequences, and extrinsic factors such as stabilizing salts and small molecules. Increased protein thermostability seems to result from sequence-dependent features that include increased hydrophobic interactions in ‘buried’ portions of proteins that may result from increased hydrophobic interactions
between amphipathic α-helices, and increased alanine content in all α-helices that leads to reduced flexibility and therefore conformational rigidity (Menéndez-Arias and Argos, 1989; Davies et al., 1993; Russell et al., 1994). In several thermostable enzymes, the lengths of unstructured loops that connect defined protein secondary structural elements are reduced, relative to their mesophile-derived homologs, which could increase conformational rigidity and lead to more compact, stable structures (Davies et al., 1993; Russell et al., 1994). Ion pairs ('salt bridges') can stabilize secondary structural elements, and the stabilization of the dipole moments in an α-helix, by charged residues appropriately positioned at the N- and C-termini of the helix, may also be an important thermostabilizing factor (Davies et al., 1993; Kelly et al., 1993). A theme emerging from the three-dimensional structural studies of a thermostable rubredoxin and ferredoxin from *Pyrococcus furiosus* is the lack of extended, disordered N-terminal tails (Teng et al., 1994; Cavagnero et al., 1995; Starich et al., 1995). Similarly, the three-dimensional structure of rHMFtB has very short, unstructured N- and C-termini and apparently is also stabilized by two unique α-helical ‘proline N-caps’ and several interhelical hydrogen bonds (Starich et al., 1995).

Some of the features that are thought to provide heat resistance to proteins isolated from thermophiles appear also to result in increased stability towards high salt concentrations. For example, malate dehydrogenase from the halophilic archaeon *Haloarcula marismortui* is stabilized in high (~4 M) salt apparently by having increased α-helical Ala content, by helix dipole stabilization, and by increased internal hydrophobicity, features that appear to provide heat resistance to the homologous enzymes from *Thermus flavus* and *Bacillus stearothermophilus* (Dym et al., 1995). This is very interesting, as salt has been proposed in several cases as a significant extrinsic factor in providing enzyme thermostability. The archaeal species *Mt. fervidus, Mt.*
sociabilis, Mt. kandleri, and P. woesei all contain very high intracellular potassium concentrations (0.5-2.3 M), counterbalanced in Mt. fervidus, Mt. sociabilis, and Mp. kandleri by the unusual organic phosphate anion cyclic-2, 3-diphosphoglycerate (cDPG), and by di-myo-inositol-1, 1'-phosphate in P. woesei (Hensel and König, 1988; Huber et al., 1989; Scholz et al., 1992). These compounds clearly increase the heat resistance of Mt. fervidus and Mt. kandleri enzymes more than other salts (Hensel and König, 1988; Breitung et al., 1992; Hensel and Jakob, 1994), and therefore are extrinsic factors in the thermostability of these enzymes.

DNA thermal stability

Thermophiles, in addition to having heat resistant enzymes, must also protect their nucleic acids from heat denaturation and destruction. DNA molecules are subject to a wide range of destructive reactions such as the generation of apyrimidinic sites, and deamination of cytosine and 5-methylcytosine, that are accelerated in vitro, and presumably also in vivo, at high temperature (Lindahl, 1993). DNA strand separation should also be a problem at high temperatures, although thermal denaturation of linear, double-stranded DNA molecules is prevented by a high concentration of salt. The very high intracellular salt concentrations in hyperthermophiles such as Mt. fervidus probably therefore prevent the heat denaturation of their genomes, regardless of the growth temperature (Hensel and König, 1988; Bowater et al., 1994; Grayling et al., 1995b). DNA binding by proteins such as HMf might still be important for the thermostabilization of the genomic DNA of Mt. fervidus, but may not be essential. Marguet and Forterre (1994) have demonstrated that relaxed, or negatively or positively supercoiled, covalently closed circular plasmid DNA molecules are inherently resistant to heat denaturation up to
107 °C. They therefore concluded that spontaneous DNA strand separation should not be a problem \textit{in vivo} for hyperthermophiles.

\textbf{Goals of this study}

The HMf proteins are related to the eukaryal histones H2A, H2B, H3, and H4. This is apparent in their primary sequences, and secondary and tertiary structures. Furthermore, these archaeal proteins form DNA-protein complexes that resemble nucleosomes, and are apparently localized by intrinsically curved DNA sequences. It is hypothesized, therefore, that the HMf family of proteins are prokaryotic histones and should be described as \textit{archaeal histones}, and four projects were defined to investigate the extent of this homology between the archaeal and eukaryal histones. The goals were: (1) to characterize in detail, the biochemical properties of the HMfA and HMfB protein components of the HMf-based NLS; (2) to characterize the structural properties of HMfA and HMfB with respect to their solution ionic environments and to heat; (3) to define the DNA binding preferences of the HMf proteins; and (4) to define the overall structure of the intact HMf-DNA complexes that form the NLS visualized by EM.
CHAPTER II
PURIFICATION AND BIOCHEMICAL CHARACTERIZATION OF
NATIVE AND RECOMBINANT HMF PROTEINS

INTRODUCTION

Native HMF protein was originally purified from stationary phase *Mt. fervidus* cultures that were not grown optimally and did not exhibit a sustained exponential growth phase or result in significant cell yields (final OD$ _{580} \leq 0.8$) (Sandman *et al.*, 1994b). The original purification method relied on size exclusion chromatography (SEC) as a first step, and was therefore limited to preparation of small amounts (0.5-2 mg) of HMF (Krzycki *et al.*, 1990). Furthermore, the purification process was monitored by screening for the DNA binding activity of HMF by agarose EMSA (Sandman *et al.*, 1990), which was tedious and not necessarily reliable due to the presence of other DNA binding proteins. Using SEC, HPLC, and acetic acid-urea PAGE, native preparations of HMF purified using stationary phase *Mt. fervidus* cultures were found to exist as dimers and to be composed of a mixture of HMF$_A$ and HMF$_B$ (previously denoted HMF1 and HMF2, respectively), in approximately equal amounts (Krzycki *et al.*, 1990). It was assumed, from these and subsequent studies, that the significant form of HMF *in vivo* was a heterodimer of the HMF$_A$ and HMF$_B$ monomers since these were always observed in an ~1:1 ratio. Alternatively, it was suggested that native HMF could exist as homodimers of HMF$_A$ and HMF$_B$, or as a mixture of homodimers and heterodimers (Krzycki *et al.*, 1990).
Biochemical and biophysical experiments were planned for HMf, and therefore a large amount (25-50 mg) of highly purified protein was required. A major goal and necessary first step for this project was therefore to optimize both the fermentation conditions for *M. fervidus*, and to develop a reliable purification method amenable to scale-up. During development of the purification method, variations in the HMfA/HMfB ratio were observed that seemed to correlate with the growth stage at which the cells were harvested. Therefore, a systematic study was undertaken to investigate the apparent growth-phase variation of the HMfA/HMfB ratio. The results of this study indicated that the levels of HMfA and HMfB are growth-phase dependent (Sandman *et al.*, 1994b), and meant that preparations of native HMf were therefore complex mixtures and *not* suitable for detailed biochemical or biophysical analyses because of their variable compositions.

The genes *hmfA* and *hmjB* that encode HMfA and HMfB have been cloned and sequenced (Sandman *et al.*, 1990; Tabassum *et al.*, 1992). Expression of these in *E. coli* was therefore used to purify recombinant (r) forms of HMfA and HMfB, to circumvent the variability problem encountered with native protein preparations (Sandman *et al.*, 1994b). However, it was found, by using the non-denaturing gel system developed here, that rHMfA preparations also were non-homogeneous. This chapter documents the development of fermentation and protein purification conditions, the growth phase-dependence of the HMfA/HMfB ratio, and the discovery of the heterogeneity problem associated with rHMfA expression in *E. coli*, as well as characterization of HMfA and HMfB by crosslinking and PAGE methods.
MATERIALS AND METHODS

Chemicals and reagents

Chemicals and reagents were purchased from Sigma Chemical Co. (St. Louis, MO).

Growth of *Methanothermus fervidus*

*Mt. fervidus* (DSM #2088) was obtained from Dr. H. Hippe, at Deutsche Sammlung für Mikroorganismen (Göttingen, Germany). Small-scale cultures of 20 ml or 200 ml were grown anaerobically at 83 °C in stoppered 100 ml serum bottles or 11 ‘Wheaton’ glass bottles, under a headspace of 0.2 MPa (15 psi) 4:1 (v/v) H2:CO2, using the rich medium described below. The type strain was passaged once through rich growth medium, and stored both frozen anaerobically under nitrogen with glycerol (50% v/v) in liquid nitrogen, and in the original serum bottle at 4 °C, before being used to seed fermentor or further small-scale cultures.

Fermentor cultures of *Mt. fervidus* were grown anaerobically at 83 °C in a medium that contained (per l): 0.3 g K2HPO4; 0.3 g KH2PO4; 0.3 g (NH4)2SO4; 0.6 g NaCl; 5 g NaHCO3; 65 mg MgSO4; 50 mg CaCl2; 2.5 g Na acetate; 3.3 mg Na2WO4-H2O; 1 mg resazurin; 2 g yeast extract; 2 g tryptone and 10 ml of trace metal solution (Nölling *et al.*, 1991). After sterilization, 0.5 g of Na2S-9H2O and 0.5 g of cysteine-HCl were added and the pH adjusted to 6.5. Cultures (20 l) were grown in a MicroFerm model CMF-128S fermentor (New Brunswick Scientific Co., New Brunswick, N.J.) inoculated with 5% (v/v) of a *Mt. fervidus* culture grown to an optical density at 580 nm of ~1. The fermentor vessel was pressurized to 0.2 MPa (15 psi) with 4:1 (v/v) H2:CO2 and stirred slowly (impeller setting of 50 rpm) for 4 h. Following the onset of
exponential growth, the culture was sparged with 4:1 (v/v) H2:CO2 at a flow rate of 1 l/min and stirred at an impeller setting of 250 rpm. Regulation of pH was achieved by automated addition of glacial acetic acid to maintain the culture pH between 6.6 and 6.8. Growth was monitored by following OD_{580}. Cells were concentrated aerobically by tangential flow filtration to ~2 l (Pellicon; Millipore Corp., Milford, MA), followed by centrifugation and freezing in liquid nitrogen, and storage at -70°C.

**Growth of *Escherichia coli***

Strains of *E. coli* JM105 (Yanisch-Perron *et al.*, 1985) used were obtained from K. Sandman (this laboratory), and contained derivatives of pKK223-3 (Brosius and Holy, 1984). Genes cloned in the polylinker region of this plasmid are placed under the control of the *tac* promoter, and are therefore IPTG inducible. The *E. coli* strains used are designated KS1124, containing plasmid pKS354; KS1138, containing pKS366, KS1183, containing pKS395; and KS1076, containing plasmid pKS323. Construction of these plasmids is described in Tabassum *et al.* (1992), and Sandman *et al.* (1994b; 1995). Briefly, pKS354 includes the *hmfA* gene and part of a downstream ORF encoding a ferredoxin or hydrogenase. For some reason, this 3´ sequence results in extreme overexpression of HMfA, to such an extent that KS1124 cells begin to lyse 6 h post-induction with IPTG. Plasmid pKS366 contains *hmfA* with the 3´ region truncated before the second ORF. KS1138 strains synthesize ~5-fold less HMfA, and do not lyse post-induction. Plasmid pKS395 also contains the *hmfA* gene, excluding the 3´ ORF, plus the *E. coli* methionine aminopeptidase-encoding *map* gene. Plasmid pKS323, contains the *hmfB* gene.

Cultures of these *E. coli* strains were grown aerobically at 37 °C in Luria-Bertani (LB) medium (pH 7) containing 100 µg ampicillin/ml (Sambrook *et al.*, 1989), in 10 l
fermentors (New Brunswick Scientific Co., New Brunswick, N.J.), using a 1% (v/v) inoculum grown in the same medium. The expression of hmfA or hmfB in strains KS1124, KS1138, and KS1076 was induced by adding IPTG (400 μM final concentration) to cultures growing at an OD600 ~0.6-0.8, and growth was continued for 6-16 h before the cells were harvested as described for Mt. fervidus. Expression of hmfA from KS1183 was carried out with trimethoprim (TMP) and thymidine (Tdr) additions to the medium, according to Sandman et al. (1995).

**HMf purification methods**

*Mt. fervidus* or *E. coli* cells were resuspended in a high salt buffer (HS buffer; 3 M NaCl, 50 mM Tris-HCl, pH 8), ruptured by passage through a French pressure cell at 138 MPa (20,000 psi) and cleared supernatants were prepared from the lysates by centrifugation (25,000 x g, 4 °C, 30 min, then 125,000 x g, 25 °C, 90 min). Following overnight dialysis of the supernatant at 4 °C against a low salt buffer (LS buffer; 100 mM NaCl, 50 mM Tris-HCl, pH 8), MgCl₂ and PMSF (5 mM and 100 μM final concentrations, respectively) and DNaseI (50-100 μg/ml) were added, and the mixture incubated at 37 °C for 2-4 h. Solid NaCl (to 3 M final concentration) was then added and the mixture incubated at 95 °C for 10 min. Denatured proteins were removed by centrifugation and filtration through a 0.45 μm Millipore membrane. The resulting solution was dialyzed against LS buffer. To obtain preparations of the native HMf from *Mt. fervidus* lysates, proteins in this solution were adsorbed to the matrix of a dsDNA-cellulose column equilibrated with LS buffer, washed with LS buffer, and eluted with 1 M NaCl, 50 mM Tris-HCl (pH8). To obtain preparations of rHMfA or rHMfB from *E. coli* lysates, proteins were adsorbed to the matrix of a Hi-Trap heparin-sepharose column (Pharmacia, Piscataway, NJ) equilibrated with LS buffer, washed with
LS-buffer and then eluted with a 0.1 M to 1.5 M linear NaCl gradient in 50 mM Tris-HCl (pH 8).

For purification of native HMf for measurement of HMfA/HMfB ratios, control experiments were performed which used a modified LS buffer containing 200 mM NaCl, instead of 100 mM NaCl. In addition, for some of the dialyzed, DNaseI-treated, cleared *M. fervidus* lysates, native HMf was purified by SEC on Sephacryl S-100 HR matrix (Pharmacia, Piscataway, NJ) equilibrated with the modified LS buffer. Fractions from the DNA cellulose, Heparin-Sepharose, and Sephacryl S-100 HR columns that contained native HMf, rHMfA or rHMfB were identified by SDS-PAGE, combined and then concentrated by ultrafiltration in an Amicon mini-stirred cell (10 ml) using YM-1 membranes (Amicon Inc., Beverly, MA). Purity of final preparations was monitored by tricine-SDS-PAGE.

**Protein quantitation**

Native HMf, rHMfA, and rHMfB concentrations were determined using a modified Bradford assay (Congdon *et al.*, 1993). However, because HMfA and HMfB contain a larger than average proportion of lysine and arginine residues for their size (13/68 and 15/69 residues, respectively), Bradford quantitations overestimated the concentrations of these proteins. Protein concentrations were therefore determined in parallel for each protein, by amino acid analysis in triplicate, using a Waters Picotag amino acid analysis system (Millipore Corp., Milford, MA). Correction factors to offset the overestimation of the Bradford assay thus calculated were 0.73 and 0.69 for rHMfA and rHMfB, respectively, produced by strain KS1183 (grown with TMP + Tdr) and strain KS1076.
**Reverse phase HPLC**

Protein preparations were diluted in 0.1% (v/v) trifluoroacetic acid (TFA) and were adsorbed to a Microsorb-MV C8-reverse phase column [100Å pore size; 25cm length; (Rainin Instrument Co., Emeryville, CA)] in a mobile phase that contained 5% solvent B (70% CH3CN; 0.085% TFA) in solvent A (0.1% TFA). HPLC, using a 55% to 75% gradient of solvent B in solvent A, resolved HMfA from HMfB and the peak assignments were verified by electrophoresis of fractions through acetic acid/urea polyacrylamide gels that contained 0.8% (v/v) Triton X-100 (Smith, 1984), using rHMfA and rHMfB as standards (data not shown). At least two chromatograms were recorded for each sample. Peak areas were calculated by numerical integration with *Kaleidagraph* (Abelbeck Software), using a valley-to-valley approximation for chromatogram baselines.

**Protein-protein crosslinking**

Crosslinking using either dimethyl suberimidate (DMSI) or formaldehyde (HCHO), was done with crosslinker stock solutions of 10 mg/ml or 400 mM, respectively, made fresh in 100 mM of the same buffer used in the crosslinking reaction. The results documented here were obtained by the addition of 1 µl of crosslinker stock solution to 9 µl of a solution containing rHMfA or rHMfB (0.5-1.5 µg), KCl or NaCl (diluted as appropriate from a 3.5 M stock solution), and 25 mM of the appropriate buffer (diluted from a 100 mM stock solution). The buffers used were potassium phosphate, pH 6; potassium phosphate, pH 7.5; triethanolamine-HCl, pH 8; and sodium borate, pH 10. The reaction mixtures were incubated at 25°C for 1 h (unless otherwise indicated in the legend to Figure 11), and then quenched by addition of 1 µl of 0.5 M ammonium bicarbonate, and continued incubation for 15 min at 25°C. The reaction mixtures were
then dialyzed against 100 mM NaCl, 25 mM Tris-HCl (pH 7.5), mixed with an equal volume of sample buffer [8% (w/v) SDS, 20% (v/v) glycerol, 100 mM Tris-HCl, 4% (v/v) 2-mercaptoethanol, 0.01% (w/v) bromophenol blue (pH 6.8)] and resolved by tricine-SDS PAGE.

**Formation of rHMfA•rHMfB heterodimer-containing mixtures *in vitro***

Protein mixtures containing rHMfA•rHMfB heterodimers were prepared by incubating equimolar amounts of homogeneous preparations of rHMfA and rHMfB in either 30 mM histidine, 30 mM MES, 200 mM NaCl (pH 6.1), or in 100 mM KCl, 10 mM Tris-HCl (pH 7.5), at 95°C for 5 min, followed by slow cooling to room temperature (~30 min). These heat-treated mixtures were used for both electrophoresis studies and for the CD studies described in Chapter 3.

**Electrophoresis techniques**

**A. One dimension, denaturing.**

Initial denaturing separations used standard SDS-PAGE, according to Laemmli (1970). However, it was observed that the HMf proteins migrated with or were disturbed by the ion front in Laemmli gels, even at high (20% T) acrylamide concentrations, and this caused confusion in some assays. Therefore, a tricine-SDS-PAGE system with a 16.5% T, 3% C acrylamide resolving gel and a 10% T, 3% C spacer gel, which separates small proteins such as HMf from the ion front, was used for most subsequent denaturing PAGE analyses (Schägger and Von Jagow, 1987). Tricine-SDS PAGE was carried out at 2-2.5 V/cm for 1 h, then at 8.5-10.5 V/cm for 2.5-3.5 h. Sample loading buffers contained 2-mercaptoethanol.
B. One dimension, non-denaturing.

Two different non-denaturing (native) electrophoresis systems were developed to separate the dimers in HMf preparations on the basis of differences in their net positive charges. A continuous, single phase system was used to resolve the components of the heat treated mixtures of rHMfA and rHMfB. This system used a 30 mM histidine, 30 mM MES, pH 6.1, (H-MES) electrode buffer, a 16% T, 3% C polyacrylamide resolving gel containing H-MES and 10% glycerol, and did not incorporate a stacking gel. Protein samples were dissolved in 2X H-MES buffer that contained 20% (v/v) glycerol and 0.01% (w/v) pyronin Y, and electrophoresed at 7 V/cm. A discontinuous, two phase system was developed from first principles, as discussed by Williams and Reisfeld (1964) and Richards et al. (1965), and employed a 4% T, 3% C acrylamide stacking gel, and a 16.5% T, 3% C resolving gel. This discontinuous system gave better, more reproducible resolution of the different HMf dimer populations than the continuous system. The stacking gel contained 30 mM PIPES, 10% (v/v) glycerol, adjusted to pH 7.4 with KOH, and the resolving gel contained 70 mM PIPES, 10% (v/v) glycerol, adjusted to pH 6.5 with KOH. The electrode buffer was 84 mM histidine, 37 mM PIPES, (pH 6.1). Protein samples were mixed with an equal volume of 30 mM PIPES, 20% (v/v) glycerol, 0.005% (w/v) Pyronin Y (adjusted to pH 7.4 with KOH), before being loaded on the gel. Electrophoresis with the discontinuous system was carried out at 10-12 V/cm.

C. Two dimensions.

Electrophoresis in a second dimension was used to separate and identify the components of the dimer bands that formed during non-denaturing electrophoresis using the discontinuous system. A denaturing acetic acid-urea PAGE system that incorporated 0.8% (v/v) Triton X-100 (AUT-PAGE; Smith, 1984) was used for this second
dimension. Proteins resolved in the second dimension were quantified by densitometry using known amounts of rHMfA and rHMfB as standards.

D. Gel staining

Resolved proteins were stained using either Coomassie Brilliant Blue R-250 (CBB; Ausubel et al., 1990) or silver (Morrissey, 1981).

Acid hydrolysis ('deformylation') of rHMfA formylated methionine residues

Preparations of rHMfA (3.5 mg/ml) were incubated at 25 °C in 0.6 N HCl. Aliquots were removed at increasing times, neutralized by mixing with an equimolar amount of NaHCO₃, and the resultant protein mixtures separated by non-denaturing PAGE, using the histidine-PIPES discontinuous system.

RESULTS

Development of fermentation and protein purification methods and conditions

Initially, fermentation conditions for the growth of *Mt. fervidus* were based on media and growth conditions formulated by Weil et al. (1988), and resulted in only a brief exponential phase followed by an extended period (up to 60 h) of linear growth with a doubling time of ~18 h, and a final OD₅₈₀ of ≤0.8. Cell yields from these cultures were only 0.7-1.0 g (wet weight)/l. Linear growth has been reported for other methanogens grown on H₂ (see, for example, Bonacker et al., 1992), and is apparently related to H₂ limitation. Since the gross solubility of hydrogen decreases with increasing temperature (Wilhelm et al., 1977), substrate limitation is likely to be a problem for the growth of hyperthermophilic methanogens.
Medium and growth conditions were systematically optimized by variation of medium components (based on Stetter et al., 1981; Balch et al., 1979; Weil et al., 1988; Nölling et al., 1991), pH, \( \text{H}_2 \) sparge rate, impeller rpm, and redox level (modulated by addition of \( \text{Na}_2\text{S} \), cysteine, or titanium (III) citrate). In agreement with Stetter et al. (1981), \( \text{M}t. \text{fervidus} \) was found to be extremely sensitive to pH and redox conditions, but only until cultures had reached an OD\(_{580}\) of ~0.4, after which a lack of sulfide or pH >7 in the culture had a minimal effect on growth (data not shown). Hydrogen limitation is a major factor in achieving exponential growth of \( \text{M}t. \text{fervidus} \), since increasing the impeller stirring speed resulted in reduced doubling times and increased final OD\(_{580}\) values, up to a maximum setting of 250 rpm. Presumably, above 250 rpm increased mass transfer of hydrogen to cells is offset by the increased shearing forces that they must experience.

The optimal growth conditions achieved are described in the Methods section of this chapter, and resulted in reproducible growth curves, as shown in Figure 6A. Both pH-controlled and pH-uncontrolled cultures generated identical growth curves, although when pH was not maintained between 6.6 and 6.8, it varied (reproducibly) over a range from 6.4 during late stationary phase, to 7.6 during exponential growth. Presumably, this variation reflects the balance between the rates of dissolved CO\(_2\) uptake by cells (greatest during exponential phase, leading to alkalinization), and CO\(_2\) dissolving in the medium (greatest during late stationary phase when cells are no longer growing, leading to acidification). Using either pH condition, using the optimized medium and fermentation conditions, doubling times for the cultures were reduced to 9-10 h (Figure 6A), with final cell yields of ~2.5 g (wet weight)/l. This doubling time is, however, still much longer than the optimal 2.8 h reported by Stetter et al. (1981).
**Figure 6.**

Optimized fermentation and protein purification conditions. A. Growth curves for *Mt. fervidus* grown using optimized conditions with (triangles) or without (circles) pH control at pH ~6.7, using automated addition of glacial acetic acid. Optical densities (OD) at 580 nm (solid symbols) and the pH of the cultures (outlined symbols) were monitored. B. Purification of native HMf from lysed *Mt. fervidus* cells. Lysates, first treated with DNaseI and heat, were adsorbed to a dsDNA cellulose column equilibrated with LS buffer (see text). Bound proteins were eluted with 1 M NaCl, 50 mM Tris-HCl, pH 8 in a single step, after extensively washing the column with LS buffer. Samples (~1 μg total protein-except for sample 5, where 20 μl was used), taken at each step during the purification process were analyzed by tricine-SDS-PAGE and silver staining; 1, DNaseI-treated clarified lysate; 2, after heat treatment and centrifugation [pellet]; 3, after heat treatment and centrifugation [supernatant]; 4, flow-through from DNA cellulose column; 5, wash from DNA cellulose column; 6, eluate from DNA cellulose column, after buffer exchange concentration. C. Heparin-sepharose purification of rHMfA from heat-treated *E. coli* cell lysate (strain KS1138). DNaseI and heat-treated lysate (P) was adsorbed to a ‘Hi-Trap’ heparin-sepharose column equilibrated with LS buffer, and the flow through (F) collected. After brief washing with LS buffer, adsorbed proteins were eluted with a 0.1-1 M linear NaCl gradient in 50 mM Tris-HCl (pH 8), and fractions were analyzed by tricine-SDS-PAGE and CBB staining. rHMfA eluted as a broad peak with a center corresponding to ~500 mM NaCl (~600 mM NaCl for rHMfB). The protein size standards (S) were 14.4, 21.5, 31.0, 45.0, 66.2, and 97.4 kDa.
Figure 6.
Protein purification

Native HMf

Biophysical studies requiring 20-50 mg of protein were planned for a major part of this project, initially using native HMf protein. To facilitate scale-up, the original purification of Krzycki et al. (1990) was revised. Many proteins isolated from thermophilic organisms are highly thermostable and a significant fold-purification can often be achieved in a single step by simply heating cell lysates (Englard and Seifter, 1990). This is particularly true of thermophile-derived proteins expressed in *E. coli*. Since the SEC step in the original protocol also separated DNA fragments from HMf protein, both DNasel treatment of clarified cell lysate, and a subsequent heat treatment step were therefore substituted. This procedure was effective (Figure 6B, lanes 2 and 3), and optimal heat-based removal of non-HMf proteins from DNasel-treated *Mt. fervidus* lysates was achieved by incubating at 95 °C for 10 min, in the presence of 3 M NaCl. Under these conditions, many *Mt. fervidus* proteins were also precipitated, and were removed by centrifugation and filtration through 0.45 μM membranes. Ammonium sulfate precipitation was also tested as an alternative to heat treatment as a first purification step, but was not as effective, with >95% of the total protein still being soluble at 80% saturated (NH$_4$)$_2$SO$_4$ (data not shown). Dialysis of the heat treated protein solutions into LS buffer (see Methods) and absorption to dsDNA cellulose, followed by extensive washing with LS buffer, and elution with 1 M NaCl resulted in a single HMf band estimated to be >97% pure, when preparations were analyzed by SDS-PAGE and stained with silver (Figure 6B).
Recombinant HMfA and HMfB

The same purification procedure was attempted for rHMfA and rHMfB expressed in *E. coli*, however it resulted in copurification of the HMf proteins with several low molecular weight *E. coli* proteins, even when a NaCl gradient was used to elute bound protein from the DNA cellulose. *E. coli* possesses several 'histone like' DNA binding proteins which are expressed at relatively high level and are intrinsically heat-resistant, such as HU, H-NS, and IHF (Drlica and Rouvière-Yaniv, 1987; Schmid, 1990). Presumably, these proteins can survive the heat treatment step and bind DNA cellulose, co-purifying with the HMfA and HMfB. The DNA cellulose affinity chromatography step was therefore replaced by a heparin sepharose affinity chromatography step, using pre-packed 5 ml columns (Pharmacia, Piscataway, NJ). Heparin sepharose is a polyanionic matrix with a much greater protein binding capacity than DNA cellulose, and is more stable and more economical. Elution of bound protein from this matrix using a 0.1-1 M NaCl gradient resulted in a broad peak containing >99% HMfA or HMfB, as determined by CBB staining (Figure 6C). Salt gradients of 0.1-1 M KCl and 0.03-0.3 M K3-citrate gave similar results. Native HMf could also be purified using these conditions, but best results were obtained with the K3-citrate gradient.

Because the HMf proteins migrate with the ion front in standard SDS-PAGE systems, the tricine-SDS-PAGE system of Schägger and Von Jagow (1987), which separates such small proteins from the ion front, was used to resolve the HMf proteins (Figure 6B, C).

Storage and solution properties of purified HMf proteins

Purified native HMf, rHMfA and rHMfB were generally stored in 1 M NaCl (or 1 M KCl), 50 mM Tris-HCl, pH 7.5-8, at -20 °C, but were stable in this buffer at 4 °C.
for at least 6 months. No changes in stored proteins were evident by tricine-SDS-PAGE or HPLC analysis after 6 months at 4 °C, and no nucleolytic or proteolytic activities were ever detected. However, storage at 4 °C in buffers that contained <0.2 M NaCl resulted in partial precipitation within two weeks. Precipitates could be re-dissolved by increasing the salt concentration to >0.6 M, and warming to 37 °C, with >90% of the original DNA binding activity recoverable. HMF protein solutions were generally stored at concentrations of 5-10 mg/ml, since above 10 mg/ml precipitation was likely, even in the 1 M NaCl buffer. Proteins stored at -20 °C in buffer containing either low (<0.2 M) or high (≥1 M) salt concentrations were stable for at least 5 years, and probably longer. As is the case for most proteins, repeated freeze-thaw cycles resulted in a gradual loss of HMF activity, as measured by the agarose EMSA.

Both rHMFα and rHMFβ are soluble in 70% ethanol, as is native HMF (Stroup and Reeve, 1992), and this provides an easy way to separate HMF-DNA complexes from unbound protein. Also, since the HMF proteins are highly thermostable (see Chapter 3), inadvertent incubation at room temperatures overnight did not result in detectable losses of DNA binding activity.

**Growth phase dependence of the HMFα/HMFβ ratio**

During optimization of *Mt. fervidus* growth conditions, HMF samples were purified from several different cultures and analyzed by AUT-PAGE and were found to vary in the relative amounts of component HMFα and HMFβ monomers. A systematic study of this phenomenon was therefore undertaken, using the optimized fermentation conditions described. Native HMF was purified (using the DNA cellulose method) from cell samples taken at different times during the growth of a *Mt. fervidus* culture, and HMFα and HMFβ monomers present in these samples were resolved by reverse phase
HPLC. HMfA was found to predominate in HMf preparations isolated from cultures in the exponential growth phase, but the relative amount of HMfB then increased as cultures approached the stationary phase, with HMfB eventually comprising ~50% of HMf preparations isolated from stationary phase cultures (Figure 7). This result was reproducible, with HMfA comprising as much as 80% of the total HMf monomer present in early exponential phase cultures (data not shown).

To confirm that this growth phase dependent variation was real, and not an artifact of the protein purification methods used, control experiments were conducted using both the method that included heat treatment, and a non-denaturing method that substituted a SEC step (Figure 8, methods 1 and 3, respectively). Results from HPLC analyses of native HMf preparations purified by these two methods were compared to results from HPLC analyses of post-DNaseI-treated cleared lysates (Figure 8, method 2). The data resulting from duplicate runs of each sample are summarized in Figure 9, and indicate that the heat treatment/DNA cellulose purification method used did not significantly affect the HMfA/HMfB ratio after the DNaseI treatment step, within the error limits of the HPLC data.

To determine the HMfA/HMfB ratio in vivo, attempts were made to produce mouse monoclonal antibodies specific for HMfA and for HMfB, so that the absolute amounts of HMfA and HMfB in lysed cells could be measured by a competitive ELISA. Clones producing antibodies that were specific for HMfA, or that cross-reacted with both HMfA and HMfB were isolated, however, despite two attempts, clones were not isolated that produced HMfB-specific monoclonal antibodies.

The small peaks with longer retention times than the HMfA and HMfB peaks that they followed (Figure 9) were examined by AUT-PAGE of collected fractions, although this analysis was inconclusive (data not shown). It seemed possible that these small
Figure 7.

Growth phase dependent synthesis of HMfA and HMfB. A. Reverse phase HPLC chromatogram of native HMf preparations isolated from *Mt. fervidus* cells taken from the culture at the timepoints indicated below in panel B. Peak heights on the chromatograms are proportional to absorbance measured at 215 nm. The numbers under the peaks are the percent of the HMf preparation that was formed by that peak [HMfA (peak A), and HMfB (peak B)], calculated from the peak areas.

B. Time course of a *Mt. fervidus* fermentation, with the circled numbers indicating when the samples were taken that were used to generate the protein preparations analyzed in panel A.
Figure 7.
Figure 8.

Design of control experiments used to demonstrate the non-selectivity of the protein purification method used in the growth curve analysis (Figure 7). Three purifications of native HMf were carried out as controls that began with the same initial steps to produce DNaseI-treated, cleared *Mt. fervidus* lysates. Subsequent purifications employed either a heat treatment/DNA cellulose protocol (1), or a SEC protocol (3), followed by reverse phase HPLC analysis of purified products. In addition, a DNaseI-treated lysate was directly analyzed by reverse phase HPLC (2).
Disrupt *M. fervidus* cells in 3 M NaCl, 50 mM Tris-HCl, pH 8

- Low speed centrifugation (30 min; 25,000 x g)
- High speed centrifugation (90 min; 125,000 x g)
- Dialyze into 200 mM NaCl, 50 mM Tris-HCl, pH 8
- Add DNasel to 100 µg/ml; MgCl₂ to 5 mM; PMSF to 0.2 mM; Digest for 2 h at 37 °C
- Adjust [NaCl] to 3 M
- Heat for 15 min at 95 °C
- Remove particulates (centrifugation and 0.45 µm filter)
- Dialyze into 200 mM NaCl, 50 mM Tris-HCl, pH 8
- Bind to DNA cellulose; wash; and elute with 1 M NaCl
- Exchange buffer to 200 mM NaCl, 50 mM Tris-HCl, pH 8 and concentrate (ultrafiltration)
- Run sample on Sephacryl S-100 HR column, equilibrated with 200 mM NaCl, 50 mM Tris-HCl, pH 8
- Assay for HMf containing fractions by agarose gel EMSA
- Concentrate (ultrafiltration)
- HPLC analysis

**Figure 8.**
Figure 9.

Validation of heat-treatment/DNA cellulose purification method as a means to establish HMfA/HMfB ratios in native HMf preparations. Reverse phase HPLC chromatograms of the native HMf preparations, prepared from *Mt. fervidus* cells taken from timepoint 3 in Figure 7B, that were purified either using the size-exclusion (SEC) or heat-treatment/DNA cellulose (HD) methods outlined in Figure 8, or were purified only to the DNaseI-treatment step (Crude). Duplicates of each protein purification (#1, #2) were analyzed. The table lists the percentage of each HMf preparation that was formed by HMfA monomer, as calculated from the HPLC peak areas.
Figure 9.
peaks could represent post-translationally modified populations of HMfA or HMfB monomers, however only the expected dimer mobilities were observed in an analysis of these native HMf preparations by non-denaturing PAGE (data not shown; see Discussion).

**Protein-protein crosslinking studies**

**Confirmation of dimeric state of the HMf proteins**

Chemical crosslinking can be used to determine the quaternary structures of soluble protein complexes, and can often be used under conditions that are inappropriate for other methods, such as SEC or sedimentation analysis (Ji, 1983; Wong, 1991). Native HMf preparations, in 100 mM potassium phosphate (pH 6.9), were shown previously by low-resolution, size exclusion chromatography to consist of polypeptide dimers (Krzycki et al., 1990). Chemical crosslinking by dimethyl suberimidate (DMSI) or by formaldehyde (HCHO) was used to confirm this observation, and to examine the oligomeric state of the HMf proteins under a wide variety of solution conditions. DMSI is a bifunctional crosslinking reagent specific for free amine groups which, on the HMf proteins in the pH range 8-10, are primarily lysine residues. DMSI has a C₆ linker between the two functional moieties, allowing crosslinking of lysine residues up to 11 Å apart in three dimensional space. In contrast, HCHO is a non-specific crosslinking reagent, reactive with cysteine, tyrosine, histidine, tryptophan, and arginine residues, in addition to amines, and it will only crosslink reactive groups that are adjacent in space (Ji, 1983). Both DMSI and HCHO crosslinked HMfA, HMfB, and native HMf in 100 mM NaCl, 50 mM TEA-HCl (pH 8) to form products with the electrophoretic mobility expected for dimers, when analyzed by tricine-SDS-PAGE (Figure 10A; HCHO data not shown). HCHO was therefore used for all subsequent crosslinking studies, as it is more
Specificity and concentration dependence of HMf crosslinking. A. Specificity of DMSI crosslinking in 100 mM NaCl, 25 mM triethanolamine-HCl (pH 8). Proteins (1.5 μg) were incubated in the presence (+) or absence (−) of 1 mg/ml DMSI, for 1 h at 25 °C, prior to quenching. Native HMf (N), rHMfA (A), and rHMfB (B) containing reactions were compared to negative (cytochrome C, a monomer; Cyt C) and positive (E. coli single-stranded DNA binding protein, a tetramer; United States Biochemical, Cleveland, OH) controls by tricine-SDS-PAGE, followed by CBB staining. B. Concentration dependence of HCHO crosslinking in 150 mM KCl, 25 mM KPi, pH 7.5. rHMfA or rHMfB (1.5-5 μg) were crosslinked with HCHO (40 mM) for 1 h at 25 °C, at decreasing concentrations of 4.5, 0.45, 0.045, and 0.0045 mg/ml. Aliquots of the quenched reaction mixtures were resolved by tricine-SDS-PAGE, followed by CBB staining. Non-crosslinked rHMfA and rHMfB monomers (0), and molecular weight standards (S) of 6.2, 8.2, 10.6, 14.4, 17.0, 21.5, 31.0, 45.0, 66.2, and 97.4 kDa were used. The positions of the monomer, dimer, and tetramer forms of rHMfA and rHMfB are indicated.
Figure 10.
stable and reactive over a wider range of solution conditions than DMSI. That HCHO will crosslink as effectively as DMSI suggests that reactive residues on the protein monomers, which can only be lysine, arginine and histidine in (r)HMfA and (r)HMfB, must be positioned very close in space.

Crosslinking of HMf is specific and concentration dependent

The specificity of HMf crosslinking was examined in 100 mM NaCl, pH 8, using cytochrome C (a monomer of 12.4 kDa) and E. coli single-stranded DNA binding protein (a tetramer of 18.9 kDa monomers; Sancar et al., 1981; Chase and Williams, 1986) as negative and positive controls, respectively. HMf crosslinking was specific, and only dimers were formed for rHMfA, rHMfB, and native HMf proteins (Figure 10A). This was the case even with the addition of excess DMSI or HCHO, or prolonged incubation of reaction mixtures (data not shown). Crosslinking of rHMfA and rHMfB at different protein concentrations demonstrated that at very high concentration (4.5 mg/ml, and to a lesser extent, at 0.45 mg/ml), both proteins could be crosslinked to form products with mobilities expected for tetramers and possibly octamers, in addition to the major dimer product (Figure 10B). However, tetramers were not formed in the concentration range at which all other studies in this work were carried out (0.02-0.2 mg/ml).

rHMfA and rHMfB are dimers over a wide range of solution conditions

HCHO was used with a wide range of salt, pH and temperature conditions, to crosslink rHMfA and rHMfB preparations. For both proteins, crosslinking generated products with the electrophoretic mobility expected for dimers (MW ~15 kDa), under all solution conditions tested (Figure 11). The variations observed in the extent of crosslinking were consistent with the dependence of the crosslinking reaction on the
Figure 11.

Tricine-SDS PAGE of the products of HCHO crosslinking. Preparations of rHMfA and rHMfB (1.5 μg samples) were crosslinked: A. In the presence of 0, 50, 100, 500, 1000, 1500 mM KCl at pH 7.5. B In the presence of 0.1 M or 1 M KCl, at pH 7.5, at 25, 37, 65, 83°C. C. Timecourse of crosslinking of rHMfA and D. of rHMfB, at pH 6, 8, and 10, at 25°C, in 100 mM KCl for 20, 40, or 80 min. Symbols indicate: O, no crosslinker added (100 mM KCl, pH 7.5); S, protein size standards. The positions of the monomer and dimer forms are indicated.
Figure 11.

A. rHMfA  
B. rHMfA
C. rHMfB
D. rHMfB

kDa
66
45
31
22
17
14
11
8.2
6.2

Dimer
Monomer
proportion of lysine residues expected to be unprotonated at the reaction pH (Figure 11C, 11D), and with the reaction temperature (Figure 11B).

*Mt. fervidus* grows optimally at 83°C, and has an internal potassium concentration of ~1 M with the major counterion being cyclic 2,3-diphosphoglycerate (cDPG) (Hensel and König, 1988). K₃cDPG is not commercially available, and is difficult to purify in preparative amounts from *Mt. fervidus*. KCl was therefore substituted to generate conditions that approximated the *in vivo* situation [1 M K⁺, 83°C, (pH 7.5)]. Under these conditions (Figure 11B), and in fact at all KCl concentrations tested, both rHMfA and rHMfB were polypeptide dimers in solution (Figure 11A). Interestingly, the crosslinking efficiency of rHMfA, but not rHMfB, appeared to be enhanced by increasing salt concentration (Figure 11A, 11B).

The doublet bands apparent in some gel lanes in Figure 11 (e.g. Figure 11A) appear to be an artifact resulting from use of the tricine-SDS-PAGE system with rHMfA and rHMfB or from properties inherent in these protein preparations. They do not result from the crosslinking, as doublet bands also appear in lanes with samples that were not crosslinked (e.g. Figure 11A, lanes ‘0’). Oxidation of methionine residues in the HMf proteins might result in such an artifact. The eukaryal histones and many other proteins are known to undergo oxidation of methionine residues during preparation and storage, and it is probable that methionine-oxidized proteins can bind SDS aberrantly and migrate differently from unoxidized proteins during SDS-PAGE (Ivanchenko and Georgieva, 1988; Hames, 1990; Volkin *et al.*, 1995). Alternatively, any variability in the reagents used that creates a non-homogeneous population of proteins before, during, or after stacking during electrophoresis could result in differential migration of some percentage of the protein population.
Development of non-denaturing gel systems to resolve (r)HMfA and (r)HMfB

One goal of the biochemical analysis of the HMf proteins was to determine the homodimer and/or heterodimer composition of native HMf. Initially, a two dimensional PAGE system was envisaged, which involved PAGE in a first dimension in the absence of 2-mercaptoethanol, using native HMf crosslinked with a thiol cleavable crosslinker (dimethyl-3,3'-dithiobispropionimide), followed by PAGE in a second dimension in the presence of 2-mercaptoethanol to cleave and resolve monomer components of crosslinked dimers separated in the first dimension. This approach was however unsuccessful because of reactivity problems between native HMf and the crosslinking reagent. An alternative approach using non-denaturing PAGE was therefore developed.

HMfA and HMfB have very similar sizes but isoelectric points of 9.2 and 10.4, respectively, and therefore can only be resolved under near-native pH conditions on the basis of their net positive charges. Since no suitable non-denaturing PAGE system to resolve very basic proteins was described in the literature, two systems were developed ad hoc, that separate the HMf proteins on the basis of their net positive charges. The first was a continuous buffer system without a stacking gel that used a histidine-MES buffer (pH 6.1) described by McLellan (1982), and incorporated 10% glycerol in the resolving gel to improve band resolution and minimize band ‘edge-trailing’ (Figure 12A). This system was, however, sensitive to protein loading, and smearing of the leading band resulted if insufficient protein was loaded, or if the electrophoresis used too high a field strength (> 7 V/cm; Figure 12B. The second non-denaturing PAGE system used a biphasic, discontinuous buffer system and was developed from principles and equations derived by Williams and Reisfeld (1964), and by Richards et al. (1965). Potassium and histidine were selected as leading and trailing ions, respectively, and PIPES was selected as a weak acid for the reservoir buffer. The final design, as described in the Materials
Figure 12.

Comparisons of non-denaturing gel systems and resolution of native HMf preparations. A. HMf proteins (4.5 μg) separated by continuous, non-denaturing PAGE at pH 6.1. The lanes contained: 1, rHMfA; 2, rHMfB; 3, native HMf (isolated at timepoint 1 of Figure 7B); 4, a mixture of rHMfA and rHMfB; 5, identical mixture as in lane 4, but incubated at 95°C for 5 min before electrophoresis (see text). The bands formed by the HMfA homodimers (A•A), HMfB homodimers (B•B), and predicted to be HMfA•HMfB heterodimers (A•B) are identified. B. Kinetics of formation of the predicted (r)HMfA•(r)HMfB heterodimer. Continuous, non-denaturing PAGE showing dimers of rHMfA (lane 1, 4.5 μg), rHMfB (lane 2, 4.5 μg), and equimolar mixtures of rHMfA and rHMfB dimers (9 μg), incubated for 45 min at 25 °C, for 3 h at 25 °C, and heat treated as in A (lanes 3, 4, and 5, respectively). C. Discontinuous, non-denaturing PAGE, showing the improvement of band resolution for dimers of rHMfA (lane 1, 3 μg), rHMfB (lane 2, 3 μg), and a heat-treated, equimolar mixture of rHMfA and rHMfB (lane 3, 6 μg). B and C. Homodimers of rHMfA (A2), rHMfB (B2), and predicted heterodimers (A•B) are indicated.

Note that both gel systems, but particularly the continuous system in panel A, are sensitive to protein loading. This causes the leading bands (B2 homodimers) not to comigrate, as in lanes 4 and 5. In these non-denaturing PAGE systems, rHMfB also stains more intensely with CBB than does rHMfA (eg. panel A; lane 4). Protein migration was from + to – electrodes (top and bottom of the figures, respectively).
and Methods section, incorporated 10% glycerol in both the 4% T stacking gel and in the 16.5% T resolving gel. This discontinuous system gave much better, more reproducible resolution of the different HMf dimer populations than the continuous system (Figure 12C). The continuous system, however, was useful in that it had no detectable effect on the kinetics of rHMfA•rHMfB heterodimer formation (see below), in part due to the absence of a stacking gel.

**Analysis of native and recombinant HMf dimer populations**

**Resolution in one dimension and kinetics of heterodimer formation**

The continuous, non-denaturing gel system was used to resolve both native and recombinant HMf dimer populations. As shown in Figure 12A, this system clearly separated homodimers of rHMfA (lane 1) from homodimers of rHMfB (lane 2), and demonstrated that preparations of native HMf from *Mt. fervidus* (lane 3) contained both HMfA and HMfB homodimers, and a third population predicted to be HMfA•HMfB heterodimers. In support of this hypothesis, an equimolar mixture of rHMfA and rHMfB homodimers (lane 4) that was heated at 95 °C for 5 min, followed by slow cooling to room temperature, resulted in the appearance of a band that co-migrated with the band predicted to contain heterodimers (lane 5). The kinetics of the predicted heterodimer formation (Figure 12B) were shown to be slow at 25 °C, with the heterodimer band barely detectable after 45 min (lane 3), and requiring at least 3 h (lane 4) to achieve an intensity comparable to a mixed homodimer sample that had undergone rapid equilibration by incubation at 95 °C for 5 min (lane 5).
Native HMf composition by two dimensional PAGE

The composition of the dimer bands predicted to contain HMfA•HMfB or rHMfA•rHMfB heterodimers was established conclusively by two dimensional PAGE. Components of the dimers separated by non-denaturing (discontinuous) PAGE in the first dimension were resolved as denatured monomers in the second dimension by AUT-PAGE. Predicted heterodimer bands in both a native HMf preparation (Figure 13A), and in a heat treated equimolar mixture of rHMfA and rHMfB homodimers (Figure 13B) were indeed resolved into equal amounts of (r)HMfA and (r)HMfB monomers by the second dimension gel (Figure 13). Therefore, in vivo, native HMf is most likely composed of a mixture of (HMfA)₂ and (HMfB)₂ homodimers and (HMfA•HMfB) heterodimers, in rapid equilibrium, since the optimal growth temperature of *Mt. fervidus* (83 °C) is consistent with the rapid kinetics of heterodimer formation that were observed in vitro. Also, the net charges predicted for HMfA•HMfA, HMfA•HMfB, and HMfB•HMfB dimers, at pH 6.1, are +3.7, +5.7, +7.7, respectively, and as their molecular masses are very similar, an even distribution of electrophoretic mobilities was expected, but was not observed (for example, Figure 12A, lanes 3 and 5). The different monomer pairs must therefore have different conformations, or dimerization in one or more cases involves electrostatic interactions that influence electrophoretic mobility.

Analysis of N-terminal processing of rHMfA in *E. coli*

During scale-up of expression and purification of rHMfA from *E. coli* strain KS1124, protein preparations were found to be non-homogeneous and to consist of 3 populations (3 bands) (Figure 14A). These 3 populations were suspected to consist of formylmethionyl-rHMfA (fmA) homodimers, methionyl-rHMfA (mA) or rHMfA (A) homodimers, and the corresponding heterodimers, based on the fact that formylation
Figure 13.

Two dimensional PAGE of A. native HMf isolated at timepoint 1 of Figure 7B, and B. an equimolar mixture of heat-treated rHMfA and rHMfB homodimers. The dimers separated in the first dimension (1D), by discontinuous, non-denaturing PAGE were resolved in the second dimension (2D) by denaturing AUT-PAGE. The dimers are identified in the 1st dimension as in Figure 12. Lanes 1 and 2 contained the same samples as in lanes 1 and 2 of Figure 12B, and lane 3 contained either native HMf (9 µg, in panel A), or the same sample as in lane 3 of Figure 12B (in panel B). Denatured rHMfA (A) and rHMfB (B) monomers are indicated in the second dimension. The different staining intensities of rHMfA and rHMfB in B result from differences in binding CBB in this PAGE system. That equimolar quantities of rHMfA and rHMfB were used was determined by quantitative amino acid analysis in triplicate.
Figure 13.
reduces the net positive charge of a protein by 1. Mild acid hydrolysis using 0.6 N HCl is a standard procedure used to 'de-block' (deformylate) the N-termini of proteins to facilitate their Edman degradation (LeGendre et al., 1993), and was used to test this hypothesis. As shown in Figure 14A, hydrolysis resulted in the loss of the two slower migrating bands with a concomitant increase in the intensity of the fastest migrating band. The composition of the dimers that formed the intermediate band was determined using the two dimensional PAGE system described above, and were shown indeed to be heterodimers of formylated and non-formylated monomers (Figure 14B).

All proteins synthesized in *E. coli* are initiated with an N-terminal formyl-methionyl residue which is later processed. A deformylase first removes the formyl group, and a methionine amino peptidase (MAP) may, or may not then remove the resulting methionyl residue depending on the amino acid residue located at the second position of the polypeptide chain (Sherman et al., 1985). These two post-translational events must occur sequentially, as polypeptides with an N-terminal formyl-methionyl residue are not substrates for the MAP (Adams, 1968; Mazel et al., 1994).

Overexpression of heterologous genes in *E. coli* may overwhelm the endogenous deformylase and MAP activities in induced cells, and this seems to be the case for *hmfA* expression in strain KS1124. Because of the N-terminal processing sequence, the 3 dimer bands documented in Figure 14 must actually be composed of (fmA*fmA) homodimers, (fmA*mA plus fmA•A) heterodimers, and (mA•mA plus mA•A plus A•A) dimers. Expression of rHMfA at lower levels using strain KS1138 was attempted, however, this failed to resolve the heterogeneity problem (data not shown).

To solve the mixed-dimer problem, a copy of the *E. coli* MAP-encoding *map* gene was cloned into pKK223-3 (Brosius and Holy, 1984) containing *hmfA*, to create plasmid pKS395 and strain KS1183, that overexpresses both MAP and rHMfA (Sandman et al.,...
Figure 14.

Analysis of rHMfA expression in *E. coli*. A. Time course of the acid hydrolysis of a rHMfA preparation. A rHMfA preparation isolated from *E. coli* KS1124 cells was incubated in 0.6N HCl and aliquots, removed after the times (in h) indicated above each lane, were subjected to continuous, non-denaturing electrophoresis and then stained with CBB. B. Two dimensional PAGE separation of the polypeptides in a rHMfA preparation from strain KS1124, carried out in the same way as described in Figure 13. Lane 1, rHMfA from strain KS1124; Lane 2, rHMfA from strain KS1183 grown in the presence of TMP and Tdr. C. Discontinuous, non-denaturing PAGE separation of rHMfA samples prepared from *E. coli* strains grown in the absence or presence of TMP and Tdr. Lane 1, native HMf from *Mt. fervidus*; lanes 2 and 3, rHMfA preparations from strain KS1138 (contained only endogenous MAP) grown in the absence and presence of TMP and Tdr, respectively. Lanes 4 and 5, rHMfA preparations from strain KS1183 (over-expressing MAP) grown in the absence and presence of TMP and Tdr, respectively.

The abbreviations fmA, mA and A are used to indicate the f-met-rHMfA, met-rHMfA and rHMfA polypeptides, respectively (see text). A•A, B•B, and A•B indicate (r)HMfA homodimers, (r)HMfB homodimers, and (r)HMfA•(r)HMfB heterodimers, respectively. Bands indicated by δ and ε in panel C correspond to fmA•A and fmA•fmA dimers, respectively.
Figure 14.
1995). To circumvent the additional need for increased deformylase activity, it was found that growth of KS1183 in LB medium that contained TMP and Tdr could force *E. coli* to initiate protein synthesis with met-tRNA\textsubscript{i}, rather than the usual f-met-tRNA\textsubscript{i}, since TMP inhibits dihydrofolate reductase activity which is required for formylmethionine production (Tdr must be added to allow DNA synthesis under these growth conditions; Mazel *et al.*, 1994). Expression of MAP and rHMfA in KS1183, using these growth conditions, therefore led to correct N-terminal processing of the rHMfA (Figure 14C), and Edman degradation verified that the processed peptide has the same N-terminal amino acid sequence as that of native HMfA (Sandman *et al.*, 1995).

**DISCUSSION**

The DNA binding and wrapping properties of rHMfA and rHMfB had been found previously to be very different, as assayed by agarose EMSA and topoisomer formation, even though their amino acid sequences are very similar (Sandman *et al.*, 1994b). For rHMfA, saturation of DNA binding and wrapping occurred at much lower protein/DNA mass ratios than for rHMfB. The actual extent of binding and wrapping at saturation was, however, much greater for rHMfB, and much more compact structures were formed than for rHMfA (Sandman *et al.*, 1994b). In addition, binding and wrapping was cooperative for rHMfB, but not for rHMfA. The growth phase dependent regulation of HMfA and HMfB levels *in vivo* in *Mt. fervidus* cells documented here presumably takes advantage of these differences, and these observations can be combined into a model for growth phase dependent regulation of the *Mt. fervidus* chromosome structure (Sandman *et al.*, 1994b). In exponentially growing cells, some genome compaction must be needed to meet space constraints but the genome structure must still allow DNA replication and
rapid access to genes for transcription. HMfA homodimers predominate during exponential growth and their DNA binding properties are consistent with limited genome compaction. HMfA homodimers may also participate in replication and transcription by directing DNA bending, by binding specifically to 'bent DNA' regulatory sequences, and by forming NLS that could have both local and global effects on the superhelicity of the *Mt. fervidus* genome. As *Mt. fervidus* cells enter the stationary growth phase, they increase the synthesis of HMfB and this should result in much greater genome compaction. This increased 'archiving' of the *Mt. fervidus* genome into NLS by HMfB is presumably appropriate for the ensuing period of very limited replication and transcription. Growth phase dependent synthesis of DNA binding proteins has also been documented in *Bacteria*. For example, in *E. coli*, there is an increase in the nucleoid-associated DNA binding proteins H-NS and IHF as cells enter the stationary growth phase (Dersch *et al.*, 1993; Ditto *et al.*, 1994), and synthesis of the FIS protein increases following dilution of stationary phase cells into fresh growth medium but then decreases when growth begins (Ninnemann *et al.*, 1992).

The roles of (HMfA*HMfB) heterodimers are uncertain. Currently, homogeneous populations of these heterodimers cannot be obtained in the absence of 'contaminating' (HMfA)₂ and (HMfB)₂ homodimers. (HMfA*HMfB) heterodimers could have DNA binding and wrapping properties intermediate between those of the two homodimers. Heterodimer formation *in vivo* is likely to be in equilibrium with homodimer formation, since heterodimerization at the *in vivo* growth temperature of 83 °C is probably very rapid. If all possible combinations of homodimers and heterodimers can also associate on DNA *in vivo* to form tetramers (the likely DNA bound form of HMf; see Chapter 5), then the *Mt. fervidus* genome could contain six different types of NLS. Regulating the levels of synthesis of the HMfA and HMfB polypeptides, and
therefore the relative amounts of the different tetramer-containing NLS, could provide *Mt. fervidus* cells with a mechanism to 'fine tune' genome compaction and the regulatory functions of HMf.

Post-translational modifications of eukaryal histones play important roles in controlling their regulatory functions. The modified amino acids are however located in the N-terminal ‘tails’ of these proteins (Chapter 1; Wolffe, 1992) that are not present in the HMf proteins. If post-translational modification of HMfA or HMfB occurs *in vivo* in *Mt. fervidus*, this should alter their net positive charges and therefore result in altered electrophoretic mobilities in non-denaturing PAGE analyses. However, populations of HMfA or HMfB dimers with altered electrophoretic mobilities have not been observed during non-denaturing PAGE analysis of native HMf preparations.

Formaldehyde crosslinking demonstrated that rHMfA and rHMfB are polypeptide dimers in solution over very wide ranges of salt, pH, and temperature conditions, including the 1 M K+, 83°C, pH 7.5 condition that was used to approximate the *in vivo* environment (Figure 11). However, at very high protein concentrations, such as are probably found inside the cells of most microorganisms (Goodsell, 1991), both rHMfA and rHMfB could be crosslinked to form tetramers and possibly even octamers, although these were minor crosslinked products (Figure 10B). Therefore, even though the dimer is likely to be the most stable HMf oligomer, it may exist *in vivo* in a state that is *competent* for tetramerization or octamerization. The higher order structures observed at high protein concentration could, however, also be artifacts of non-specific crosslinking resulting from the close proximity of many reactive lysine residues, a problem that can be encountered when concentrated protein solutions are crosslinked (Wong, 1991).
CHAPTER III

CONFORMATIONAL STABILITY OF THE HMf PROTEINS

INTRODUCTION

The basis of protein chemical and thermal stability remains a central, unsolved problem in protein biochemistry (Menéndez-Arias and Argos, 1989; Adams, 1993), and one for which suitable model proteins for investigation are lacking. In principle, the HMf protein family provides an ideal set of such proteins that are small (67-69 residues), closely related in sequence, and have similar biological properties (DNA binding and wrapping), but that differ in their intrinsic thermal and chemical stabilities (Darcy et al., 1995; Grayling et al., 1995a). Because native HMf is thermostable, one of the major goals of this project was to define the thermostabilities of the HMfA and HMfB protein components, and their dependence on the ionic environment. With the E. coli expression problems for rHMfA solved (Chapter 2), larger amounts of rHMfA and rHMfB were produced and biophysical studies initiated to approach this goal.

CD spectroscopy is a unique light absorption method that is very sensitive to the secondary structural content (α-helix, β-sheet, random coil) of proteins in solution. As a solution-based technique, it is a simple way to monitor protein conformational changes that result from changes in salt, pH, temperature and quaternary structure (Chapter 1; Johnson, 1988). A series of CD spectroscopy-based experiments was therefore initiated to characterize the solution secondary structures and stabilities of rHMfA and rHMfB under a variety of pH, salt, and temperature conditions. Initial studies examined the salt-
dependent stability of rHMfA and rHMfB and the conformation of the rHMfA-rHMfB heterodimer as compared to the individual HMf homodimers. These studies showed that the solution conformational structures of both HMfA and HMfB were highly dependent on salt concentration, and CD-based thermal denaturation assays were then used to define the thermal unfolding of these proteins in different salt environments. Thermal unfolding assays were also carried out using rHfoB, a variant of the HMf proteins derived from the mesophile, *Mb. formicicum*, expressed in *E. coli*. This chapter documents the salt-dependent and thermal stabilities of rHMfA and rHMfB in comparison with rHfoB, as determined by CD spectroscopy. The data presented confirm previous suggestions (Darcy *et al.*, 1995) that although the amino acid sequences of members of the HMf family of proteins are very similar (Chapter 1), the differences in their sequences are sufficient to confer major differences in their thermostabilities. In addition, thermal unfolding assays revealed that these proteins undergo cold denaturation under only very mildly denaturing conditions, a rarely observed phenomenon that is nevertheless predicted for all proteins.

**MATERIALS AND METHODS**

**Reagents and proteins**

Chemicals were purchased from the Sigma Chemical Co. (St Louis, MO). Guanidine-HCl used was Sigma product # G-3272 (lot # 64H5771).

**Protein production**

Preparations of rHMfA and rHMfB were purified as described in Chapter 2, from *E. coli* strains KS1183 and KS1076, respectively (Sandman *et al.*, 1994b), and contained >98% rHMfA or rHMfB, as determined by silver-staining following SDS-PAGE, and by
reverse-phase HPLC. Amino acid microsequencing demonstrated that >95% of the recombinant protein in these preparations had the same N-terminal sequence as the respective native HMfA and HMfB proteins obtained from *M. fervidus*. The rHMfA and rHMfB polypeptides do not contain UV-absorbing residues suitable for use in protein quantitation. Protein concentrations were therefore determined in triplicate by quantitative amino acid analysis, using a Waters Picotag amino acid analysis system (Millipore Corporation, Milford, MA). The recombinant proteins (10-20 mg/ml) were stored at -20°C in 1 M KCl (or NaCl), 25 mM Tris-HCl (pH 7.5), and diluted as appropriate for each experiment. rHFO protein was obtained from K. Sandman, and was quantitated by the method of Bradford (1976). All protein concentrations stated refer to the concentration of rHMfA, rHMfB or rHFO monomers.

**Production of rHMfA•rHMfB heterodimer containing mixtures**

Mixtures containing (rHMfA•rHMfB) heterodimers were prepared by incubating homogeneous preparations of rHMfA and rHMfB in 100 mM KCl, 10 mM Tris-HCl (pH 7.5), in equimolar amounts at 95°C, followed by slow cooling to room temperature, as previously described (Chapter 2; Sandman *et al.*, 1994b). These heat-treated mixtures were used for both electrophoresis and CD studies.

**Electrophoresis techniques**

The single phase, continuous buffer system described in Chapter 2 was used to resolve the dimer components of the heat-treated mixtures of rHMfA and rHMfB that were analyzed by CD spectroscopy. This non-denaturing PAGE system had no detectable effect on the kinetics of rHMfA•rHMfB heterodimer formation, in part due to the absence of a stacking gel. For the same reason, however, the system was sensitive to protein loading, and leading bands tended to smear.
**CD spectroscopy**

**Instrumentation**

CD spectra were measured using a Spex Dichrograph CD6 spectropolarimeter and associated software (Jobin-Yvon, France). At least 1 h of lamp warm-up time was allowed before making measurements, for maximal baseline stability. For all measurements, instrument baseline drift was < 0.1 mdeg per day and baseline deviation from 0 mdeg in non-absorbing spectral regions was < ± 0.5 mdeg. Thermal denaturation measurements made using the same machine used a manually programmed Hewlett Packard 89090A temperature controller and a thermoelectrically heated/cooled cell holder, modified to fit the spectropolarimeter sample chamber. The thermoelectric cell holder was attached to a modified, water-cooled heatsink, to maximize heating and cooling efficiency. The external water circulator contained a 1:1 ethylene glycol-water mixture, and was maintained at 5-40 °C. The temperature regulation accuracy for samples in the cell holder was ± 0.1 °C.

**Standard data acquisition conditions**

All measurements were made by diluting stock protein solutions in buffers that had been passed through 0.2 μM filters. Final salt concentrations were corrected for addition of salt from the stock protein solution. For acquisition of full spectra (200-300 nm), data were collected at 0.5 nm intervals, using a fixed slit width of 2 nm and a 0.5 sec integration time, and 4-16 scans were accumulated for each spectrum. For thermal denaturation analysis, the wavelength was fixed at 222 nm or 230 nm (2 nm bandpass), and 12 data points were manually recorded, using an integration time of 5 sec.
Heat treatment and salt titration experiments

For the heat treatment experiment, full spectra of protein samples (35 μM) in 100 mM KCl, 10 mM Tris-HCl, pH 7.5, were measured in a cylindrical cell with 0.1 mm pathlength, at 25 °C. For the salt titrations, protein stock solutions were diluted directly into filtered 0.01 N HCl (pH 2) or 10 mM Tris-HCl (pH 7.5) to give a final concentration of 5 μM. For each titration, 2.3 ml of the appropriate protein solution was dispensed into a 1 cm pathlength cuvette, and 3.5 M KCl was serially added to give the desired final KCl concentration. The contents of the cell were mixed and incubated at ambient temperature (25 °C) for 10 min after each salt addition, to allow equilibrium to be attained before measurement. The pH variation over the salt range of the titrations was < 0.1 unit, and overall errors for these data varied from < ± 1% for KCl concentrations ≤ 0.2 M, to < ± 0.3% for KCl concentrations ≥ 0.2 M. Errors for the unsmoothed data points in the region 218 nm to 230 nm were < ± 3%.

Thermal denaturation assays

Protein stock solutions were diluted directly into the appropriate, filtered buffers, to give final concentrations of 5 μM. The buffers used were 0.2 M and 1 M KCl in 10 mM KPi, pH 7.5 containing 1 M gdm-cl; 1 M KCl, K₂SO₄, or KPi in 25 mM KPi, pH 7.5, with or without 1.5 M gdm-cl; 1.025 M KPi, pH 7.5, with or without 2.75 M gdm-cl. For each of the 1 M salt-containing buffers, [K⁺] was fixed at 1.025 M, and anion concentrations varied. Changes in the pH of solutions over the temperature range 0-80 °C were < 0.3 units. Each protein solution (2.3 ml) was dispensed into a 1 cm pathlength cuvette, and equilibrated at 25 °C for 10 min. Baseline spectra were then acquired (4-6 scans), followed by re-equilibration of samples at 0 °C for 30 min. Thermal denaturation assays were then carried out by acquiring data at 222 nm or 230 nm (for samples in 2.75 M gdm-cl), at 2 or 4 °C intervals, up to a maximum of 76-84 °C. Sample equilibration
was followed using the intensity/time function of the spectropolarimeter software. For temperatures below 20 °C, the sample chamber was purged with dry N\textsubscript{2}, to prevent condensation on the cuvette faces, and samples were allowed 10-20 min to equilibrate, after each temperature increment. At higher temperatures, equilibration was complete within 2-5 min. Full spectra were acquired at the temperature maximum of the assay, and 20 min after cooling to 25 °C (4-6 scans), and were compared to the baseline spectra acquired at the beginning of the assay. Errors in the measured CD were estimated as ± 1σ of the average of the 12 data points collected at each temperature increment, and were ≤ ±3% for samples containing up to 1.5 M gdm-cl. Errors for samples containing 2.75 M gdm-cl were larger and more variable.

**Refolding kinetics**

Samples at their maximal denaturation temperatures, that were completely unfolded, were rapidly cooled (> 8 °C/min) to 25 °C, and changes in their CD at 222 or 230 nm over time were monitored using the Kinetics function of the Dichrograph software. Data points were sampled at 6 sec intervals, using a 5 sec integration time (2 nm bandwidth). Errors in measured CD were ≤ ±3% for all samples.

**Data analysis**

All CD data were expressed as either the raw ellipticity values (in mdeg), the pathlength and concentration-independent form (molar ellipticity, [θ], in deg cm\textsuperscript{2} dmol\textsuperscript{-1}), or were converted to fractional unfolding values. Spectral data were averaged from the results of 4-16 scans. Data from the heat treatment experiment were smoothed with a 5-point sliding Fourier filter (*Passage*; written by D. Guenther and C. Tigges, 1987). Where other data smoothing was required, it was accomplished using a 9-point sliding quadratic-cubic filter. A software routine (in C-source code) was written specifically for
this purpose (Figure 15), and was based on the algorithm of Savitzky and Golay (1964). Assymptotes to sigmoidal curves for the calculation of ion uptake at pH 2, or to protein folding curves for the calculation of fractional unfolding, were determined by nonlinear regression using the curve fitting function of Kaleidagraph (Abelbeck Software). Melting temperatures were determined using an average method, as \( T_m = T_{[\theta]_m} \), with the corresponding molar ellipticity values at transition midpoints calculated from curve assymptotes as \( [\theta]_m = \left( \frac{[\theta]_f - [\theta]_u}{2} + [\theta]_u \right) \), where \( [\theta]_f \) and \( [\theta]_u \) represent the values of \( [\theta] \) for the folded and unfolded protein, respectively (lower and upper assymptotes, respectively). Kinetic data were modelled with a four-parameter logistic function of the form \( f(x) = \frac{a - d}{1 + \left( \frac{x}{c} \right)^b} + d \), to generate sigmoidal fits, using the same software.

RESULTS

**Secondary structures of rHMfA and rHMfB and analysis of a rHMfA•rHMfB heterodimer containing mixture**

The CD spectra of rHMfA and rHMfB homodimers in 100 mM KCl at pH 7.5 are shown in Figure 16A. They are very similar, but are not identical, and no significant changes in either spectrum were observed using protein concentrations ranging from 1 \( \mu \text{M} \) to 150 \( \mu \text{M} \) (Figure 17A). On the basis of the dichroism at 222 nm, \((\text{rHMfA})_2\) molecules have slightly more helical character than \((\text{rHMfB})_2\) molecules (Table 2).

At pH 6.1, \((\text{rHMfA})_2\), \((\text{rHMfA})•\text{rHMfB}\), and \((\text{rHMfB})_2\) dimers are predicted to have net charges of +3.7, +5.7, and +7.7, respectively (Devereux *et al.*, 1984). When resolved using the continuous, non-denaturing PAGE system (Chapter 2), these dimers should therefore give rise to an even distribution of electrophoretic mobilities, since their
Figure 15.

C-source code for the 9-point quadratic-cubic smoothing routine described in the text. The code was compiled using Think C 5.0 for the Macintosh. Input and output files consisted of a single spreadsheet column of text containing up to 220 y-values that corresponded to evenly-spaced x-values. The algorithm used was based on that suggested by Savitzsky and Golay (1964).
include <stdio.h>
define MAXSTRING 32
define window_size 9
define norm_factor 231

void smooth(FILE *in, FILE *out);

main()
{
  FILE *ifp, *ofp;
  char c, in_file_name[MAXSTRING], out_file_name[MAXSTRING], query;
  int i;

  /* read input string */
  printf("Input file name ? *");
  for(i = 0; (c = getchar()) != '\n'; ++i)
    in_file_name[i] = c;
  in_file_name[i] = '\0';

  /* check for existing input file */
  if((ifp = fopen(in_file_name, 'r')) == NULL) {
    printf("File: %s, does not exist*");
    exit(1);
  }

  /* read output string */
  printf("Output file name ? *");
  for(i = 0; (c = getchar()) != '\n'; ++i)
    out_file_name[i] = c;
  out_file_name[i] = '\0';

  /* check for existing output file */
  if((ofp = fopen(out_file_name, 'w')) == NULL) {
    printf("Creating file: %s");
    ofp = fopen(out_file_name, "w");
  }
  else {
    printf("Overwriting : %s");
    scanf("%c", &query);
    if(query == 'n' || query == 'N') {
      close(ifp);
      close(ofp);
      printf("QUITING !");
      exit(1);
    }
  }

/* temporary segment */
  printf("Overwriting : %s");
  ofp = fopen(out_file_name, "w");
}
Figure 15 (continued)

```c
smooth(ifp, ofp); 
fclose(ifp); 
fclose(ofp); 
printf("\n\n%s", "Program completed successfully");
}

/* SMOOTHING SUBROUTINE */

void smooth(FILE *ifp, FILE *ofp)
{
  double window[window_size], in_data[220], out_data[220],
  window_sum;
  int coefficient[window_size] = [-21, 14, 39, 54, 59, 54, 39, 14,
  -21];
  int count = 0, data_size, smooth_data_size, now, prev;
  int i, j, k, array_top = (window_size - 1);

  /* READ IN DATA: HAS EXTRA LINE TO PRINT DATA AS READ IN */
  printf("Enter file size (# of data points): ");
  scanf("%d", &data_size);
  while(fscanf(ifp, "%lf", &in_data[count]) != EOF) {
    printf("%6d%3s%18.8e", (count+1), " : ", in_data[count]);
    ++count;
  }
  if(data_size != count) {
    printf("%3d%3d%3s", "Input file size (", data_size,
    ") does not match actual file size (", count, ")!");
    data_size = count;
  }

  /* INITIALIZE SMOOTHING VARIABLES */
  smooth_data_size = (data_size - window_size) + 1;
  for(i = 1; i <= array_top; ++i) {
    j = i - 1;
    window[i] = in_data[j];
  }

  /* MAIN SMOOTHING LOOP */
  for(i = 0; i <= (smooth_data_size - 1); ++i) {
    j = i + array_top;
    for(now = 0; now <= (array_top - 1); ++now) {
      prev = now + 1;
      window[now] = window[prev];
    }
    window[array_top] = in_data[j];
    window_sum = 0.0;
    for(k = 0; k <= 8; ++k)
      window_sum = window_sum + (window[k] * coefficient[k]);
    out_data[i] = (window_sum / norm_factor);
  }

  /* WRITE DATA TO FILE */
  for(count=0; count <= (smooth_data_size - 1); ++count) {
    fprintf(ofp, "%e", out_data[count]);
    putc('n', ofp);
    printf("%5.8e%3d%3s", out_data[count], " : point ",
      (count+1), " written to file");
  }

  /* END OF SMOOTH ROUTINE */
}
```
Figure 16.

CD spectra of mixtures of (rHMfA)$_2$ and (rHMfB)$_2$ dimers and non-denaturing PAGE analysis of rHMfA•rHMfB heterodimer formation. A. CD Spectra of protein samples (35 μM) in 100 mM KCl, 10 mM Tris-HCl, pH 7.5, were measured in a cylindrical cell with 0.1 mm pathlength. The spectra are (rHMfA)$_2$ (---), (rHMfB)$_2$ (---), an equimolar mixture of (rHMfA)$_2$ and (rHMfB)$_2$ (——), and an identical, but heat-treated, equimolar mixture of (rHMfA)$_2$ and (rHMfB)$_2$ (•••). The spectra shown were generated from raw data that were smoothed using a 5-point sliding Fourier filter. B. Continuous, non-denaturing PAGE showing, lane 1, rHMfA homodimers (A$_2$, 4.5 μg); lane 2, rHMfB homodimers (B$_2$, 4.5 μg); lanes 3, 4, and 5, equimolar mixtures of rHMfA and rHMfB homodimers (9 μg), incubated for 45 min. at 25 °C, for 3 h at 25°C, and heat treated at 95°C (see text). The band containing rHMfA•rHMfB heterodimers is indicated (A•B).
Figure 16.
Table 2. Physical Properties of rHMfA and rHMfB

<table>
<thead>
<tr>
<th>Protein</th>
<th>[θ]$_{222}$ pH 2</th>
<th>[θ]$_{222}$ pH 7.5</th>
<th>% helix, pH 2</th>
<th>% helix, pH 7.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>rHMfA$^d$</td>
<td>-7,900</td>
<td>-20,400</td>
<td>12 (31)</td>
<td>62 (70)</td>
</tr>
<tr>
<td>rHMfA$^e$</td>
<td>-26,100</td>
<td>-23,100</td>
<td>84 (88)</td>
<td>72 (78)</td>
</tr>
<tr>
<td>rHMfB$^d$</td>
<td>-5,800</td>
<td>-18,700</td>
<td>3 (24)</td>
<td>55 (65)</td>
</tr>
<tr>
<td>rHMfB$^e$</td>
<td>-20,800</td>
<td>-21,300</td>
<td>63 (71)</td>
<td>65 (73)</td>
</tr>
</tbody>
</table>

$^a$ The secondary structure of the rHMfB polypeptide is known from NMR data to consist of only α-helical regions and small segments of β-strand structure (Starich et al., 1995). The values of molar ellipticity used to calculate helical content were therefore only taken at 222 nm, and are indicated in units of deg cm$^2$ dmol$^{-1}$.

$^b$ Circular dichroism values used were taken from the same data sets used to generate Figures 17C and 17D.

$^c$ Calculated on the basis of 100% α-helix = -30,000 (Chen et al., 1974) and 100% coil = -5,000 deg cm$^2$ dmol$^{-1}$ (Becktel and Schellman, 1987). The values listed in brackets indicate the % helix calculated using +2,000 deg cm$^2$ dmol$^{-1}$ as an alternative 100% coil value (Becktel and Schellman, 1987).

$^d$ Proteins examined at 5 mM KCl.

$^e$ Proteins examined at 1 M KCl.
molecular weights are very similar. This was not observed (Figure 16B, lane 5), and the prediction was made that the different monomer pairs might have different conformations, or that dimerization in one or more cases could result in electrostatic interactions that influence electrophoretic mobility. Therefore, rHMfA•rHMfB heterodimer-containing dimer mixtures (Chapter 2) were examined to determine if there were any obvious CD-observable conformational differences. Both heat treated and equivalent, non-heat treated mixtures of (rHMfA)₂ and (rHMfB)₂ were examined by CD spectroscopy (Figure 16A). Identical CD spectra were obtained for both mixtures, and these lie between the spectra obtained for the homodimers alone. The continuous, non-denaturing PAGE system was therefore used to confirm that heterodimer formation had occurred in the heat treated mixture, and that there was negligible heterodimer formation in the non-heated mixture during the time needed to acquire its CD spectrum (Figure 16B).

**Salt titration of rHMfA and rHMfB**

The acidic and basic amino acid residues of proteins become completely protonated below ~pH 3 (Voet and Voet, 1990), leading to electrostatic repulsion and often unfolding of protein secondary and tertiary structures. Titration of an unfolded protein at pH 2 with salt can stabilize and/or destabilize the protein structure, and therefore can provide information regarding the electrostatic contributions to protein stability, and regarding ion binding by proteins during folding (Becktel and Schellman, 1987).

The stabilities of the rHMfA and rHMfB secondary structures were determined by CD spectroscopy, as a function of KCl concentration, at both pH 2 and pH 7.5. A subset of the CD spectra obtained for rHMfA at pH 2, are shown in Figure 17B, and very similar spectra were obtained for rHMfB (data not shown). At pH 2, rHMfA and rHMfB
**Figure 17.**

Dependence of CD spectra on rHMfA and rHMfB concentration and salt titrations of rHMfA and rHMfB at pH 2, and pH 7.5. A. Concentration dependence of rHMfA and rHMfB spectra. CD spectra of proteins, at concentrations of 1 μM (—), 5 μM (•••), and 150 μM (---) were measured in 100 mM KCl, 10 mM Tris-HCl, pH 7.5, in cells with pathlengths of 1 cm (for 1 and 5 μM protein) or 0.1 mm (for 150 μM protein). The upper and lower groups of curves correspond to rHMfB and rHMfA, respectively. Raw data were smoothed using a 9-point sliding quadratic-cubic filter to generate the curves shown. B. A subset of the CD spectra for rHMfA at pH 2, measured in 0.006 M KCl (diamonds), 0.086 M KCl (squares), 0.126 M KCl (circles), 0.166 M KCl (triangles), 0.406 M KCl (square with + inscribed). C. and D. Circular dichroism measured at 222 nm as a function of KCl concentration, at pH 2 and pH 7.5, respectively. The data were collected using protein concentrations of 5 μM, in a 1 cm pathlength rectangular cell for rHMfA (diamonds), and for rHMfB (squares).
exhibited large changes in dichroism at 222 nm, showing cooperative transitions with increasing KCl concentration. Complete denaturation was achieved at pH 2, but only at KCl concentrations <0.1 M (Figure 17C).

At pH 7.5, the changes in the 222 nm dichroism of rHMfA and rHMfB proteins with increasing KCl concentration were greatly reduced, but the transitions observed were more cooperative and occurred at lower salt concentrations than those observed at pH 2 (Figure 17D). Although the helicity of both proteins increased with increasing KCl concentration at pH 7.5, at salt-saturation (>0.6 M KCl) the dichroism of rHMfA at 222 nm reached only -23 100 deg cm$^2$ dmol$^{-1}$, a value 13% less than the dichroism of rHMfA measured at saturating salt conditions at pH 2 (-26 100 deg cm$^2$ dmol$^{-1}$).

Measured dichroism values and estimated helicities for the salt titrations are summarized in Table 2. At pH 2, rHMfA showed a 230% increase in dichroism at 222 nm over the salt range from 5 mM to 1.3 M KCl, but only a 13% increase was observed at pH 7.5. For rHMfB, the increases were 260% and 14%, respectively.

**Cooperativity of ion uptake by rHMfA and rHMfB at pH 2**

The effect of KCl concentration on the extent of folding of rHMfA and rHMfB can be estimated from the slope of a Hill-type plot of $\ln(K_{app}^f)$ versus $\ln(c)$, where $c$ is the KCl concentration, and $K_{app}^f$ is the apparent equilibrium constant for renaturation. The value of the slope obtained is a measure of the cooperativity of ion uptake during the folding reaction (van Holde, 1985). Plots of the fractional unfolding of rHMfA and rHMfB versus the KCl concentration are sigmoidal, suggesting a single transition for both proteins at pH 2 (Figure 18A). Therefore, if a unimolecular, two-state reaction is assumed for the proteins at pH 2, the renaturation reaction can be written as $U \xrightarrow{K_{app}^f} F$, where $U$ and $F$ are the concentrations of unfolded monomers and folded dimers,
Figure 18.

Ion uptake calculations for the folding reactions of rHMfA and rHMfB at pH 2. 
A. Fraction of protein folding ($f_o$) versus KCl concentration. B. Least squares fit of the natural logarithm of the apparent equilibrium constants for conversion between the low and high salt (unfolded and folded) forms of the proteins versus the natural logarithm of KCl concentration. Data in both plots correspond to rHMfA (diamonds) and rHMfB (squares). The slopes calculated were 2.7 and 3.1 for rHMfA and rHMfB, respectively.
Figure 18.
respectively. The apparent equilibrium constants (at each salt concentration) for the refolding process can therefore be calculated as

$$K'_{\text{app}} = \frac{f_t}{1 - f_t}$$

(1)

and

$$f_t = \frac{[\theta] - [\theta]_u}{[\theta]_t - [\theta]_u}$$

(2)

where $[\theta]_u$ is the dichroism of the unfolded (ion unbound) form, $[\theta]_t$ is the dichroism of the folded (binding site saturated) form, and $f_t$ is the fraction of the folded form. A plot of the $\ln(K'_{\text{app}})$ vs $\ln(c)$ yields linear fits with slopes of 2.7 and 3.1 for rHMfA and rHMfB, respectively (Figure 18B), indicating that these are highly cooperative transitions.

**Thermal stability of rHMfA and rHMfB in KCl**

Thermal denaturation curves for rHMfA and rHMfB in 0.2 M and 1 M KCl, 10 mM KPi, pH 7.5, are shown in Figure 19. These KCl concentrations were chosen, as they correspond to the lower boundary of maximal stability and approximate the *in vivo* potassium concentration, respectively (Figure 17D). Transition points for thermal denaturation, which occur at the protein melting temperatures ($T_m$ values), were only attainable in the temperature range of the thermoelectric cell holder by the addition of gdm-cl to samples, to a final concentration of 1 M. $T_m$ values were 56, 65, 71, and 79 °C, for rHMfA and rHMfB in 0.2 M and 1 M KCl, respectively, indicating that the 0.8 M increase in KCl concentration resulted in a ~14-15 °C increase in melting temperature.
Figure 19.

Dependence of the thermal stability of HMf proteins on salt concentration.

Thermal denaturation assays were carried out with both rHMfA (rA) and rHMfB (rB), using a 1 M gdm-cl, 10 mM KPi, pH 7.5 buffer that contained 0.2 M or 1.0 M KCl. CD data, measured at 222 nm, were collected at 5 °C intervals. Errors in [θ] were ≤ 2.5% and are not shown for clarity.
Figure 19.
Anion dependence of rHMfA and rHMfB salt-based thermostabilization

*Mt. fervidus* cells, grown at the temperature optimum of 83 °C, contain ~1 M K+ and ~300 mM cDPG3° (Hensel and König, 1988). K₃cDPG is not commercially available, and is difficult to purify in sufficient amounts at a purity adequate for CD spectroscopy. Thermal denaturation curves for rHMfA and rHMfB were therefore determined in KCl, K₂SO₄, and KPi, using a basal buffer containing 1.5 M gdm-cl, 25 mM KPi, pH 7.5. Potassium concentrations were fixed at 1 M K⁺, and anion concentrations varied accordingly. The results, shown in Figure 20, indicate that for both proteins, thermostabilization is greatest in the presence of phosphate, with almost no unfolding observed, and decreases in the order Pi > SO₄²⁻ > Cl⁻. These results suggest that phosphate has a unique interaction with these proteins, and are in agreement with previous studies that have documented the anion and cation-dependent thermostabilization of the activities of enzymes isolated from thermophilic methanogens. For example, the activities of glyceraldehyde-3-phosphate dehydrogenase and malate dehydrogenase from *Mt. fervidus*, and N-formylmethanofuran: tetrahydromethanopterin formyltransferase from *Mp. kandleri* (a related methanogen a t<sub>opt</sub> of 98 °C) are thermostabilized most effectively by potassium phosphate, and least effectively by potassium chloride and sodium salts (Hensel and König, 1988; Breitung et al., 1992; Hensel and Jakob, 1994).

Although rHMfA and rHMfB could not be unfolded in 1 M KPi using 1.5 M gdm-cl, they could be unfolded within the temperature range of the instrument in 1 M KPi by the addition of 2.75 M gdm-cl to samples (Figure 21). At this concentration, gdm-cl absorbs strongly at 222 nm, and therefore CD measurements were made at 230 nm. Even at this higher wavelength, there is so much salt present that the signal to noise ratio for samples is dramatically decreased (Figure 21), although T<sub>m</sub> values can still be calculated (Table 3).
Figure 20.

Anion-dependence of rHMfA and rHMfB thermostability. Thermal denaturation assays were done for rHMfA (A) and rHMfB (B), using a 1.5 M gdm-cl, 25 mM KPi, pH 7.5 buffer that contained KCl (squares), K$_2$SO$_4$ (diamonds), or KPi (circles). Each salt addition to the buffer was made so that the final potassium concentration in the sample solution was 1.025 M. CD data (measured at 222 nm) were collected at 2 °C intervals for the KCl curves and at 4 °C intervals for the remainder of the curves. Errors in [θ] were ≤ 3% (not shown).
Figure 21.

Thermal denaturation of HMf proteins in buffer containing 2.75 M gdm-cl. CD Data recorded at 230 nm for rHMfA (diamonds) and rHMfB (squares) were collected at 4 °C intervals using a sample buffer that contained 2.75 M gdm-cl, 1.025 M KPi (K⁺ fixed at 1.025 M), pH 7.5. Errors in [θ], expressed as ± 1σ of the average of 12 data points collected per temperature increment, are plotted as vertical bars above and below data points.
Figure 21.
Melting temperatures determined for each of the unfolding curves (Figures 19, 20, 21; Table 3) indicate that rHMfB was ≥9 °C more thermostable than rHMfA, for each salt condition tested. However, with the buffer conditions used, maximal [θ]222 values were only 70-85% of those measured in the absence of gdm-cl (Figure 17). In 1 M KCl, containing 1.5 M gdm-cl, both proteins showed partial denaturation at temperatures between 0 and 15 °C. This phenomenon, termed cold denaturation, has been observed for only a limited number of proteins, although it is thermodynamically predicted for all proteins (Dill et al., 1989; Privalov, 1990; Antonino et al., 1991; Azuaga et al., 1992).

That the variations in [θ] measured did not include a component due to changes in guanidine absorbance with temperature was verified by recording spectra of the 1.5 M gdm-containing buffer at 0, 25, and 70 °C (Figure 22A; 0 °C spectrum not shown). For the wavelengths of interest (220-230 nm), there was no change in absorbance of the buffer over the temperature range 0-70 °C, except that there was less thermal noise in the spectra recorded at lower temperatures (Figure 22A).

**van't Hoff analysis of rHMfA and rHMfB thermal denaturations**

A simple thermodynamic analysis can be carried out for complete denaturation curves where proteins undergo a two-state, cooperative, reversible transition from a folded state (f) to an unfolded state (u), \( F \rightleftharpoons U \), where F and U are the concentrations of f (folded dimer) and u (unfolded monomer) forms, respectively. This is the case for thermal denaturation of rHMfA and rHMfB in 1 M KCl, 1.5 M gdm-cl, 25 mM KPi, pH 7.5 (Figures 20, 23). The apparent equilibrium constants for the unfolding reaction under these conditions can thus be calculated as

\[
K_{app}^u = \frac{f_u}{1 - f_u}
\]  

and
Figure 22.

Dependence of buffer on temperature and Van't Hoff analysis of rHMfA and rHMfB thermal unfolding. A. CD spectra were recorded for a buffer containing 1.5 M gdm-cl, 1 M KCl, 25 mM KPi, pH 7.5, at 0, 25, and 70 °C (0 °C spectrum not shown). Significant absorbance due to gdm-cl that precludes data collection is apparent at wavelengths below 218 nm. Each spectrum shown is the average of 5 scans. B. Van’t Hoff plots for rHMfA (diamonds) and rHMfB (squares) based on thermal denaturation data recorded in 1.5 M gdm-cl, 1 M KCl, 25 mM KPi, pH 7.5 (Figure 20). Only data points in the sigmoidal region of the curves around each thermal transition (40-64 °C) were used. The slopes calculated from least-squares fits of \( \ln(K'_{app}) \) [\( \ln(K_u) \)] versus 1/T were -13250 and -9100, which correspond to \( \Delta H''_{app} \) values of +110 and +76 kJ/mol (dimer) for rHMfA and rHMfB, respectively.
Figure 22.
by analogy with the previous derivation (equations 1 and 2). Applying standard thermodynamic theory (van Holde, 1985),

\[ \Delta G_{\text{app}}^u = \Delta H_{\text{app}}^u - T \Delta S_{\text{app}}^u \]  

(5)

where \( x_{\text{app}}^u \) are the apparent quantities for the unfolding reaction, and

\[ \Delta G_{\text{app}}^u = -RT \ln(K_{\text{app}}^u) \]  

(6)

Combining equations 5 and 6 gives

\[ \ln(K_{\text{app}}^u) = \frac{-\Delta H_{\text{app}}^u}{RT} + \frac{\Delta S_{\text{app}}^u}{R} \]  

(7)

Therefore, a plot of \( \ln(K_{\text{app}}^u) \) versus \( 1/T \), known as a Van’t Hoff plot, should give a straight line with slope \( -\Delta H_{\text{app}}^u / R \) (van Holde, 1985). Although this Van’t Hoff analysis makes several assumptions, in particular that the unfolding reaction is two-state and unimolecular, and that the enthalpy change for unfolding is independent of temperature, it nonetheless provides a useful estimate of \( \Delta H_{\text{app}}^u \) for the unfolding reaction. Applying this analysis to rHMfA and rHMfB yields linear fits (Figure 22B) and \( \Delta H_{\text{app}}^u \) values of +110 and +76 kJ/mol (dimer), respectively, which are within the range of values typically observed for heat denaturation of globular proteins [0.4-2.0 kJ/mol (amino acid residue); Murphy, 1995].

**Reversibility of rHMfA and rHMfB thermal denaturation and rates of refolding**

Spectra were measured in 1.5 M gdm-cl, 1 M KCl, 25 mM KPi, pH 7.5 and in 2.75 M gdm-cl, 1.025 M KPi, pH 7.5 before denaturation (at 25 °C), after denaturation (at 76-84 °C), and after cooling (at 25 °C). Thermal unfolding for rHMfA and rHMfB was completely reversible in both buffers (Figure 23), indicating that no CD-detectable
Figure 23.

Reversibility of rHMfA and rHMfB thermal denaturation. CD spectra were measured for rHMfA (A, C) and rHMfB (B, D) in either 1.5 M gdm-cl, 1 M KCl, 25 mM KPi, pH 7.5 (A and B) or 2.75 M gdm-cl, 1.025 M KPi, pH 7.5 (C and D). Spectra were measured at 25 °C, prior to denaturation (—), after 3.5-4.5 h when the maximal denaturation temperature of 76-84 °C was achieved (---), and 20 min after cooling to 25 °C (•••). Each unsmoothed spectrum shown is the average of 4-6 scans.
Figure 23.

A. $[\theta] \times 10^{-4}$ deg cm$^2$ dmol$^{-1}$

B. $[\theta] \times 10^{-4}$ deg cm$^2$ dmol$^{-1}$

C. $[\theta] \times 10^{-4}$ deg cm$^2$ dmol$^{-1}$

D. $[\theta] \times 10^{-4}$ deg cm$^2$ dmol$^{-1}$
changes resulted from the heat denaturation that prevented subsequent refolding into the original secondary structural conformation.

The rates of refolding of the HMf proteins in these buffers were found to be rapid (Figure 24) which was not unexpected, given the small size and simple tertiary structures of these proteins (Starich et al., 1995). Calculated $t_{1/2}$ values for refolding were 93, 163, and 146 sec, for rHMfB, and rHMfA and rHMfB, in the 1.5 M gdm-cl and 2.75 M gdm-cl buffers, respectively. These values could, however, be overestimates if refolding proceeded as rapidly as the sample cooling would allow. The slower refolding rates at the higher gdm-cl concentrations were expected, as gdm-cl and other protein denaturants preferentially interact with the denatured state, thus stabilizing it (Privalov, 1990).

**Thermal stability of rHFoB**

Darcy et al. (1995) purified three members of the HMf protein family, HFoA1, HFoA2, and HFoB, from the mesophilic methanogen *Mb. formicicum*, and cloned and sequenced the encoding genes. Native HFo preparations were found to contain a mixture of all three polypeptides, and the DNA binding activity of these preparations was shown to be significantly less thermostable than that of native HMf or HMt preparations isolated from *Mt. fervidus* or *Mb. thermoautotrophicum*, respectively (Darcy et al., 1995). The HFoB encoding gene, *hfoB*, has been expressed in *E. coli*, and rHFoB protein purified (K. Sandman, personal communication). The thermal stability of rHFoB was therefore determined in the presence and absence of 1.5 M gdm-cl, in the buffer and salt conditions found to be most stabilizing for the HMf proteins (1.025 M KPi, pH 7.5). The respective thermal denaturation curves are compared in Figure 25A, and indicate that rHFoB unfolded completely, even in the absence of gdm-cl, but with gdm-cl added the $T_m$ was reduced by $\sim 13 ^\circ C$. In addition, rHFoB in the buffer containing 1.5 M gdm-cl
Kinetics of rHMfA and rHMfB refolding. Refolding of rHMfA (diamonds) and rHMfB (squares) from the denatured state at the temperature maximum of the assay (80-84 °C) to the renatured state at 25 °C was monitored by measurement of the CD at 222 or 230 nm as a function of time. Buffers used were 1.5 M gdm-cl, 1 M KCl, 25 mM KPi, pH 7.5 (A) and 2.75 M gdm-cl, 1.025 M KPi, pH 7.5 (B). Measurement of CD and cooling of the sample were initiated simultaneously, and data points were taken every 6 sec, using a 5 sec integration time. Curves shown were least-squares fitted using a four-parameter logistic function, and gave values for t1/2 (refolding) of 93, 163, and 146 sec for rHMfB in the 1.5 M gdm-cl buffer, and for rHMfA and rHMfB in the 2.75 M gdm-cl buffer, respectively. Correlation coefficients for the fits were >0.98.
Figure 24.
Figure 25.

Thermal denaturation of rHFoB, and comparison to rHMfA and rHMfB. A. CD data were measured at 222 nm, in 1.025 M KPi, pH 7.5 buffer that contained (circles) or did not contain (triangles) 1.5 M gdm-cl. B. Comparison of denaturation curves for rHFoB (circles), rHMfA (diamonds) and rHMfB (squares), in 1.5 M gdm-cl, 1.025 M KPi, pH 7.5. Fractional unfolding values ($f_u$) for each protein were calculated from the maximal folded and unfolded $[\theta]$ or raw ellipticity values that were determined as described in the Methods section of this chapter. For clarity, errors in $[\theta]$ or $f_u$, which were $\leq 3\%$, and are not shown.
Figure 25.
was already ~35% unfolded at its maximum state of folding (at 25 °C), relative to its maximal state of folding in the buffer that contained no gdm-cl (at 12 °C), and underwent almost complete denaturation at both extremes of the 0-80 °C temperature range tested (Figure 25A). The concentration of rHFOB used in these experiments was not determined accurately enough to compare [θ] values to those of rHMfA and rHMfB, however fractional unfolding values could be calculated and were used for direct comparisons with previous results (Figures 20, 25B). Under the conditions at which rHMfA and rHMfB became only 20% unfolded (1.025 M KPi, 1.5 M gdm-cl, pH 7.5; 0-80 °C), rHFOB underwent two transitions (a ‘cold’ and ‘hot’ denaturation), corresponding to an almost complete thermal denaturation curve, as predicted for protein unfolding (Privalov, 1990). The ΔT_m between rHFOB and the rHMf proteins is at least 33 °C (Figure 25B), and extrapolation of curves suggest this value is closer to 50 °C (data not shown).

Heat denaturation of rHFOB was almost completely reversible, with ≥95% refolding after cooling from 80 °C in both the absence (Figure 26A) and presence (Figure 26B) of 1.5 M gdm-cl. The kinetics of rHFOB refolding (Figure 27) were not as rapid as for the rHMf proteins (Figure 24), even in the absence of gdm-cl, with calculated t_1/2 values for refolding of 168 and 288 sec, in the absence and presence of 1.5 M gdm-cl, respectively.

DISCUSSION

The CD data presented here demonstrate that rHMfA and rHMfB have very stable secondary structures, provided that sufficient salt is present. At pH 7.5, there is little change in the helical content of either of these proteins above ~100 mM KCl (Figure
Figure 26.

Reversibility of rHFoB thermal denaturation. CD spectra were measured in 1.025 M KPi, pH 7.5 buffer that did not contain (A) or contained (B) 1.5 M gdm-cl. Spectra were measured at 25 °C, prior to denaturation (---), after 3.0-3.5 h when the maximal denaturation temperature of 80 °C was achieved (···), and 20 min after cooling to 25 °C (---). Refolding of rHFoB in the absence (A) and presence (B) of 1.5 M gdm-cl was ~95% and >97%, respectively. Each unsmoothed spectrum shown is the average of 4-6 scans.
Figure 26.
Figure 27.

Kinetics of rHFoB refolding. Refolding, from the denatured state at 80 °C to the renatured state at 25 °C in the absence (− gdm-cl; diamonds) and presence (+ gdm-cl; squares) of 1.5 M gdm-cl in 1.025 M KPi, pH 7.5 buffer, was monitored by measurement of the CD (ellipticity) at 222 nm as a function of time. Measurement of CD and cooling of the sample were initiated simultaneously, and data points were taken every 6 sec, using a 5 sec integration time. Curves shown were least-squares fitted using a four-parameter logistic function, and gave values for $t_{1/2}$ (refolding) of 168 and 288 sec for rHFoB, in the absence and presence of 1.5 M gdm-cl, respectively. Correlation coefficients for the fits were $>0.99$. 

121
Figure 27.
Table 3. Melting temperatures of rHMfA, rHMfB, and rHFoB in different buffers

<table>
<thead>
<tr>
<th>Protein</th>
<th>0 M gdm-cl</th>
<th>1 M gdm-cl</th>
<th>1.5 M gdm-cl</th>
<th>2.75 M gdm-cl</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 M KPi</td>
<td>0.2 M KCl</td>
<td>1 M KCl</td>
<td>1 M KCl</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 M KCl</td>
<td>1 M K2SO4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 M KPi</td>
<td>1 M KPi</td>
</tr>
<tr>
<td>rHMfA</td>
<td>NDd</td>
<td>56</td>
<td>71</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>77</td>
<td>&gt;80e</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>61</td>
<td>61</td>
</tr>
<tr>
<td>rHMfB</td>
<td>NDd</td>
<td>65</td>
<td>79</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&gt;80e</td>
<td>&gt;80e</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>rHFoB</td>
<td>60</td>
<td>NDd</td>
<td>NDd</td>
<td>denaturedb</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NDd</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9c, 47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NDd</td>
</tr>
</tbody>
</table>

*a* Melting temperatures in °C, calculated by average method

*b* rHFoB is almost completely unfolded under these conditions

*c* rHFoB undergoes two thermal transitions, corresponding to 'cold' and 'heat' denaturation

*d* Melting temperature not determined for this condition

*e* Above the temperature range of the spectropolarimeter
124

18C), indicating that experiments carried out in vitro at neutral pH in the presence of ≥ 100 mM KCl should generate results that are relevant to the HMf structure in vivo.

Heat treatment of (rHMfA)_2 and (rHMfB)_2 mixtures was used to generate heterodimer containing mixtures similar to those found in native HMf preparations from Mt. fervidus (Chapter 2). These heat treated mixtures and identical, non-heat treated mixtures had identical CD spectra, and exhibited dichroism values between those of (rHMfA)_2 and (rHMfB)_2. Heterodimer formation therefore does not result in a conformational change that differs detectably from the homodimeric states, indicating that results of studies of rHMfA and rHMfB homodimers may be extrapolated legitimately to infer structural properties of the heterodimers. This is important as heterodimers are present in vivo, but without procedures to generate homogeneous preparations of heterodimers, their structure cannot be investigated directly.

Many of the primary sequence differences between rHMfA and rHMfB are conservative amino acid substitutions that lie within regions of the proteins that are known or predicted to form amphipathic α-helices (Figure 28A) (Grayling et al., 1994; Starich et al., 1995). There are, however, four non-conservative substitutions (N14D, E37R, R64V, and M66R) in these regions that should cause a net charge difference between rHMfA and rHMfB. Differences in electrostatic interactions within these proteins are therefore likely to be responsible for the differences in their salt-dependent stabilities at low pH (Figures 17C, D).

Estimates of the role of such electrostatic interactions on the stabilities of HMfA and HMfB can be made by examining pH titration curves, predicted by making the assumption that each titrable group within these proteins titrates independently, and with the same pKₐ as the corresponding free amino acid. The curves obtained indicate that the net charges on HMfA and HMfB should remain parallel, as a function of pH, with HMfB
Figure 28.

Predicted differences in the stabilities of HMfA and HMfB. A. An alignment of the amino acid sequences of HMfA and HMfB. Residues that differ between the two sequences are boxed, where shaded boxes indicate non-conservative differences. The secondary structure of the HMfB protein, known to contain α-helical secondary structures (indicated by the coiled lines) and unstructured or β-strand regions (solid lines), is also predicted for HMfA (Grayling et al., 1994; Starich et al., 1995). B. pH titration curves predicted for the HMf proteins. The algorithm used is incorporated in the ISOELECTRIC program (Devereux et al., 1984), and assumes that each amino acid titrates independently with pKa values corresponding to those of the free amino acids. Solid and dashed lines indicate the formal net charges predicted for rHMfA and rHMfB, respectively, over the pH range 1-13.
A.

Helix I

Helix II

Helix III

B.

Figure 28.
maintaining approximately a +2 greater formal net charge over the entire pH range (Figure 28B). Electrostatic interactions should therefore be greater for HMfB than for HMfA at all pH values, but particularly at acidic pHs, where these proteins have high positive charges. Furthermore, added salt should stabilize both proteins through electrostatic charge shielding, and this effect should be greater for HMfB than for HMfA. These predictions are consistent with the salt titration data acquired at pH 2. Complete denaturation of rHMfB, but only partial denaturation of rHMfA was observed at pH 2 and <0.1 M KCl concentrations (Figure 17C), conditions where ion-dependent shielding of ionized groups should be least effective. Also, both proteins were stabilized at pH 2 by increasing KCl concentration, and KCl uptake was more cooperative for rHMfB than for rHMfA, as indicated by cooperativity constants of 3.1 and 2.7, respectively. This degree of cooperativity could result from site-specific ion binding (most likely by chloride ions), but there are, as yet, no experimental data that address this parameter. It is not clear, however, why at the higher salt concentrations the helicity of HMfA at pH 2 (84%) was so different from that at pH 7.5 (72%). In addition to electrostatic interactions, other factors associated with protein stabilization, such as intrahelical salt bridges (Marqusee and Baldwin, 1987), high alanine content (Menéndez-Arias and Argos, 1989), and increased hydrophobic interactions, probably also contribute to the stability profiles of these proteins at low pH. The large changes observed in helicity at pH 2, and the parallel nature of the pH titration curves (Figure 28B), indicate that studies of the HMf proteins using low salt conditions, away from neutral pH, will not generate results meaningful to the in vivo situation. Nevertheless, the conditions established in the salt titration study should be generally useful for future comparative analysis.
The salt-dependent conformational stabilities of rHMfA and rHMfB determined by salt titration are consistent with the high internal salt environment present in vivo in *Mt. fervidus* (~1 M K⁺, ~300 mM cDPG³⁻), since presumably these proteins have evolved for optimal function in this environment. That the HMf proteins are stabilized by increasing salt concentration is supported by the results of thermal denaturation experiments (Figure 19; Table 3). The melting temperatures of rHMfA and rHMfB were increased with increasing KCl concentration, by 14-15 °C for the transition from 0.2 to 1 M KCl. Melting temperatures were also found to be very dependent on the salt anion used, with phosphate having the most thermostabilizing effect (Table 3), in agreement with previous studies of enzymes derived from *Mt. fervidus* and *Mp. kandleri* (Hensel and König, 1988; Breitung et al., 1992; Hensel and Jakob, 1994). Hensel and Jakob (1994) showed that high potassium phosphate concentrations stabilize *Mt. fervidus* enzymes, both conformationally, and with respect to covalent modifications, such as deamidation of asparagine residues and peptide bond hydrolysis, even though the high ionic strength and temperature in vivo favors such chemical reactions. rHMfA and rHMfB are also likely to be protected from these chemical reactions, as their refolding to native conformation is complete after thermal denaturation, at least under the conditions of the assay (Figure 23).

Native HMf proteins must experience at least two types of phosphate interactions in the intracellular environment, since they bind DNA (presumably via the phosphate backbone), and since the dominant anion in *Mt. fervidus* cells is cDPG³⁻. It is possible, and likely, that the HMf proteins have surface charge distributions that favor coordination of the high charge density phosphate group. Potassium phosphate may therefore be a sufficiently close ‘mimic’ of DNA and/or cDPG³⁻ that it has an in vitro thermostabilizing effect on the HMf proteins that is comparable to what results in vivo.
The Van't Hoff enthalpies of unfolding ($\Delta H_{\text{app}}^u$) calculated for rHMfA and rHMfB in 1 M KCl, 1.5 M gdm-cl, 25 mM KPi, pH 7.5 were +110 and +76 kJ/mol (dimer), respectively. However, rHMfB has a greater $T_m$ than rHMfA under these, and in fact all, conditions (Table 3), and therefore the respective $\Delta G$ values for the unfolding reaction must be such that $\Delta G_{\text{app}}^u(rHMfB) > \Delta G_{\text{app}}^u(rHMfA)$. For this to be the case, $\Delta S$, which is a positive quantity for protein denaturation, must be significantly smaller for rHMfB than it is for rHMfA (see equation 5). One interpretation of these results is that although rHMfA is apparently more enthalpically stabilized than rHMfB (perhaps by virtue of an increased number of inter- and intra-molecular hydrogen bonds, van der Waals interactions, and electrostatic interactions), rHMfB could be more entropically stabilized, by having a larger entropy in the folded state which would lead to a smaller $\Delta S$ upon unfolding. A larger entropy in the folded state is likely to result from increased hydrophobic interactions within the folded protein, since solvation of hydrophobic groups during unfolding actually results in a decrease in entropy (Privalov, 1990). This interpretation would therefore infer that rHMfB is stabilized towards thermal denaturation over rHMfA largely by increased hydrophobic interactions within the folded dimers, an effect that should not be influenced by salt type or concentration. In fact, the $\Delta T_m$ between rHMfA and rHMfB is not influenced by changing salt type or concentration (Table 3). In addition, the observation that rHMfB seems to cold denature at slightly higher temperatures than rHMfA (Figure 20) is consistent with rHMfB having increased hydrophobic interactions in the folded state, relative to rHMfA (see below).

Hydrophobic interactions must be important in the thermostabilization of both HMfA and HMfB, since the recombinant forms of these proteins are stabilized towards thermal denaturation by high salt concentrations (Figure 17C, D; Table 3). Under these conditions, electrostatic interactions and hydrogen bonding should be greatly reduced in
ion accessible regions of the protein structures. A stabilizing mechanism involving very efficient hydrophobic packing is therefore likely, and this would be consistent with the high salt environment that the HMf proteins experience \textit{in vivo} (Hensel and König, 1988).

Both rHMf proteins and rHFoB undergo cold denaturation in the presence of 1.5 M gdm-cl (Figures 20, 25). Cold denaturation is now recognized as an inherent property of proteins, and both heat and cold denaturation can be described accurately by a single thermodynamic function (Privalov, 1990; DeKoster and Robertson, 1995). Although there is still significant debate, cold denaturation is generally ascribed to result from decreases in the hydrophobic contribution to stability, that result from an increased negative $\Delta G$ for hydration of hydrophobic groups as temperature decreases. In contrast, heat denaturation is accepted to result mainly from configurational entropy effects that favor protein unfolding as temperature increases (Privalov, 1990; Murphy and Friere, 1992; DeKoster and Robertson, 1995). Clearly, further study is needed, however, a current major limitation to studying cold denaturation is the difficulty of achieving significant denaturation at experimentally accessible temperatures. Most studies to date have relied on the use of moderate concentrations of denaturants such as gdm-cl or urea, low pH, or proteins intentionally destabilized by site-directed mutagenesis (SDM) (Chen and Schellman, 1989; Antonino et al., 1991; Azuaga et al., 1992; Kuroda et al., 1992). In one recent study where such conditions were not required, the protein of interest (CheY) cold denatured effectively at low gdm-cl concentrations, but its heat denaturation was only 85% reversible (DeKoster et al., 1995). The rHMf proteins, and in particular, rHFoB, appear to be ideally suited to cold denaturation studies. rHFoB cold denatures completely using only 1.5 M gdm-cl (Figure 25), and its heat denaturation is >95%
reversible. Furthermore, rHFoB is small, and its 67 amino acids are readily manipulated by SDM.

The DNA binding activity of native HFo preparations is much less thermostable (t_{1/2} at 95°C of ~1.5 h) than that of HMf preparations (t_{1/2} at 95°C of >5 h) (Darcy et al., 1995). This is supported by the thermal denaturation studies (Figure 25B) that indicate a ΔT_{m} between rHMfB and rHFoB of >33 °C (Table 3). Presumably, HFoA1 and HFoA2 are similarly less thermostable. The primary sequences of rHMfB and rHFoB are highly conserved (Figure 29), and only 8 of the 15 differences correspond to significant, non-conservative changes, with most of these present in the region that forms the long, central α-helix in rHMfB (Darcy et al., 1995; Starich et al., 1995). Presumably some or all of the differences between these proteins must be responsible for the greatly reduced thermostability of rHFoB relative to rHMfB. Several residues that might confer increased thermostability on rHMfB were identified (Starich et al., 1995), and seem to result in improved hydrophobic packing. However, no obvious features, such as loss of a salt-bridge or hydrogen bonding network have been identified in comparisons of the rHFoB and rHMfB primary sequences. The amino acids that do contribute to the thermostabilization of rHMfB over rHFoB are probably best identified by the stepwise conversion of rHMfB into rHFoB, using SDM. By examining the stability of protein variants produced using the CD-based thermal denaturation assay and conditions developed here, it should now be possible to identify which residues, individually or in combination, confer increased thermostability on rHMfB.

Based on the data presented in this chapter, the members of the HMf protein family (Chapter 1) appear to provide ideal models for the study of a number of important problems in protein biochemistry. The recombinant versions of these proteins are easy to synthesize and purify in large amounts from E. coli and are readily modified by SDM.
Figure 29.

Alignment of the amino acid sequences of HFoB and HMfB. Residues that differ between the two sequences are boxed, where shaded boxes indicate non-conservative differences. The three regions of the proteins known or predicted to have \(\alpha\)-helical secondary structures are indicated, as in Figure 28A.
They are small in size, and therefore the 'wild-type' proteins and any SDM-generated variants are readily amenable to structure-function studies (Starich et al., 1995). Their simple, almost entirely α-helical structures should, in addition, facilitate studies of the mechanisms of protein folding. Although HMfA, HMfB, and HFoB have very similar primary sequences, they differ dramatically in their thermal stabilities, and undergo both heat- and cold-unfolding under only mildly denaturing conditions, and in completely reversible manners. These proteins therefore are ideal models to investigate the thermodynamics and mechanisms of protein folding, the origins of protein thermostability, and the phenomenon of cold denaturation.
CHAPTER IV
THE SEQUENCE SPECIFICITY OF HMf-DNA INTERACTIONS

INTRODUCTION

DNA tertiary structures vary dramatically, because of intrinsic, sequence-directed curvatures that occur as a result of differences in the roll, tilt, and twist angles (the 'wedge' angles; Figure 5) between adjacent basepairs, or from curvatures that result from protein binding (Trifonov, 1985; Travers, 1990; Harrington, 1992). Sequence-directed DNA curvature depends, in magnitude, on the wedge angles between each of the 16 possible dinucleotide pairs (Bolshoy et al., 1991). Local bends arise from the 'phasing' of large wedge angles, such as those between AA, AG, and CG dinucleotide pairs with the 10.5 bp helical repeat in B-DNA (Trifonov, 1985; Bolshoy et al., 1991). The best studied DNA curvature elements are 'poly-dA' tracts, as very significant DNA curvature is associated with runs of (dA)$_n$ bp (where $n$ is $\geq 5$) phased with the helical repeat (Crothers et al., 1990).

Intrinsic DNA curvature is important in both sequence-specific and sequence-non-specific protein-DNA interactions, where it can facilitate sequence recognition by wrapping around proteins, or provide 'looped' regions that facilitate interactions between adjacent, DNA-bound proteins such as transcription factors (Kahn and Crothers, 1993; Harrington, 1992; Wolffe, 1994a). In addition, intrinsically curved DNA sequences or sequences with intrinsic ability to be curved should thermodynamically favor the binding of proteins that cause DNA bending (Travers, 1990; Harrington, 1992; Kahn and
Crothers, 1993). This seems to be the case for the eukaryal histones which appear to position nucleosome assembly in vivo, at least in part, by localizing at intrinsically curved DNA sequences (Drew and Travers, 1985; Satchwell et al., 1986; Thoma, 1992). This positioning is determined by the (H3•H4)2 tetramer component of the nucleosome (Dong and van Holde, 1991), which presumably therefore recognizes curved DNA as a positioning signal.

The HMF proteins have structural features in common with the eukaryal histones (Chapter 1), and therefore might also bind preferentially and assemble NLS at intrinsically curved DNA sequences. An initial study, using preparations of native HMF protein isolated from Mt. fervidus, demonstrated that HMF was indeed bound and localized preferentially in vitro at highly curved Crithidia fasciculata kinetoplast DNA fragments cloned in pBR322, when compared with binding to the remainder of the plasmid (Howard et al., 1992). To extend this study, and to investigate the DNA sequence determinants of preferential binding sites, DNA sequences that were protected from micrococcal nuclease (MNase) digestion by assembly in vitro into HMF-DNA complexes, were cloned and sequenced.

**MATERIALS AND METHODS**

**Chemicals and reagents**

Chemicals and MNase (Sigma #N-5386) were purchased from Sigma Chemical Co. (St Louis, MO), acrylamide from United States Biochemical (USB; Cleveland, OH), and restriction enzymes from Gibco-BRL (Life Technologies, Inc.; Gaithersburg, MD). Double-distilled water (ddH2O) was used in all experiments. Recombinant (r) HMFa and
rHMfB were purified from *E. coli* strains KS1183 and KS1076, respectively, as described in Chapter 2.

**Bacterial strains and plasmids**

All cloning and propagation of plasmid vectors was carried out in *E. coli* strain DH5α-F' (Woodcock *et al*., 1989). Plasmids used were, pLITMUS 28 (general purpose cloning vector; New England Biolabs [NEB], Beverly, MA); pUC19 [general purpose cloning vector; (Yanisch-Perron *et al*., 1985)]; pUC1318 [general purpose cloning vector containing inverted-repeat poly linker region; (Kay and McPherson, 1987)]; pBEND2 [used for analysis of curved DNA sequences; (Kim *et al*., 1989)]; pRG101 [previously designated pAC101; obtained from J. Nölling; contains methanogen plasmid pME2001 (Meile *et al*., 1983) cloned in the *Sma*I site of pUC19].

**Molecular biological methods**

Standard molecular biological procedures were carried out according to Sambrook *et al.* (1989), and Ausubel *et al.* (1990), or according to the directions provided by the manufacturer of a particular reagent, enzyme, or kit. All phenol/chloroform extractions used 200 μl of DNA solution and involved the addition first of an equal volume of a 1:1 mixture of phenol:chloroform, followed by a second extraction with 200 μl of chloroform.

**Ethanol precipitation of small DNA fragments**

Ethanol precipitations of DNA fragments <200 bp in length from aqueous solution used the modification suggested by Sambrook *et al.* (1989), with MgCl₂ added to 10 mM in addition to the usual 300 mM sodium acetate, and used 3 volumes of 95% ethanol. Precipitation was allowed on ice for 20 min, followed by Eppendorf
centrifugation at room temperature for 20 min. Precipitates of these small DNA fragments were washed with ice-cold 80% ethanol, dried under a vacuum, and dissolved in ddH$_2$O.

**Dephosphorylation reactions**

DNA fragments were dephosphorylated by using the protocol supplied with the BRL alkaline phosphatase.

**T4 DNA polymerase ‘fill-in’ reactions**

End-filling reactions, necessary as a cloning step, used 1 µg of DNA in a buffer that contained 50 mM Tris-HCl (pH 7.2), 10 mM MgSO$_4$, 0.1 mM dithiothreitol, 50 µg/ml bovine serum albumin, 100 µM of each dNTP, and 6 units T4 DNA polymerase (Promega, Madison, WI). The reactions were incubated at 37 °C for 10 min, and terminated by addition of EDTA to 25 mM, and placed on ice.

**Primers used**

Three primers were used for DNA sequencing and/or PCR amplification during cloning:

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence:</th>
<th>Source/design:</th>
<th>Use:</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1212</td>
<td>5' -GTTTTCCCAGTCAGCAGAC-3'</td>
<td>(NEB, Beverly, MA)</td>
<td>PCR</td>
</tr>
<tr>
<td>#1233</td>
<td>5' -AGCGGATAACAATTTCACACAGGA-3'</td>
<td>(NEB, Beverly, MA)</td>
<td>Sequencing of 60 bp clones in pLITMUS 28; PCR</td>
</tr>
<tr>
<td>RAG1</td>
<td>5' -TAAAAATAGGCGTATCAGCAGG-3'</td>
<td>R. A. Grayling</td>
<td>Sequencing of 60 bp clones in pBEND2</td>
</tr>
</tbody>
</table>

Primers #1212 and #1233 are the standard pUC19 universal forward and reverse sequencing primers, respectively.
Preparation of plasmid DNAs

Small-scale preparations of plasmid DNAs ('minipreps') used the alkaline lysis procedure of Sambrook et al. (1989), or the procedure of Yie et al. (1993). Large-scale plasmid DNA was obtained by using the CsCl methods of Sambrook et al. (1989) and Ausubel et al. (1990). Some of the micrococcal nuclease experiments required very large amounts of pRG101 DNA (100 µg per experiment) that were tedious to prepare using CsCl methods and therefore, a modified alkaline lysis procedure was used (Feliciello and Chinali, 1993). This procedure gave high plasmid yields (≥3 mg per 250 ml culture) at a purity equal to or better than the CsCl methods.

Micrococcal nuclease digestion

Linear DNA was used to avoid any variabilities in HMf-DNA interactions that might result from different levels of supercoiling of different plasmid DNA preparations. Either EcoRI-linearized pUC19 or SmaI-linearized pRG101 DNA (10 µg) was incubated with increasing amounts of native HMf, rHMfA, or rHMfB, in 95 µl of digestion buffer [50 mM Tris-acetate (pH 8.8), 1 mM CaCl₂, 100 mM NaCl] at 37 °C for 20 min. Micrococcal nuclease [5 µl of 0.01 U/µl, diluted in 50 mM Tris-acetate (pH 8.8), 100 mM NaCl] was then added and incubation continued at 37 °C for 30 sec to 10 min. The digestion reactions were terminated by the addition of 10 µl of 0.1 M EDTA, diluted to 200 µl with ddH₂O, extracted with phenol/chloroform, and ethanol precipitated (as described above), prior to drying. An ethanol wash was not necessary, in most cases, and the dried pellets were dissolved in 10 µl of 1X gel loading buffer [2% Ficoll 400, 10 mM EDTA (pH 8), 0.025% bromophenol blue] for analysis by PAGE or agarose gel electrophoresis. Where indicated, crosslinking before the MNase digestion used either 40 mM HCHO (for protein-only samples) or 140 mM HCHO (for protein-DNA complexes),
in a buffer that contained 100 mM NaCl, 50 mM TEA-HCl (pH 8), using the crosslinking reaction conditions described in Chapter 2. After the crosslinking, the buffer was exchanged with 100 mM NaCl, 50 mM Tris-acetate, pH 8.8 (or Tris-HCl, pH 8) by microdialysis or ultrafiltration (Microcon units; Amicon, Beverly, MA), before MNase digestion. Crosslinked and MNase-digested samples were phenol/chloroform extracted only after digestion first with 20 μg/ml proteinase K (37 °C for 30 min). This was needed to remove protein crosslinked to DNA, because otherwise the complexes partitioned into the organic phase during the extraction.

MNase dilutions and all MNase digestion reactions used 500 μl plastic microfuge tubes that had been pre-treated with Sigmacote (Sigma #SL-2), to prevent the enzyme absorbing to the tubes.

**Electrophoresis techniques**

The same 10X sample buffer [20% Ficoll 400, 0.1 M EDTA (pH 8), 0.25% bromophenol blue] was used for both agarose gel electrophoresis and for PAGE. Both types of gel were stained using 1 μg/ml ethidium bromide (EtBr) dissolved in 1X TAE buffer (40 mM Tris-acetate, 2 mM EDTA, pH 8.5) or 1X TBE buffer (89 mM Tris base, 89 mM boric acid, 2 mM EDTA, pH 8), followed by destaining in the same buffer without EtBr.

**Agarose gel electrophoresis**

Electrophoretic separation of DNA fragments <1 kb in length, dissolved in 1X sample buffer, used 4 % NuSieve GTG agarose (FMC BioProducts, Rockland, ME) polymerized in 1X TAE buffer as directed by the manufacturer, and 3 to 7 V/cm.
DNA samples that had been extracted with phenol/chloroform and ethanol precipitated were dissolved in 1X sample buffer and loaded onto 8% T or 10% T polyacrylamide gels polymerized in 1X TBE buffer, using a vertical apparatus (Idea Scientific, Minneapolis, MN). 33:1 (3% C) and 75:1 (1.3% C) acrylamide:bisacrylamide ratios were used for the analysis of MNase digestion products and the analysis of DNA fragments for intrinsic curvature, respectively. These gels were electrophoresed at 5 V/cm or 4.3 V/cm (for curvature analyses) in 1X TBE buffer, for 2.5 to 4.5 h, prior to EtBr staining.

**Gel purification of 30 and 60 bp DNA fragments**

Bands containing ~30 or ~60 bp DNA fragments, resolved by PAGE or agarose gel electrophoresis, were excised and purified using either the crush and soak method of Sambrook *et al.* (1989) (PAGE-resolved fragments), or a Mermaid kit [agarose gel-resolved fragments (BIO101, La Jolla, CA)].

**DNA sequencing**

DNA samples were prepared for sequencing directly from overnight 1.5 ml cultures of plasmid-containing *E. coli* DH5α-F' cells using the method of Yie *et al.* (1993). Double-stranded DNA sequencing was carried out using the dideoxy chain termination method (Sanger *et al.*, 1977) with deoxycytidine 5'-[α-35S]-thiotriphosphate (ICN Biomedicals, Irvine, CA), and Sequenase Version 2.0 (USB, Cleveland, OH), according to the supplier's instructions. The #1233 and RAG1 primers were used for sequencing DNA fragments cloned in pLITMUS 28 and pBEND2, respectively. The DNA sequences obtained were analyzed by using the Wisconsin Sequence Analysis
PCR amplification conditions

PCR amplifications in 100 µl reaction mixtures contained 10 pmol of each of #1212 and #1233 primers, 6 mM MgCl₂, 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.25 mM of each dNTP, ~1 ng plasmid DNA template, and 5 U Taq DNA polymerase. Mineral oil overlays (100 µl) were added to prevent evaporation. Amplification reactions started with denaturation at 95 °C for 3 min, followed by 30 cycles of 1.5 min denaturation at 94 °C, 1.5 min annealing at 47 °C, and 1 min extension at 72 °C, followed by a final 5 min extension at 72 °C before termination of the reactions by incubating at 4 °C. Amplified DNA was phenol/chloroform extracted and ethanol precipitated before use.

End-labelling procedure

The 30 and 60 bp DNA fragments generated by MNase-digestion were 5' end-labelled with deoxyadenosine 5'-[α-³²P] triphosphate using either the 'forward' or 'exchange' reactions of T4 polynucleotide kinase (Ausubel et al., 1990; Sambrook et al., 1989). Reaction mixtures were terminated by the addition of EDTA to 20 mM, diluted to 200 µl with ddH₂O, phenol/chloroform-extracted, and unincorporated nucleotides were removed by two successive ethanol precipitation and wash steps. The resulting air-dried, ³²P-labelled DNAs (probes) were dissolved in ddH₂O, and had specific activities of ≥ 5 x 10⁸ dpm/µg.
DNA hybridization analysis

Blotting and hybridization with labelled DNA probes (Southern, 1975) used variations of the alkaline transfer and hybridization procedures as suggested by Chomczynski and Qasba (1984) and Meinkoth and Wahl (1984).

Membrane blotting procedure

Restriction fragments, separated by agarose gel electrophoresis, were depurinated by soaking the gels in 0.25 M HCl for 30 min, followed by rinsing with ddH2O and denaturation for 30 min in transfer buffer (0.4 M NaOH, 0.6 M NaCl). Alkaline transfer of the fragments to Zetaprobe charged nylon membranes (Bio-Rad Laboratories, Hercules, CA) was achieved by overnight capillary action (Southern, 1975) using transfer buffer. The membranes were then neutralized in 0.5 M Tris-HCl (pH 7.5), blotted dry on paper towels, and either used immediately or stored moist at 4 °C, covered with plastic wrap. Restriction fragments separated by PAGE were depurinated (as above), the gels were neutralized, and the DNA molecules were electrophoretically transferred to Zetaprobe membranes under neutral buffer conditions using a ‘submerged’ transfer apparatus (‘Transphor’; Hoefer Scientific Instruments, San Francisco, CA) with 0.5X TBE buffer (8 h, 35 V). Denaturation and fixation of membrane-bound DNA used the above conditions, and transfer buffer.

Hybridization and washing

Membranes were incubated in prehybridization buffer [50 % formamide, 50 mM Tris-HCl (pH 7.5), 1 M NaCl, 2.5 % SDS, 100 µg/ml sonicated, heat-denatured calf thymus DNA] at 60 °C for 15 to 30 min with agitation (to remove residual bound agarose or acrylamide), followed by continued incubation at 38 °C for 15 to 30 min. Hybridizations were then initiated by the addition of heat-denatured probe DNA using fresh prehybridization buffer (final radioactive concentration of 5-10 x 10^5 dpm/ml) that
contained 10% polyethylene glycol 8000, and continued by incubation of the membranes in this buffer at 38 °C for 16 h. Following hybridization, the membranes were washed twice (15 min each) at 48 °C in 2X SSC buffer (300 mM NaCl, 30 mM sodium citrate), followed by two washes (30 min each) at 48 °C in 0.1X SSC (60 bp probe) or 0.5X SSC (30 bp probe). The stringency of the final wash conditions was calculated to allow ~10% base-mismatch for both probes. Washed membranes were blotted dry on paper towels, covered with plastic wrap, and used to expose X-ray film (BioMax MR; Kodak, Rochester, NY) in the presence of intensifying screens (Lightning Plus, DuPont NEN, Wilmington, DE), at -70 °C.

**DNA bending predictions and bend analysis**

**Basic software**

The DNA trajectories (in three-dimensional space) of DNA fragments were predicted and analyzed by using the program **AUGUR** (Tan *et al.*, 1988) which allows the choice of any one of six models that predict the helical parameters (bp roll and tilt) of B-DNA with a fixed twist angle, for a given sequence. In this work, the ApA wedge model (Ulanovsky and Trifonov, 1987), currently the most popular and best documented predictor, was used. **AUGUR** predicts the trajectory of a given sequence, displays this graphically as a stereo image, and calculates two parameters, the base-pair ratio (bp-ratio) and $d_{\text{max}}$, that relate to the predicted curvature of the particular sequence. The bp-ratio is the ratio of the distance (in 3D space) between the two ends of the DNA molecule and the physical contour length. This must be $\leq 1$, with a value of 1 corresponding to an absolutely straight trajectory. The $d_{\text{max}}$ is the largest distance (in Å) between a base pair in the molecule and a straight line drawn to connect the ends of the DNA molecule. This is 0 for a straight fragment, as illustrated below in Figure 30.
Figure 30.
Calculation of DNA bending parameters by AUGUR.

Calculation of local curvature for pRG101 sequence

Figure 30 also illustrates the procedure used to calculate a 'bending index' along the entire length of pRG101. AUGUR was used to calculate bp-ratio values for overlapping 60 bp sequences that were staggered by 10 bp. The value generated for each 60 bp sequence was assigned to the central bp position of that sequence, thereby generating the predicted local curvature of pRG101 at a 10 bp resolution. The bp-ratio values thus obtained were scaled so that a value of 0 corresponded to a straight DNA sequence, and a value of 100 corresponded to a (circular) DNA sequence in which the ends must meet. The 60 bp size for analysis was chosen to correlate with the size of the protein-protected DNA fragments obtained following MNase digestion of NLS.
Densitometry and the determination of relative electrophoretic mobilities of DNA restriction fragments in polyacrylamide gels used the *NIH Image* software (National Institutes of Health, Bethesda, MD). Images, obtained using a UVP Gel Documentation System (Ultraviolet Products Inc., San Gabriel, CA), were stored in 8-bit Tagged Image File Format (TIFF), and imported directly into the *NIH Image* program for analysis. Relative mobilities were calculated by using a 404 bp fragment in a *MspI* digest of pUC19 as a reference. Mobility versus polylinker site curves, generated in the pBEND2 clone circular permutation analysis (Figure 43), were fit with a second-order polynomial function, using the curve-fitting option of *Cricketgraph* (Computer Associates International, Inc.).

**RESULTS**

**MNase digestion of HMf-DNA complexes assembled in vitro**

Micrococcal nuclease is a sequence non-specific nuclease that cleaves single- and double-stranded DNA and RNA molecules yielding oligo- and mono-nucleotides with 3' phosphates (Alexander et al., 1961). MNase digestion is inhibited by proteins bound to the DNA, and MNase digestion of eukaryal histone H1-depleted chromatin and analysis by PAGE results in a characteristic pattern of regular bands corresponding to multiples of 140-150 bp (Noll, 1974a, b; Noll, 1978). This was shown to be the length of DNA protected by a histone octamer in a nucleosome (Chapter 1), and MNase digestion of histone-DNA complexes assembled in vitro also resulted in the protection of similar multiples of 140-150 bp when analyzed by PAGE (Thomas and Butler, 1978). The mean
size of the DNA molecules in these bands has now been refined to 146 ± 1 bp (Noll, 1974a, b; Wolffe, 1992).

If the HMf proteins resemble the eukaryal histones in terms of their DNA binding, then they should also protect DNA molecules from MNase digestion. MNase digestion-protection studies were therefore undertaken, using pUC19 DNA linearized with EcoRI to eliminate plasmid supercoiling effects on the HMf-DNA interactions, and protection from MNase digestion by native HMf was observed (Figure 31A). HCHO-crosslinked and non-crosslinked HMf-DNA complexes gave the same digestion patterns, except that the crosslinked complexes were slightly more resistant to digestion (data not shown). Protection was not specific to pUC19 DNA, as MNase digestion of pRGI01 DNA linearized with Smal (Figure 31B) and pBR322 DNA linearized with PstI (data not shown) gave similar patterns. At high protein/DNA molar ratios (≥ 200; Figure 31A), or using short digestion times or low MNase concentrations (Figure 31B), bands corresponding to multiples of ~60 bp were observed. At very low protein/DNA molar ratios (≤ 2; Figure 31A), or very low MNase concentrations (Figure 31B), and additional ~30 bp protected band was observed.

By analogy with the eukaryal nucleosome, the ~60 bp DNA fragment remaining after extensive MNase digestion (Figure 31A, lane ‘20’; Figure 31B) is predicted to correspond to the length of DNA protected in a single HMf tetramer-DNA complex (Chapters 1, 5), and the multiples of 60 bp DNA are predicted to result from protection by closely spaced, adjacent HMf-DNA complexes. The ~30 bp protected fragment was the only protected species observed at very low protein/DNA molar ratios (≤ 0.2) and this therefore probably resulted from protection against MNase digestion by an HMf dimer, since at such low molar ratios, DNA wrapping and tetramer formation is very rare (HMf is a dimer in solution; Chapter 2).
Figure 31.

MNase digestion of HMf-DNA complexes assembled *in vitro*. A. Protection of linear pUC19 at different native HMf /DNA molar ratios. *Eco*RI-linearized pUC19 DNA (10 µg) and native HMf were mixed at protein/DNA molar ratios of 0.2, 2, 20, and 200 and then exposed to MNase (0.5 mU/µl) for 2 min (0.2, 2 ratios) or 10 min (20, 200 ratios), followed by extraction with phenol/chloroform and PAGE analysis on 8% T, 3% C gels. The approximate lengths (in bp) of protected DNA fragments, and their presumed origins are indicated in the cartoon to the right of the figure (see text). B. Time course of MNase digestion of *Sma*I-linearized pRG101-native HMf complexes formed at a protein/DNA molar ratio of 200. The complexes were exposed to MNase (0.25 mU/µl) for 3, 6, and 8 min, before phenol/chloroform extraction and electrophoresis of duplicate samples of each reaction mixture on a 4% agarose gel. The approximate lengths of protected fragments are indicated. M, pUC19/*Msp*I size standards (501, 489, 404, 331, 242, 190, 147, 110, 67, 34, 26 bp). The controls used in panels A and B were linearized pUC19 and pRG101 DNA, respectively, that was treated with (+) or without (−) MNase for 2 min.
Figure 31.
Sequence specificity in HMf-DNA binding

Two approaches to assess the sequence specificity of HMf binding to DNA were used with both pUC19 and pRG101 DNA, with native and rHMfA protein, respectively (Figure 32). The first qualitative approach used hybridization of $^{32}$P-end-labelled, MNase-protected 60 bp DNA fragments isolated from each of the two plasmids, to probe restriction digests of the same plasmid. The second approach was to clone and sequence such ~60 bp DNA fragments. A similar technique has been used to analyze DNA fragments protected against MNase digestion by their incorporation into eukaryal nucleosomes, and is referred to as ‘statistical sequencing’ (Satchwell et al., 1986; Ambrose et al., 1990).

Hybridization analysis of 30 and 60 bp DNA fragments

If some 60 bp fragments predominated in the end-labelled probe population, due to localized HMf binding, then this should be detected by increased intensity of hybridization to some restriction fragments, not consistent with just the size-dependency of hybridization intensity. This was indeed observed for both pUC19 (Figure 33) and pRG101 (Figure 34). For pUC19, probes were generated from both 30 bp and 60 bp MNase-protected DNA fragments, as outlined in Figure 32, and when hybridized to restriction fragments from several pUC19 digests, the 30 bp (Figure 33A, B) and the 60 bp (Figure 33C, D) probes gave similar, but not identical, intensity patterns on autoradiograms. Hybridization to pUC19 DNA digested with only MspI confirmed this observation (Figure 33E). These results demonstrated that the 30 bp and 60 bp probes hybridized to predominantly the same sequences, and therefore that the MNase-protected ~30 bp DNA fragments were mostly a subset of the ~60 bp sequences. Variations in
Figure 32.

Approaches used to assess the sequence specificity of rHMfA-DNA binding. pRG101 DNA (20 μg), linearized with Smal, was incubated with rHMfA at a protein/DNA molar ratio of 200, and then digested with MNase (0.5 mU/μl) for 5 min. The protected ~60 bp DNA fragments formed a band following electrophoresis in a 4% agarose gel that was excised, and the fragments were purified using a Mermaid kit (BIO101, La Jolla, Ca). They were then end-labelled with $^{32}$P-dATP for use in Southern blot analyses, or end-filled with T4 DNA polymerase for cloning. Cloning was accomplished using dephosphorylated EcoRV-digested pLITMUS 28 vector, at a 150:1 insert:vector molar ratio and ligation at room temperature for 12 h. The LB-outgrowth step following transformation was limited to 30 min, to minimize the chance of recovering sibling clones. The same procedure was used to generate 30 bp and 60 bp probes from EcoRI-linearized pUC19-native HMf complexes, for use in Southern analyses (Figure 33).
**SmaI-linearized pRG101**

- rHMfA bound

**rHMfA-pRG101 complexes**

- digested with micrococcal nuclease and electrophoresed through 4% agarose gel

**protected 60 bp fragments excised from gel**

- 5'-end labeled with $^{32}$P-dATP
- used as the probe against restriction digests of pRG101
- Southern Analysis

- ends filled with T4 DNA polymerase
- blunt-end ligated into the EcoRV site of pLITMUS28
- Sequencing
Figure 33.

Southern blot analysis of pUC19 restriction digests, using 30 and 60 bp probes derived from pUC19-native HMf complexes, generated by the procedure outlined in Figure 32. A. and C. EtBr-stained polyacrylamide gels. pUC19 DNA was digested with MspI (lane 1), TaqI (lane 2-partial digest), Rsal (lane 3), PvuI (lane 4), Hinfl (lane 5), HaeIII (lane 6), DdeI (lane 7-partial digest), Sau3AI (lane 8), EcoRI (lane 9), and resolved by PAGE through 8% T, 3% C gels, followed by staining with EtBr. Sheared, salmon sperm DNA was used as a negative control (lane 10). B. and D. Autoradiograms of charged nylon membranes that contained the resolved DNA fragments were transferred from the gels shown in panels A and C. Membranes were probed with $^{32}$P-end-labelled 30 bp probe (B) or 60 bp probe (D). E. MspI-digested pUC19 DNA was resolved by PAGE through 8% T, 3% C gels in duplicate lanes. The resolved DNA fragments were transferred and hybridized (as above) to either the 30 bp or the 60 bp probe, followed by autoradiography. Single and double-headed arrows indicate DNA fragments that hybridized predominantly to one or both of the probes, respectively.
Figure 33.
Figure 34.

Southern blot analysis of pRG101 restriction digests. pRG101 digests obtained with HaeII+BglII+BstEII (lane 1), AvaI (lane 2), HincII (lane 3), RsaI (lane 4), StyI (lane 5), BamHI+EcoRI (lane 6), SmaI (lane 7), were resolved by electrophoresis through a 4% agarose gel (A). The DNA fragments were transferred to a nylon membrane and hybridized to the 60 bp probe, generated from pRG101-rHMfA complexes as outlined in Figure 32. The membrane was exposed to X-ray film to generate the autoradiogram shown in (B).
Figure 34.
band intensities could have resulted if a 60 bp sequence spanned a restriction site and thereby hybridized to two restriction fragments, whereas two 30 bp fragments from the same 60 bp sequence hybridized to a only one of the restriction fragments in these digests.

Only 60 bp probes were used to hybridize to restriction digests of pRG101 (Figure 34), and again, some sequences were clearly over-represented in the 60 bp probe population, indicating some localization of rHMfA binding (Figure 34B).

**Comparison of MNase protection by rHMfA and rHMfB**

Both rHMfA and rHMfB were tested for their ability to protect DNA sequences from MNase digestion, and both gave similar results, with ~60 bp DNA fragments the sole products after extensive MNase digestion (Figure 35A; lanes 3, 4). To determine if protein-protein crosslinking had an effect on HMf-DNA interactions, rHMfA and rHMfB were crosslinked with HCHO before being incubated with DNA and digested with MNase. In both cases, this crosslinking prevented DNA protection from MNase digestion (Figure 35A; lanes 1, 2). Crosslinking therefore either prevents DNA binding by (rHMfA)₂ or (rHMfB)₂ homodimers, or prevents their assembly into MNase-resistant HMf-DNA complexes (probably DNA-wrapped tetramers) after dimer binding (see Chapter 1). Both alternatives could result from HCHO neutralization of lysine positive charges, which presumably are required to bind negatively charged DNA, or from a crosslinking-mediated ‘stiffening’ of the dimer configuration which might prevent conformational flexibility required for complex formation.
Figure 35.

Comparison of DNA protection from MNase digestion by rHMfA and rHMfB, and analysis of the population of ~60 bp fragments used for cloning. A. rHMfA (rA) or rHMfB (B) protein was either crosslinked (lanes 1, 2) or not crosslinked (lanes 3, 4) with HCHO before incubation with Smal-linearized pRG101 DNA at a protein/DNA molar ratio of 100, and digestion with MNase (0.5 mU/μl) for either 1 min (lanes 1, 3) or 2 min (lanes 2, 4). Digestion products were resolved on a 4% agarose gel, and formed a single dominant ~60 bp band. Lanes M, +, and – contained size standards and the positive, and negative controls, respectively, as described in Figure 31. B. A sample of the population of ~60 bp DNA fragments used for cloning, end-labelled with 32P-dATP and resolved on a DNA sequencing gel. A sequencing ladder (A, G, C, T; generated using the #1233 primer and pLITMUS 28) provided the size standards indicated in bp to the right of the figure. The densitometric tracing of the resolved fragment population shown alongside the autoradiogram indicates that the center of the size distribution (---) is at 59 ±1 bp.
Figure 35.
Cloning of protected 60 bp DNA fragments

MNase-protected 60 bp DNA fragments, generated from rHMfA-pRG101 complexes, were cloned and sequenced to assess the specificity of rHMfA-DNA interactions (Figure 32). Plasmid pRG101, used for these experiments contains the entire *Mb. thermoautotrophicum* strain Marburg plasmid, pME2001 (Meile *et al.*, 1983) cloned in pUC19. *Mb. thermoautotrophicum* and *Mt. fervidus* are related methanogens (Woese, 1987) that contain very similar proteins, HMt and HMf, respectively (Chapter 1; Grayling *et al.*, 1994), and therefore, pRG101 DNA should be 'biologically relevant' for HMf binding. This plasmid must contain archaeal regulatory sequences and probably also binding sites that are recognized specifically by HMf.

A sample of 32P-end-labelled 60 bp fragments produced from a MNase digest of rHMfA-pRG101 complexes was analyzed by electrophoresis through a DNA sequencing gel, with an adjacent DNA sequencing ladder used as a size reference (Figure 35B). Densitometry indicated that the population of fragments had an approximately Gaussian size distribution centered at 59±1 bp (Figure 35B). An aliquot from this same population of DNA fragments was end-filled using T4 DNA polymerase, and ligated with *EcoRV*-digested, dephosphorylated pLITMUS 28 plasmid DNA. Initial cloning attempts using pUC19 as the vector, and using *SmaI* or *HindIII* as blunt cloning sites in either pUC19 or pLITMUS 28 vectors, were unsuccessful. Success required high concentrations of the 60 bp fragment DNA in the ligation reactions, and the use of pLITMUS 28 digested with *EcoRV*. Plasmid DNA isolated from ampicillin resistant colonies of *E. coli* DH5α-F', transformed with the ligated DNA, were screened by restriction digestions, since blue/white colony screening was unreliable with such small insert fragments. Of ~200 colonies that were screened, 53 contained pLITMUS 28 clones that were isolated and their insert DNA sequences were determined. After eliminating non-readable sequences
and sequences that contained deletions, 49 sequences remained, which included 44 unique sequences (Table 4). Some of these sequences were overlapping (e.g. numbers 8, 15, and 16), and some, located in different regions of the pRG101 plasmid (numbers 10a, 10b; 34a, 34b), were cloned together forming chimeric DNA fragment inserts in pLITMUS 28 (Table 4).

**Analysis of cloning data**

The size distribution of the cloned DNA fragments (Figure 36A) resembles that of the population of ~60 bp fragments that was added to the ligation reactions (Figure 35B) and can be fit with a Gaussian curve having a mean of ~60 bp. This is very similar to the pre-cloning fragment size distribution, although the post-cloning distribution is biased slightly towards clones < 60 bp in size, possibly resulting from the co-migration of smaller fragments within the larger ~60 bp band that was gel purified. A slight bias towards smaller fragments may also have been present in the pre-cloning size distribution, as the densitometry-determined background level for DNA fragment sizes < 60 bp was greater than for sizes > 60 bp (Figure 35B).

Every cloned sequence was analyzed by the AUGUR software for sequence-directed (intrinsic) DNA curvature, and the bp-ratio and $d_{\text{max}}$ values predicted for each sequence are listed in Table 4. Bending indices, which are large for sequences predicted to have substantial intrinsic curvatures, were also calculated for each cloned sequence, and are depicted in Figure 36B. Eighteen of the 44 unique sequences had a bending index $\geq 5$, which indicates a significant predicted curvature, and the remaining sequences were predicted to have very small, probably negligible intrinsic curvatures. All of the 49 cloned DNA sequences, positioned on a map of pRG101, are compared to a continuous local bending index, calculated for the entire pRG101 molecule, in Figure 37. This
Table 4. Sequences and analysis of cloned pRG101 DNA fragments, protected from MNase by rHMfA binding.

<table>
<thead>
<tr>
<th>Clone</th>
<th>Sequence</th>
<th>Size</th>
<th>bp-ratio</th>
<th>$d_{max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2&lt;sup&gt;a&lt;/sup&gt;</td>
<td>CGATGTAAACCCACTCTGTGACACCAACTGATCTTCCACATCTTTTACCTTTTACCAGCGGTTTCTGGTGAGCAAAAACAGGA</td>
<td>80</td>
<td>0.96</td>
<td>18.9</td>
</tr>
<tr>
<td>3&lt;sup&gt;b&lt;/sup&gt;</td>
<td>TCCTCAAATCCCTACCAGAATACTCTGTAGTGGTTATATAACACGATACTAATCTCAATCT</td>
<td>60</td>
<td>0.94</td>
<td>23.5</td>
</tr>
<tr>
<td>4&lt;sup&gt;c&lt;/sup&gt;</td>
<td>ATTCCTACCTGGGGTGTCATCCCCCACTATGGGTGTGTTTTTTCTGTGGT</td>
<td>63</td>
<td>0.94</td>
<td>27.4</td>
</tr>
<tr>
<td>5</td>
<td>ACTGCAGGATTTAAGGGTTGAATCCCTCAAGACTGCTAGTGGTTGTCTGTTGCGTGCGTACATGGGATGCCCT</td>
<td>75</td>
<td>0.77</td>
<td>69.7</td>
</tr>
<tr>
<td>6&lt;sup&gt;a&lt;/sup&gt;</td>
<td>ACTGGACCTGAGAGGGTATAGTAACAGGGACATCAAGCACATAGTTGGACTTGTGAAAGGGCCGTGACATGGTACAGGCAACAT</td>
<td>82</td>
<td>0.95</td>
<td>35.4</td>
</tr>
<tr>
<td>7&lt;sup&gt;a&lt;/sup&gt;</td>
<td>AAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCTTTACCTGTGGATGACCGGTGAACCTCTCGA</td>
<td>66</td>
<td>0.92</td>
<td>32.2</td>
</tr>
<tr>
<td>8&lt;sup&gt;a&lt;/sup&gt;</td>
<td>TTCTGTCTGAGAAAGCAAAAAATACAAAGAAAATCCGTGGA</td>
<td>41</td>
<td>0.93</td>
<td>21.3</td>
</tr>
<tr>
<td>9&lt;sup&gt;a&lt;/sup&gt;</td>
<td>AGGCCGAGTGGAAAGGAAATCTACGTAGGATTTGCTGATGAGATTATCAAAAAATCAGTCCCCTTACCTGTGAAGGAGGATTTTGAGGAATAGATG</td>
<td>69</td>
<td>0.89</td>
<td>38.1</td>
</tr>
<tr>
<td>10&lt;sup&gt;a&lt;/sup&gt;</td>
<td>GGGCATCTTCTAGATCGACCTGAGCCATGCAAGCTTGCC</td>
<td>41</td>
<td>0.99</td>
<td>7.2</td>
</tr>
<tr>
<td>10&lt;sup&gt;b&lt;/sup&gt;</td>
<td>CCCCCGCTACGAGTTGTAAGACGACCGCTG</td>
<td>33</td>
<td>0.95</td>
<td>14.7</td>
</tr>
<tr>
<td>11&lt;sup&gt;a&lt;/sup&gt;</td>
<td>ACCTCGCACAACCTCGGGAGGATCCGAAATAAAGGAATACACAGATGAAAACTCCACACCT</td>
<td>63</td>
<td>0.98</td>
<td>14.7</td>
</tr>
<tr>
<td>12&lt;sup&gt;a&lt;/sup&gt;</td>
<td>TGGAATAGTGTATGCGCGACCCAGATGCTTCCTTTGGCCGGCGTGCAATACGGGATAATACCCGCCACATAG</td>
<td>73</td>
<td>0.95</td>
<td>33.5</td>
</tr>
<tr>
<td>Clone</td>
<td>Sequence</td>
<td>Size</td>
<td>bp-ratio</td>
<td>$d_{max}$</td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>13&lt;sup&gt;1&lt;/sup&gt;</td>
<td>AGCACATCTCTCTCGACGGGTGCTCCTACCCACGAGGGAGGTTATTGAAACCCCATCATTC</td>
<td>65</td>
<td>0.98</td>
<td>9.1</td>
</tr>
<tr>
<td>14&lt;sup&gt;2&lt;/sup&gt;</td>
<td>AGGCCGTGACATGGTACAGGCAACATAAAAGGAGGAGGAGGGAGTTATTGTCGCTGAGGTGGCTGACATTAGGAGCTGACATTGAGTTAGGATG</td>
<td>114</td>
<td>0.98</td>
<td>18.3</td>
</tr>
<tr>
<td>15&lt;sup&gt;4&lt;/sup&gt;</td>
<td>AGCCCCAAACATCAAGAAATCACGTGGGAGCTCATGAGCACTTCATTTTGAATCCGTCT</td>
<td>61</td>
<td>0.85</td>
<td>40.9</td>
</tr>
<tr>
<td>16&lt;sup&gt;4&lt;/sup&gt;</td>
<td>CCTGACGAGCTTACCACTGACGTGGCTCAAGTGAGGCGAACCCTGCGAGGATGACATAGG</td>
<td>76</td>
<td>0.95</td>
<td>22.5</td>
</tr>
<tr>
<td>17, 18&lt;sup&gt;3&lt;/sup&gt;</td>
<td>TGATCATATCAGGCATGTACCTGCGCTGAGGTGGTATGTTGCTGGCTGCATGTACGCTAGGTTGCTT</td>
<td>69</td>
<td>0.96</td>
<td>27.5</td>
</tr>
<tr>
<td>19, 20</td>
<td>TGGTGTTGGTTGTTGATCTCGAGGCTGATCAGCTGACATTGAGCAGGCTGAGGATGAGCTCG</td>
<td>61</td>
<td>0.96</td>
<td>20.8</td>
</tr>
<tr>
<td>21, 22&lt;sup&gt;4&lt;/sup&gt;</td>
<td>TGGTAATATATCTATATCCCACTCATTGAGATTATAATTTTGGGAATTTGAGGGTTGAAACCCCTGCTT</td>
<td>67</td>
<td>0.79</td>
<td>60.3</td>
</tr>
<tr>
<td>23</td>
<td>TGGTGTCCTCAATGAATCGTCTGCGAGAGCA</td>
<td>37</td>
<td>0.98</td>
<td>8.3</td>
</tr>
<tr>
<td>24&lt;sup&gt;1&lt;/sup&gt;</td>
<td>AAGCCATAGCTGACCAGTTAATCAATGATGACAGCATCATAGTCATTACCATCATTAGGACGACGACATCGACATTGACATCATTCCATCAACGAG</td>
<td>87</td>
<td>0.92</td>
<td>44.5</td>
</tr>
<tr>
<td>25</td>
<td>TGGGGAGGAGCAACAGAGAGCGCGGTCAAGGGCGCTGAGGGGTGTTGCGGGTGCCTGGCTTAACCT</td>
<td>69</td>
<td>0.98</td>
<td>11.1</td>
</tr>
<tr>
<td>26</td>
<td>TGGACCTGTTGGTGACTCTTCCTGTTGAGGAG</td>
<td>30</td>
<td>0.98</td>
<td>7.9</td>
</tr>
<tr>
<td>27</td>
<td>TGCTACAGGCATCGTGGTGCTAGCCTGCGTCTGTTGATGCTTTGCTCATTCAGCCTCCGCTTCACGAGCA</td>
<td>67</td>
<td>0.95</td>
<td>31.0</td>
</tr>
</tbody>
</table>
Table 4 (continued)

<table>
<thead>
<tr>
<th>Clone</th>
<th>Sequence</th>
<th>Size</th>
<th>bp-ratio</th>
<th>$d_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>282</td>
<td>TGATGTCAGGTTTACAACCCCTACCTCCAACCTGACTGAGAGGTATATGTAACGACACATGACTGGGTCAGGGACATCAAGCACATAGTTGGACTTGTGAAGGCCGTGACATGGTACAGGCAACAATAAAGAGGAGGAAGGGAGGGAGTTATGCG</td>
<td>145</td>
<td>0.94</td>
<td>55.4</td>
</tr>
<tr>
<td>29</td>
<td>TCCAAATCACAAAGCCTCAAGATACCCCCTAAGGTCAGTTGAACTCTGGGATGTCATCAAGGT</td>
<td>71</td>
<td>0.94</td>
<td>32.2</td>
</tr>
<tr>
<td>302</td>
<td>TAAGAGGATCATGGAGGCTACGAGGTCACTCCAGGGCTGTCTATGAGCGATTGA</td>
<td>54</td>
<td>0.96</td>
<td>17.6</td>
</tr>
<tr>
<td>31</td>
<td>CTGTAACGGCGATCCGGGAGACGACAAAGCAGGCAGCTCGAGCTCGAGCTCGGGGGTGTGGGC</td>
<td>75</td>
<td>0.98</td>
<td>9.5</td>
</tr>
<tr>
<td>321</td>
<td>ACAAGCTACGCTGACCACTCAATTAAATGATGACAGCACATCGAACATCTCGA</td>
<td>52</td>
<td>0.97</td>
<td>13.4</td>
</tr>
<tr>
<td>33</td>
<td>CGTTCATCCATAGTGCTGTGACTCCGGCTGTGATGATATACCTGAT</td>
<td>48</td>
<td>0.98</td>
<td>8.2</td>
</tr>
<tr>
<td>34a6</td>
<td>AGTCACGACGTGTAAAACGACGCGCATG</td>
<td>32</td>
<td>0.96</td>
<td>16.7</td>
</tr>
<tr>
<td>34b6</td>
<td>GGGATCTCTCTAGAGTCAGCTAGGCAAGCGTCAAGCGTGTTTGGG</td>
<td>42</td>
<td>1.00</td>
<td>0.2</td>
</tr>
<tr>
<td>35</td>
<td>TCCAACCCGTTAGACACAGCAGTTATGCGGACTGGAACGCGCACACTGGGT</td>
<td>50</td>
<td>0.98</td>
<td>5.7</td>
</tr>
<tr>
<td>36</td>
<td>TGTTGACTGGAATTGAAAAATGCAAGGGAAAATGGTAGGCTTTTGGGAAAAAAACCATTGCTATAGTGGT</td>
<td>69</td>
<td>0.82</td>
<td>57.6</td>
</tr>
<tr>
<td>37,</td>
<td>TGAGTTATGGAAGGAATTGGAAGGTGAGGTTGCGGGAGATAGGGATATGCAGGCGGGTGGT</td>
<td>66</td>
<td>0.98</td>
<td>13.0</td>
</tr>
<tr>
<td>383</td>
<td>GGTGATCCAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>392</td>
<td>AGCACATAGTTGGAACCTGTAGAAGGAGGAGGACATGGTACAGGCAACATAAAAGGAGGA</td>
<td>69</td>
<td>0.97</td>
<td>18.4</td>
</tr>
</tbody>
</table>
Table 4 (continued)

<table>
<thead>
<tr>
<th>Clone</th>
<th>Sequence</th>
<th>Size (^7)</th>
<th>bp-ratio (^8)</th>
<th>(d_{max} (^9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>40(^1)</td>
<td>AGCACATCCTCCTCGCAGGTAGGTTGCTCCACGCCAGGAGGGATGTTATTGAACCCCATCAGT</td>
<td>64</td>
<td>0.98</td>
<td>8.6</td>
</tr>
<tr>
<td>41(^5)</td>
<td>CAGGGTTTCTGGGTAGCAGGAAACACAGAGGAAATGCCGACAAAAAGGAATAGGCCGACGGA</td>
<td>68</td>
<td>0.85</td>
<td>47.3</td>
</tr>
<tr>
<td>42</td>
<td>TGGGACAGGCTGGAAATCCCTCAAGCAGGAGATAGCATGATTTGTTGTTGACCA</td>
<td>66</td>
<td>0.97</td>
<td>10.2</td>
</tr>
<tr>
<td>43</td>
<td>TGAAAAAGCGTTGTTGAGGTTGGAATACCGAACCACAAGGAAATACCCGACACCCACCACACAGACACACA</td>
<td>68</td>
<td>0.94</td>
<td>27.9</td>
</tr>
<tr>
<td>44</td>
<td>AGGCCTTTCCCCCTGGAAGCTCCCTCGTCCGCTTCCGACCCTGCGCAACTTA</td>
<td>64</td>
<td>0.96</td>
<td>23.0</td>
</tr>
<tr>
<td>45(^3)</td>
<td>TGTGCTGGGATGGCTCTGCTCTGGATAGTTATTTTACTGATGTGGAGTTACTCAACCAG</td>
<td>52</td>
<td>0.96</td>
<td>17.0</td>
</tr>
<tr>
<td>46</td>
<td>ACATCCTGCGGAGGGTTGAGGTGTCCACCGAGGAGGGATGTTATTGAACCCCAATCA</td>
<td>60</td>
<td>0.98</td>
<td>7.7</td>
</tr>
<tr>
<td>47(^3)</td>
<td>ACGTCTGCGGGGCAAAACTACTCGGAGTCTTACCAGTGTGAGATCCAGGATGGAGATGGAGATGGAGATGGATGAT</td>
<td>74</td>
<td>0.91</td>
<td>36.5</td>
</tr>
</tbody>
</table>

\(^1-5\) Superscripts identify clones that belong to clusters 1 to 5

\(^6\) Cloned as a contiguous DNA fragment, but 'a' and 'b' were originally at different locations in the pRGlOl sequence (see Figure 37)

\(^7\) Length of cloned DNA fragment, in bp

\(^8, 9\) bp-ratio and \(d_{max}\) calculated using the AUGUR software (see text)
Figure 36

Size distribution and predicted intrinsic curvatures of the cloned ~60 bp fragments. A. Histogram of the size distribution of the DNA fragments cloned in the statistical cloning experiment. By ignoring the 30-45 bp and 110-115 bp data points, the histogram fits reasonably well to a Gaussian distribution with mean=~60 bp, σ=~10 bp (not shown). B. Histogram of the curvature predicted for each of the 44 unique sequences that were cloned, expressed in terms of a bending index. This is defined as (1 - bp-ratio) x 100, with the bp-ratio calculated using the program AUGUR (see text). Sequences having a bending index of ≥ 5 are predicted to have significant curvature.
Figure 36.

Bending index

Number of clones

Clone #

Size (bp)

0-5
5-10
10-15
15-20
20-25
25-30
30-35
35-40
40-45
45-50
50-55
55-60
60-65
65-70
70-75
75-80
80-85
85-90
90-95
95-100
100-105
105-110
110-115
115-120
120-125
Figure 37.

Summary of the statistical cloning results. Each of the 49 cloned sequences shown in Table 4 were positioned (short, thin horizontal lines) on a map of the pRG101 plasmid (long, thick horizontal line). The local bending index, calculated at a 10 bp resolution for the entire pRG101 sequence (see Methods) is shown above the pRG101 map. Cloned sequences positioned within 100 bp of each other were defined as forming a Cluster (indicated by boxes). The pRG101 plasmid consists of plasmid pME2001 from *Mb. thermoautotrophicum* strain Marburg, ligated into the *SmaI* polylinker site of pUC19 (vertical lines). Open reading frames in the pME2001 sequence (A-D) and known genes in pUC19 and the origin of replication (*lacZ, bla, ori*) are indicated. An A/T-rich region predicted to encompass the pME2001 origin of replication is indicated by the shaded box inscribed with ‘AT’ (Nölling, 1992). The ordinate axis of the bending index plot is scaled in increments of 1000 bp, the complete pRG101 sequence being 7125 bp.
Figure 37.

Bending Index

Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5

pUC19 | pME2001 | AT | Smal | pUC19
demonstrates that 28 of the 49 sequences cluster (within 100 bp of each other) in five regions, and that clusters 2, 4, and 5 correspond to regions of pRG101 sequence that are predicted to be highly curved (Figure 37). Clusters 2, 4, and 5 each contained 6 overlapping, cloned sequences, indicating strongly that these were likely to be preferred sites of rHMfA binding. None of these clusters, nor the singly cloned sequences, correspond to regions of pRG101 that are known or predicted to be promoters or terminators, except cluster 2, which spans the region predicted to be the promoter for orfB, an open reading frame in the pME2001 component of pRG101 (Nölling, 1992). Clusters 4 and 5, however, correspond to regions known or predicted to be the origins of DNA replication (ori regions) of plasmids pUC19 (Selzer et al., 1983) and pME2001 (Nölling, 1992), respectively. These sequences contain AT-rich elements of the type predicted to give rise to intrinsically curved DNA (Trifonov, 1985; Crothers et al., 1990; Bolshoy et al., 1991), and clusters 2, 4, and 5 may therefore correspond to regions of curved DNA with functional importance. Aligning the 44 unique sequences (Figure 38) did not reveal any obviously conserved primary sequence motifs, although several tri-, tetra-, and penta-nucleotide sequences were statistically over-represented. For example, the sequence (A/G)GGTG would be predicted to occur ~6 times on a random basis in the total length of sequences shown in Figure 38, but it actually occurs 17 times. In similar analyses that used DNA fragments cloned from nucleosome-protected MNase digests, Satchwell et al. (1986) reported that GC and AT dinucleotides were phased with the helical repeat and that (dA)n tracts occurred preferentially at the ends of these sequences. These sequence motifs were proposed to be involved in the rotational and translational positioning of nucleosomes, respectively (see Chapter 1). Several instances of end-localized (dA)n tracts and dinucleotide phasing are present in the sequences in Figure 38, but these are not a majority.
Figure 38.

Alignment of unique, non-overlapping cloned ‘~60 bp’ DNA sequences. Of the 44 unique sequences that were cloned, several were overlapping. The central sequence of each overlap group was aligned with the other unique, non-overlapping sequences, using the program PILEUP. The output was formatted using the program PRETTY, which generated the consensus sequence (Devereux et al., 1984). In this case, a consensus sequence clearly has limited meaning, but nevertheless, features such as the (A/G)GGTG motif, that is present at above a randomly expected occurrence level (see text), is also present in the consensus sequence. Clone numbers are indicated to the left, and base numbers above the alignment, respectively.
Evaluation of the physical properties of cloned sequences that were predicted to be intrinsically curved.

Overview of technologies employed

To evaluate the accuracy of the AUGUR curvature prediction algorithm, a subset of the 44 unique cloned sequences were analyzed by two gel electrophoresis methods that detect sequence-directed DNA curvature. The electrophoretic mobility of a DNA molecule depends on its shape, in addition to its contour length, and bent molecules migrate more slowly than linear molecules of the same contour length through polyacrylamide gels (Crothers et al., 1991). Bends reduce the electrophoretic mobility of a linear DNA fragment most if they are located near the center of the molecule, rather than located near either end.

In the first method, a circular permutation mobility assay (CPMA), the DNA sequence of interest was inserted in a vector at a unique restriction site that is flanked by direct repeats of a sequence that contains multiple restriction enzyme sites. Cleavage with different restriction enzymes therefore generated DNA fragments of identical size, but with the inserted sequence positioned at different sites relative to the ends of the molecule (Figure 42; Crothers et al., 1991). From the electrophoretic mobilities of each of the permuted fragments, calculations could be made to determine the location and approximate magnitude of the inserted bend (Kim et al., 1989; Crothers et al., 1991). This method is frequently used to investigate protein-induced bending at specific DNA-binding sites, but in principle should also be applicable to the analysis of intrinsic DNA curvature. In the second method, sequence-directed DNA curvature was detected simply by anomalous mobilities of DNA restriction fragments in PAGE analyses, when
compared to non-bent size standards. Apparent/predicted size ratios were calculated, and a ratio > 1 taken to indicate intrinsic DNA curvature.

The CPMA and comparative electrophoresis assays were used to assess the intrinsic curvature of 5 of the cloned 60 bp sequences, chosen as representatives of the 44 unique sequences analyzed. These included both isolated individual clones (# 27, 36) and sequences present within clusters (# 15, 37, 41), that were predicted either to be curved (# 15, 36, 41) or not to be curved (# 27, 37), and that ranged in length from 58 to 70 bp (see Figure 44).

Construction of pBEND2-60 plasmids and predictions of the intrinsic curvatures of these constructs

The vector used for the CPMA studies, pBEND2, contains unique XbaI and SalI cloning sites flanked by direct repeats of a 17-site polylinker sequence, with a total length of 236 bp (Kim et al., 1989; Zwieb et al., 1991). The five DNA fragments chosen for analysis were transferred from pLITMUS 28 to pBEND2 using PCR, via pUC1318. This intermediate vector was needed to generate XbaI sites flanking the ~60-bp sequence, to facilitate cloning into the pBEND2 vector (Figure 39). The transfer of intact fragments was verified by DNA sequencing of the final pBEND2-60 clones, and the sequences of the XbaI fragments from these constructions are given in Figure 40, indicating the orientation of the ~60 bp insert sequences relative to their original orientations (see Table 4).

The AUGUR software was used to predict the intrinsic curvatures of the DNA sequences (pBEND2 polylinker plus XbaI fragments) that were analyzed using the circular permutation mobility assay. Stereo images of the DNA trajectories predicted for each of the sequences are shown in Figure 41. Clones 15, 36, and 59 were predicted to
Construction of pBEND2-60 derivatives. Five ~60 bp DNA fragments (dark shaded rectangles) cloned in the *EcoRV* site of pLITMUS 28 were PCR-amplified, using #1212 and #1213 primers. The amplified products were digested with *EcoRI* and *BamHI*, end-filled with T4 DNA polymerase, and then ligated with dephosphorylated, *HincII*-digested pUC1318, which contains two *Xbal* sites flanking the central *HincII* site. PCR amplification (using the same primers) of the five subclones thus produced, followed by digestion of the amplified DNAs with *XbaI*, resulted in *XbaI* fragments, which were ligated with *XbaI*-digested, dephosphorylated pBEND2 vector to produce five pBEND2-60 clones (see text). All clones were maintained or grown in LB-media that contained 100 μg/ml ampicillin. DNA sequences of the final clones were verified by dideoxy-sequencing, using the RAG1 primer. Abbreviations are, B=BamHI, E=EcoRI, Hc/E=filled *EcoRI/HincII* hybrid site, P=PstI, RV=EcoRV half-site, X=XbaI; *bla*=β-lactamase gene of vectors.
Figure 39.
Figure 40.

Sequences of the *XbaI* inserts in the five pBEND2-60 plasmids. The *XbaI* (†) and internal *PstI* (*) sites from the vector are indicated to provide the orientation of the ~60-bp insert relative to their original orientations shown in Table 4. Numbers indicate base positions, with † symbols providing a 10 bp scale.
Figure 40.
Figure 41.

Stereo images of bends predicted for the pBEND2-60 inserts. Sequence-directed DNA curvatures were predicted by using the AUGUR program for the entire pBEND2-60 insert sequences (XbaI fragments of Figure 40 plus the remainder of the pBEND2 polylinker), and for the pBEND2 polylinker without an inserted sequence. The plane of greatest curvature is shown for each three-dimensional DNA trajectory. Sequence numbers 15, 36, and 41 are predicted to have the greatest intrinsic curvature in the planes shown.
Figure 41.
have relatively large intrinsic curvatures, whereas clones 27, 37 and the pBEND2 vector poly linker region alone were predicted to be almost linear (Figure 41). These shape predictions agree well with the predictions made for the sequences of the corresponding parent ~60 bp clones in pLITMUS 28 (Table 4; Figure 36).

**Determination of curvature by circular permutation mobility assay**

The restriction sites in the pBEND2 poly linker sequence chosen to generate the permuted sets of DNA fragments for the five clones are identified in Figure 42. These position the ~60 bp fragment-containing Xbal insert at 12 locations, relative to the ends of the poly linker sequence. The 12 DNA molecules in each set have the same size to within ± 2 bp (The size differences result from restriction site ends that differ in terms of 5' and 3' overhangs and blunt ends).

The results from the PAGE of the 12 DNA inserts from each of the five clones are shown in Figure 43. There are clearly differences between the electrophoretic mobilities of DNA fragments within each set, and between the different sets of fragments corresponding to each clone. However, the pBEND2 poly linker region alone also shows significant differences in the mobilities between the different DNA fragments within its permuted set. To accommodate these effects, differential mobilities were calculated by subtracting each individual pBEND2 DNA fragment mobility from the corresponding pBEND2-60 fragment mobility, for each of the 12 fragments in each clone set. This calculation makes the assumption that the curvatures of the poly linker and insert DNA fragments are independent, and that the effects of two (or more) curved sequences are additive. The calculated mobilities of each DNA fragment in the permuted sets are plotted against their position in the pBEND2 poly linker in Figure 43. These data are all fit well by 2nd-order polynomial functions, according to Kim et al. (1989). The minimum of a
The pBEND2 polylinker region identifying the restriction sites used for the analysis of the pBEND2-60 inserts. All the polylinker restriction sites are shown and the twelve sites chosen to generate the permuted sets of DNA fragments for circular permutation mobility assays (Figure 43) were BamHI (site 1), KpnI (site 2), SspI (site 3), NruI (site 4), Smal (site 5), EcoRV (site 6), XhoI (site 7), SpeI (site 8), Clal (site 9), NheI (site 10), BgII (site 11), and MluI (site 12). Digestion at these sites positioned the insert (darkly shaded triangle) at twelve different sites relative to the ends of the polylinker.

Figure 42.

The pBEND2 polylinker region identifying the restriction sites used for the analysis of the pBEND2-60 inserts. All the polylinker restriction sites are shown and the twelve sites chosen to generate the permuted sets of DNA fragments for circular permutation mobility assays (Figure 43) were BamHI (site 1), KpnI (site 2), SspI (site 3), NruI (site 4), Smal (site 5), EcoRV (site 6), XhoI (site 7), SpeI (site 8), Clal (site 9), NheI (site 10), BgII (site 11), and MluI (site 12). Digestion at these sites positioned the insert (darkly shaded triangle) at twelve different sites relative to the ends of the polylinker.
Figure 43.

Gel mobility analysis of the five pBEND2-60 inserts. DNAs from the five clones were each digested with the twelve enzymes indicated in Figure 42. The DNA fragments obtained were electrophoresed through 10% T, 1.3% C polyacrylamide gels, as shown to the right of the figure. The images obtained after EtBr-staining were analyzed using the NIH Image software, and the mobilities of DNA fragments in each permuted set were calculated relative to the mobility of the pBEND2 polylinker-alone fragments. These relative mobilities are plotted against the positions of the restriction sites (defined in Figure 42) in each permuted set. The solid lines are 2nd-order least squares fits to the mobility data. Lanes 1 through 12 in the gels contained DNAs digested with restriction enzymes corresponding to sites 1 through 12, as defined in Figure 42. Note that digests with Clal (site 9, gel lane 9) were unsuccessful, and therefore did not contribute to the analysis. SspI (site 3, gel lane 3) also cuts once outside the pBEND2 polylinker region, and therefore generated two bands on gels shown. M, pUC19 digested with MspI or pRG101 digested with DdeI and used as size standards.
parabolic fitted-curve in such an analysis corresponds to the site of maximum curvature
within the DNA sequence, and the depth of the minimum is a qualitative indication of the
degree of curvature (bending angle). Both the positions and relative magnitudes of
curvature predicted for the five clones (Figure 41) agree well with the experimental data
presented in Figure 43. Most importantly, the sequences predicted to have the greatest
intrinsic curvatures (# 15 and 36) did indeed behave as having the greatest curvatures
experimentally.

Determination of curvature by comparative electrophoresis

Comparative electrophoresis in 10% T, 1.3% C polyacrylamide gels was carried
out using the BamHI and EcoRV fragments of each of the five pBEND2-60 DNAs. The
BRL 123 bp ladder, pRG101 digested with DdeI, and pUC19 digested with MspI were
used as size standards (Figure 44A). The DNA fragments in the BRL 123 bp ladder have
minimal intrinsic curvature, and therefore are used as standards for comparative
electrophoresis analyses (Anderson, 1986). Apparent molecular sizes were calculated
from fragment mobilities and standard curves based on the 123 bp ladder and the
pUC19/MspI standards (Figure 44A). Apparent/actual size ratios ($R_{app}$) were thereby
calculated, and are listed in Figure 44B, along with a summary of the features of each of
the ~60 bp sequences cloned in pBEND2. Although the $R_{app}$ values for fragments
predicted to be linear (clones/sequences # 27, 37, pBEND2 polylinker) are less than 1,
indicating a greater than expected mobility, the $R_{app}$ values for fragments with predicted
large curvatures (clones/sequences # 12, 36) are ≥ 1 compared to either size standard,
indicating a less than expected mobility, and thus the presence of significant intrinsic
curvature in these DNA sequences. The intrinsic curvatures of the five cloned DNA
fragments, as indicated by their comparative electrophoresis analysis, are therefore also in
**Figure 44.**

Comparative electrophoretic analysis of the pBEND2-60 inserts. A. The five DNAs (clones # 15, 27, 36, 37, 41) were digested with either *BamHI* (B) or *EcoRV* (E), which positioned the ~60 bp inserts at the end or center of the released DNA fragment, respectively (sites 1 and 6 in Figure 42). The resulting DNA fragments were electrophoresed through a 10% T, 1.3 % C polyacrylamide gel in parallel with the BRL 123 bp ladder (1), pRG101 digested with *DdeI* (2), and pUC19 digested with *MspI* (3) size markers. Several of the *DdeI*-generated restriction fragments from pRG101 had anomalous mobilities, and therefore this digestion was not used in subsequent analyses. B. Summary of the properties of the five cloned DNAs. Apparent/actual size ratios (*R_app*) for the *BamHI* and *EcoRV* released DNA fragments were calculated using BRL 123 bp ladder and pUC19/*MspI* digest fragments to generate size versus mobility curves, based on the actual mobilities measured from the gel shown in panel A. Clone lengths (bp) in the table refer to the lengths of the original ~60 bp fragments, as listed in Table 4. Although significant curvature was not predicted for the pBEND2 polylinker sequence, both the circular permutation (Figure 43) and comparative electrophoresis analyses (shown here) indicates intrinsic curvature in this sequence, noted by the ‘???’ in the ‘Bend Predicted’ column. N/A, not applicable.
**A.**

![Gene expression analysis](image)

**B.**

<table>
<thead>
<tr>
<th>Clone #</th>
<th>Clone Size (bp)</th>
<th>Position in pRG101</th>
<th>Bend Predicted?</th>
<th>Size in pBEND2 (bp)</th>
<th>( R_{app} ) (123 bp ladder)</th>
<th>( R_{app} ) (pUC19/MspI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>61</td>
<td>Cluster 4</td>
<td>YES</td>
<td>222 (BamHI)</td>
<td>1.08</td>
<td>1.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>217 (EcoRV)</td>
<td>1.11</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>68</td>
<td>Individual, near Cluster 5</td>
<td>NO</td>
<td>227 (BamHI)</td>
<td>0.91</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>222 (EcoRV)</td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>70</td>
<td>Individual; near Cluster 2</td>
<td>YES</td>
<td>230 (BamHI)</td>
<td>0.98</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>225 (EcoRV)</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>58</td>
<td>Cluster 3</td>
<td>NO</td>
<td>226 (BamHI)</td>
<td>0.92</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>221 (EcoRV)</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>69</td>
<td>Cluster 5</td>
<td>YES</td>
<td>229 (BamHI)</td>
<td>0.94</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>224 (EcoRV)</td>
<td>0.96</td>
<td></td>
</tr>
</tbody>
</table>

**pBEND2 vector**

<table>
<thead>
<tr>
<th></th>
<th>N/A</th>
<th>N/A</th>
<th>??</th>
<th>121 (BamHI)</th>
<th>0.95</th>
<th>ND</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>121 (EcoRV)</td>
<td>0.89</td>
<td>ND</td>
</tr>
</tbody>
</table>

*Figure 44.*
good agreement with the predictions for sequence-directed curvatures obtained using the ApA wedge model in the AUGUR software.

DISCUSSION

The eukaryal histone octamer protects ~146 bp of DNA from MNase digestion, in both in vivo chromatin, and in in vitro-reconstituted nucleosomes, and the so-called '146 bp ladder' thus formed is a characteristic property of eukaryal nucleosomes (Noll, 1974a, b; Wolffe, 1992). In this respect, the HMf proteins (native HMf, rHMfA, and rHMfB) resemble eukaryal histones as the complexes that they form with DNA in vitro generate a similar ladder of bands after MNase digestion and PAGE analysis, except that the fragment lengths are multiples of ~60 bp rather than ~146 bp. The eukaryal histone (H3•H4)2 tetramer is much more stable than the octamer under a wide range of conditions (Thomas and Kornberg, 1978; Thomas, 1989; Pruss et al., 1995), and it also forms complexes with DNA in vitro, but protects only ~70 bp fragments from MNase digestion (Dong and van Holde, 1991; Puerta et al., 1993; Hansen and Wolffe, 1994); intriguingly similar to the ~60 bp size of DNA fragments protected by the HMf proteins. The HMf proteins probably bind to DNA as a tetramer (see Chapter 5), and as the eukaryal histones are larger than the HMf proteins, the (HMf)4 tetramer may be analogous, or even homologous to the (H3•H4)2 tetramer.

A problem with any random cloning approach, such as the statistical cloning carried out in this work, is that an unrecognized cloning bias would result in the samples being cloned not being representative of the starting population of DNA fragments. Although the starting population of DNA fragments and the final cloned fragments had essentially the same Gaussian size distributions, this does not necessarily indicate
representative cloning. Several fragments were cloned twice (Table 4), which could indicate a cloning bias, or be a real result, reflecting sequence preference in rHMfA binding. Statistical cloning approaches similar to that described here have been used to examine nucleosome phasing and positioning in eukaryal chromatin, and these also relied on protection from MNase to generate suitable DNA fragments for cloning and sequencing. Both Satchwell et al. (1986) and Ambrose et al. (1990) noted biases in their cloned DNA fragment populations, which they attributed to ‘non-clonability’ or ‘selection’ of some DNA sequences during the cloning step. Neither of these studies, however, incorporated a polymerase-filling step (as done here) to eliminate MNase-generated single-stranded ends that could have biased the cloning results reported by these authors.

‘Non-clonability’ problems, where certain DNA sequences are difficult (if not impossible) to clone, have been attributed to many factors (Ausubel et al., 1990), but not usually to sequences having a large intrinsic curvature. A highly curved DNA fragment presumably adopts a shape in solution consistent with a stable thermodynamic state and the ligation of such a stable, curved fragment (or even of a linear fragment with a kink in one end) might require a deformation of the fragment and/or its recipient vector that is thermodynamically unlikely under typical ligation conditions. If most DNA fragments in random populations do not have intrinsic curvatures that prevent their ligation, then a competition must exist between thermodynamically favorable and unfavorable ligations which would lead to a cloning bias against highly curved and/or inflexible DNA fragments.

The AUGUR predictions of intrinsic curvature were reasonably well supported by the experimental data, although a suitable positive control was lacking. Both the circular permutation mobility assay and comparative electrophoresis data for the five clones tested
were in good agreement with the computer predictions. Newer models and algorithms such as a nucleosome-positioning-based model (Goodsell and Dickerson, 1994) have been shown to be more accurate, particularly where phased G/C sequences are involved (which are not considered by the AUGUR algorithm), but nevertheless, AUGUR appears to have been valid for the purposes of this study.

Intrinsically curved DNA clearly helps position eukaryal nucleosomes (Thoma, 1992), and therefore, by analogy, may also position the HMf proteins on DNA. Several of the cloned DNA sequences were predicted and shown to be intrinsically curved (Table 4; Figures 43, 44), and some of these may have biological significance, since they occur in clusters (2, 4, and 5) that correspond to known or predicted promoter or ori regions. Almost half of the DNA fragments examined in the statistical cloning experiment were not however predicted to have significant intrinsic curvature by the AUGUR software. The algorithm and model used in this software probably does not detect all biologically relevant types of sequence-directed DNA curvature, and therefore some of the DNA fragments not predicted to be curved may actually be curved. The ability of a sequence to be bent (its intrinsic flexibility; Harrington, 1992; Kahn and Crothers, 1993) may be equally or even more important than it being intrinsically curved for rHMfA binding, and detection of sequence flexibility is not a feature of the analysis undertaken using the AUGUR software. Alternatively, some other property, such as di- and tri-nucleotide phasing (which may set the rotational and translational orientation of DNA wrapped around the eukaryal nucleosome) could be important in directing rHMfA binding (Drew and Travers, 1985; Satchwell et al., 1986; Shrader and Crothers, 1990).

The experimental data presented in this chapter addresses almost exclusively the sequence-determined DNA binding properties of the rHMfA protein. Parallel cloning and sequencing experiments for rHMfB, and for native HMfB (or heterodimer-containing
rHMfA/rHMfB mixtures) were beyond the scope of this work, and it must therefore be assumed that the DNA binding properties exhibited by rHMfA are similar for rHMfB and native HMf. This may however not be the case, and if different sequence 'shapes' are selected by HMfA, HMfB and HMfA•HMfB dimers then this could could be used by *Mt. fervidus* cells to regulate gene expression and/or DNA packaging, particularly when coordinated with the growth-phase regulation of the HMfA/HMfB synthesis documented in Chapter 2.
CHAPTER V
CHARACTERIZATION OF HMF-DNA COMPLEXES

INTRODUCTION

The biochemical and structural properties of the HMf proteins, and the features of DNA molecules that interact preferentially with these proteins have been documented in Chapters 1, 2 and 3. In this chapter, experiments that address the composition of HMf-DNA complexes and aspects of the DNA binding events are described and the results obtained are related to the eukaryal nucleosome.

The eukaryal nucleosome core histone octamer is a tripartite protein structure, consisting of a \((\text{H}3\text{•H}4)_2\) tetramer flanked on each side by a \(\text{H}2\text{A}•\text{H}2\text{B}\) dimer (Figure 45A), that protects \(~140-150\) bp of DNA from MNase digestion. The histone octamer is a stable unit only in the presence of \(2\) M NaCl, or when bound to DNA at physiological ionic strengths as the nucleosome core (Kornberg, 1974; Thomas and Butler, 1978; Wolffe, 1992). Under other solution conditions, the octamer dissociates into \((\text{H}3•\text{H}4)_2\) tetramers and \(\text{H}2\text{A}•\text{H}2\text{B}\) dimers (Thomas and Butler, 1978; Thomas, 1989). The \((\text{H}3•\text{H}4)_2\) tetramer is responsible for organizing the assembly of nucleosomes from \((\text{H}3•\text{H}4)_2\) tetramers and \(\text{H}2\text{A}•\text{H}2\text{B}\) dimers, and for positioning intact nucleosomes on DNA sequences (Dong and van Holde, 1991; Wolffe, 1992; Pruss et al., 1995). This tetramer protects \(~70\) bp of DNA from MNase digestion, although it apparently can wrap the same \(140-150\) bp of DNA as a complete octamer (Figure 45C; Dong and van Holde, 1991; Pruss et al., 1995). The H3 and H4 histones are much more conserved
Figure 45.

Organization of the DNA around (A) the core histone [H2A•H2B (H3•H4)2 H2A•H2B] octamer and (B) the (H3•H4)2 tetramer, shown looking down the DNA superhelical axis. In (A) the approximate arrangements of the histone pairs in the nucleosome core, in the absence of histone H1, are shown with the N- and C-termini of the monomers indicated. The second H2A•H2B pair, which lies behind the plane of the page, is not shown for clarity, although both (H2A•H2B) pairs and the (H3•H4)2 tetramer are required for complete wrapping of the DNA and protection of the full ~146 bp from MNase digestion. In (B) the (H3•H4) dimers are shown associated as the (H3•H4)2 tetramer that wraps and protects ~70 bp of DNA from MNase digestion. The N- and C-termini of H3 (lightly stippled cylinders) and H4 (darkly shaded cylinders) are indicated. Drawings were adapted from Wolffe (1994a), Pruss et al. (1995), and Wolffe (1995), from original graphics provided by D. Pruss. Cylinders represent α-helical regions.
evolutionarily than are the H2A and H2B histones, probably reflecting their very critical roles in nucleosome formation. It is therefore also likely that the contemporary H3 and H4 histones most closely resemble the ancestral histone from which all extant histones have evolved (Thatcher and Gorovsky, 1994).

The structure of the eukaryal nucleosome has been defined by using a wide variety of techniques which examined the individual proteins, the DNA component, or the complexes of protein and DNA in intact nucleosomes (Wolffe, 1992). In early studies using transmission electron microscopy (EM), chromatin fibers that had been stripped of linker histone H1 by high-salt treatment to form nucleosome cores (Thomas and Butler, 1977) gave the appearance of ‘beads on a string’ (Olins and Olins, 1974) and nuclease digestions, using DNaseI or MNase, established 140-150 bp as the size of the DNA component in these cores (Noll, 1974a, b; Noll, 1978). Chemical crosslinking studies defined the octamer and tetramer components of the cores, and were used to study the path of assembly of the histone octamer in the presence of DNA (Komberg, 1974; Komberg and Thomas, 1974; Thomas and Butler, 1978; Thomas and Kornberg, 1978; Thomas 1989; Wolffe, 1992). The conclusive evidence establishing that eight histones and 140-150 bp of DNA comprise a nucleosome was obtained from sedimentation profiles and X-ray diffraction studies, which examined the DNA and protein components of intact nucleosome cores simultaneously.

EM observations of native HMf complexed with pUC19 DNA in vitro revealed that structures were formed that visibly resemble eukaryotic nucleosomes, and they were therefore termed nucleosome-like structures (NLS; Sandman et al., 1990). The experiments described in this chapter were undertaken to investigate and define the protein and DNA composition of an intact NLS, by using techniques similar to those used to define the structure of the eukaryal nucleosome. These studies were designed to
facilitate a synthesis of the data documented in Chapters 2, 3, and 4, and therefore to define an overall structure for the HMf-based NLS.

MATERIALS AND METHODS

Chemicals and reagents

Chemicals and MNase (Sigma #N-5386) were purchased from Sigma Chemical Co. (St Louis, MO), acrylamide from United States Biochemical (USB; Cleveland, OH), and restriction enzymes from Gibco-BRL (Life Technologies, Inc.; Gaithersburg, MD). Recombinant HMfA and HMfB were purified from E. coli strains KS1183 and KS1076, respectively, unless otherwise specified, and native HMf was purified from Mt. fervidus, as described in Chapter 2.

Plasmid DNA preparation and standard molecular biological methods

All experiments except the acrylamide gel EMSA experiments (Figure 55) used EcoRI-linearized pUC19 or SmaI-linearized pRGlOl plasmid DNA, prepared as described in Chapter 4. The small DNA fragments used in the acrylamide EMSA experiments were made from large-scale preparations (Chapter 4) of pLITMUS 28-60 clones #26 and #36 (Table 4, Chapter 4). Following restriction enzyme digestion and electrophoresis through a 4% agarose gel, the appropriately sized restriction fragments were purified using a Mermaid Kit (BIO101, La Jolla, CA), as described in Chapter 4. Clone #26 was digested with BamHI plus PstI, or with BamHI plus EcoRI to generate the desired 39 and 45 bp DNA fragments, respectively. Clone #36 was digested with XbaI plus PstI, to generate the desired 86 bp DNA fragment. All standard molecular biological procedures were carried out as described in Chapter 4.
**Electrophoresis techniques**

**Protein**

Native HMf, rHMfA or rHMfB proteins, crosslinked in the presence of DNA, were resolved using the tricine-SDS-PAGE system (as described in Chapter 2), except for the samples shown in Figure 46A, which were resolved using a standard Laemmli SDS-PAGE system and a 15% T, 3% C polyacrylamide gel. Gels were fixed, stained with CBB (except for the gel shown in Figure 46A, which was stained with silver), and destained as described in Chapter 2.

**DNA-agarose gels**

MNase digested, deproteinized DNA fragments were resolved by electrophoresis through 4% (w/v) NuSieve GTG agarose (FMC BioProducts, Rockland, ME) polymerized in TAE buffer (40 mM Tris-acetate, 2 mM EDTA, pH 8.5), as described in Chapter 2. Electrophoresis was at 5 to 6 V/cm in TAE buffer. The HMf-bound DNA complexes analyzed in the EMSA studies were resolved by electrophoresis at 1 V/cm for 16-24 h, through 0.8% (w/v) agarose gels (Agarose I, Amresco, Solon, OH) that were polymerized and run in TAE buffer.

**DNA-polyacrylamide gels**

EMSA for HMf-DNA complexes formed with the 39, 45, and 86 bp DNA fragments used 8% T, 1.3% C acrylamide gels polymerized and electrophoresed at 3.5 V/cm for 2 h in TBE buffer (89 mM Tris base, 89 mM boric acid, 2 mM EDTA, pH 8).

The 10X sample buffer used for both the agarose gel electrophoresis and for PAGE of DNA samples, consisted of 20% (w/v) Ficoll 400, 0.1 M EDTA (pH 8), 0.25% bromophenol blue. Both gel types were stained with 1 μg/ml EtBr and destained as described (Chapter 4).
Electrophoretic mobility shift assay (EMSA) studies

Agarose EMSA experiments investigated rHMfA or rHMfB binding to Smal-linearized pRG101 DNA (30 ng), at 25 °C for 20 min, at different protein/DNA mass ratios, in agarose EMSA binding buffer [50 mM NaCl, 25 mM Tris-HCl (pH 8), 10 mM EDTA, 2% Ficoll 400, 0.025% bromophenol blue]. The mobilities of the complexes formed were determined by electrophoresis through 0.8% agarose gels. Polyacrylamide EMSA experiments, using the 39, 45, or 86 bp DNA fragments that contained sequences known to bind rHMfA (Chapter 4), followed incubation of rHMfA, rHMfB, or bovine serum albumin with the DNA fragments (15 or 25 ng) at different protein/DNA molar ratios in acrylamide EMSA binding buffer [80 mM KCl, 25 mM Tris-HCl (pH 8), 10 mM EDTA, 2% Ficoll 400, 0.025% bromophenol blue]. The mobilities of the complexes that formed during incubation at 25 °C for 20 min, in 500 μl plastic microfuge tubes pre-treated with Sigmacote (Sigma #SL-2) to prevent protein adsorption to the plastic, were determined by electrophoresis at 3.5 V/cm through 8% T, 1.3% C polyacrylamide gels.

Protein-DNA crosslinking

To obtain the limited crosslinking of HMf proteins to DNA desirable for SDS-PAGE analyses, 40 mM HCHO was used which resulted in ≤ 5% of the protein being crosslinked to DNA. To cross-link a larger percentage of the protein to DNA, as needed for the MNase digestion experiments, 140 mM HCHO was used in the crosslinking reactions.

SDS-PAGE analyses

Native HMf (1 μg), rHMfA or rHMfB (3 μg) was incubated with increasing amounts of EcoRI-linearized pUC19 or Smal-linearized pRG101 DNA at 25 °C for 20
min in 18μl of 50 mM TEA-HCl (pH 8), in the presence of different amounts of salts [5 or 100 mM NaCl, 100 mM KCl, or 100 mM KPi (pH 7.5)]. Crosslinking was then initiated by adding 2 μl of 15 mg/ml DMSI or 400 mM HCHO, freshly made in 100 mM TEA-HCl (pH 8), and incubation at 25 °C was continued for 40 min. Crosslinking was terminated by the addition of 2 μl of 0.5 M NH₄HCO₃, and further incubation at 25 °C for 15 min, followed by microdialysis into 100 mM NaCl, 25 mM Tris-HCl (pH 7.5). The dialyzed samples (25 μl) were mixed with 25 μl of sample buffer [8% (w/v) SDS, 20% (v/v) glycerol, 100 mM Tris-HCl, 4% (v/v) 2-mercaptoethanol, 0.01% (w/v) bromophenol blue (pH 6.8)] and incubated at 95 °C for 3 min, followed by separation of the crosslinked products by tricine-SDS PAGE.

**MNase treatment/HPLC analyses**

Crosslinking of ¹⁴C-labeled rHMfA-pRG101 complexes for analysis by HPLC employed 140 mM HCHO. Labeled rHMfA (24 μg) was incubated at 25 °C for 20 min with SmaI-linearized pRG101 (20 μg) in 45 μl of 100 mM TEA-HCl (pH 8) buffer. Crosslinking was initiated by adding 5 μl of 1.4 M HCHO [diluted freshly in 100 mM TEA-HCl (pH 8)]. After incubation at 25 °C for 1 h, the reactions were terminated by adding 5 μl of 4 M ammonium acetate (pH 7.5) and continued incubation at 25 °C for 20 min. Following buffer exchange into 50 mM Tris-acetate (pH 8.8), 1 mM CaCl₂ buffer, all protein that had not been crosslinked to DNA was removed by ultrafiltration using Microcon 50 units (50 kDa molecular weight cutoff; Amicon Inc., Beverly, MA).

**In vivo crosslinking and MNase digestion**

**Crosslinked and non-crosslinked nucleoprotein samples**

Aliquots (1 l) from cultures of *M. fervidus* grown to an OD₅₈₀ of ~1, as described in Chapter 2, were cooled to 25 °C and crosslinked with 1% HCHO for 1 h in
the original growth medium that contained 0.1% Na$_2$S, under anaerobic conditions in stoppered 1 l ‘Wheaton’ bottles. The crosslinked cells were then pelleted by centrifugation (1000 x g, 10 min), resuspended in 20 ml of 0.1 M ammonium acetate (pH 7.5) and incubated at 25 °C for 15 min. This was followed by two cycles of centrifugation and resuspension in ddH$_2$O, and a further two cycles of centrifugation and resuspension in MNase digestion buffer [50 mM Tris-acetate (pH 8.8), 1 mM CaCl$_2$]. These washed, crosslinked cell pellets [~1.5 g (wet weight)] were then resuspended in 3.5 ml of MNase digestion buffer and passed once through a French pressure cell (SLM Instruments, Urbana, IL) at 69 MPa (10,000 psi). The cell lysates obtained were digested with MNase (1.5 or 6 U) at 37 °C, and aliquots (150 µl) were removed from the reaction tube at increasing times and mixed with 6 µl of 0.5 M EDTA and 1.5 µl of 20% SDS, to terminate the MNase digestion. Following the addition of proteinase K to 25 µg/ml and incubation at 37 °C for 3.5 h, the samples were incubated at 65 °C for 6 h to hydrolyze protein-DNA crosslinks (Solomon et al., 1985), and then clarified by centrifugation (12,000 rpm, 5 min, in a microfuge). The clarified samples were extracted once with phenol/chloroform and once with chloroform, and this was followed by ethanol precipitation and washing and drying of the remaining nucleic acids, as described for small DNA fragments, in Chapter 4. The dried nucleic acid pellets were dissolved in 25 µl of 10 mM Tris-HCl (pH 8) that contained 40 µg/ml RNase A and incubated at 37 °C for 30 min, diluted in 1X sample buffer, and electrophoresed through 4% agarose gels (see above). Non-crosslinked samples were prepared in the same way, except that the HCHO crosslinking and ammonium acetate wash steps were omitted.

**Deproteinized samples**

The protein-free crosslinked and non-crosslinked MNase-digested *Mt. fervidus* DNA, needed for control experiments (Figure 50), was prepared in a similar manner to
the MNase-digested, crosslinked and non-crosslinked protein-DNA (nucleoprotein) samples. The cell lysates were, however, digested with 400 µg/ml proteinase K for 4 h, and heated at 65 °C for 6 h, followed by the phenol/chloroform and chloroform extractions, before digestion with 0.75 or 0.45 U of MNase. The continued preparation of these samples was as described above.

\[14\text{C}-\text{labelling of rHMfA and rHMfB}\]

_E. coli_ strains KS1183 and KS1076 were used to produce \(^{14}\text{C}\)-labeled rHMfA and \(^{14}\text{C}\)-labeled rHMfB, respectively. Overnight cultures (20 ml), grown at 37 °C in modified glucose-M9 mineral salts medium (Ausubel _et al._, 1990) that contained (per l); 5 mg thiamine-HCl, 50 mg ampicillin, 5 g glucose, 10 g tryptone, 5 g yeast extract, were used to inoculate 1 l of essentially the same medium in pre-warmed 4 l flasks, but that contained only 2 g tryptone and 1 g yeast extract plus 50 mM MOPS-HCl (pH 7). Cultures were shaken vigorously (230 rpm), and grown for ~4 h to an OD\(_{600}\) of ~2, at which point rHMfA and rHMfB synthesis was induced by the addition of IPTG to 0.4 mM. Following growth for a further 20 min, 250 µCi of a \(^{14}\text{C}\)-labeled amino acid mixture (ICN Biomedicals, Costa Mesa, CA) diluted in 10 ml of 0.1 M MOPS-HCl (pH 7) was added and growth was allowed to continue for a further 3.5 to 4.5 h. Approximately 50% of the labeled amino acid mixture was taken up by the cells within two hours after addition, with very little subsequent uptake, as determined by scintillation counting of culture supernatants. The cells were harvested and rHMfA and rHMfB purified as previously described (Chapter 2), except that protein purity and the extent of labeling of other (E. coli) proteins was determined by SDS-tricine-PAGE, and autoradiography of dried gels. Quantitation indicated that \(\geq 50\%\) of labeled protein in
these cells was rHMfA or rHMfB and that the purified 14C-labeled rHMfA and rHMfB protein preparations both had specific activities of 1160 dpm/µg (protein).

**Size-exclusion HPLC (SE-HPLC)**

Crosslinked, 14C-labeled rHMfA-pRG101 complexes were digested with MNase, and DNA fragments generated were purified, as described in Chapter 4, before HPLC analysis. SE-HPLC used a Beckman System Gold HPLC system (Beckman Instruments, San Ramon, CA), with a 300 x 7.8 mm ProGel-TSK G2000SW XL column (Supelco, Bellefonte, PA), and a mobile phase of 150 mM NaCl, 25 mM KPi (pH 7.4). Samples, dissolved in 50 µl of the mobile phase, were applied to the equilibrated column in a 100 µl injection loop and resolved at a flow rate of 0.7 ml/min. Resolved complexes and the individual DNA and protein components of samples were detected at 260 nm (DNA only) and using an in-line model 171 Radioisotope Detector (protein only) (Beckman Instruments, San Ramon, CA). The column was calibrated using size standards of cytochrome C (12.4 kDa), carbonic anhydrase (29 kDa), bovine serum albumin (66 kDa), and alcohol dehydrogenase (150 kDa) purchased from Sigma Chemical Co. (St. Louis, MO); and gly-tyr (0.24 kDa), myoglobin (17 kDa), Fab fragment (50 kDa), IgG (150 kDa), and β-galactosidase (465 kDa) purchased from Boehringer Mannheim Biochemicals (Indianapolis, IN). Calibration curves were constructed from size-standard data by plotting the logarithm of the molecular weights versus the retention times (not shown).

**Image analysis and curve fitting**

Contrast enhancement and background reduction, for Figure 50, employed the **NIH Image** software (NIH, Bethesda, MD). Assymptotes to protein/DNA mass ratio
titration curves, used to calculate the fractional saturation values in the agarose EMSA experiments, were determined by nonlinear least squares fitting of data points to a rectangular hyperbolic function, using the curve fitting function of Kaleidagraph (Abelbeck Software). Fractional saturation values (Y) were thereby calculated as the ratio of each experimental data point to the appropriate saturation endpoint, and these values were plotted against protein/DNA mass ratio. Data were then fit with a Hill-type ligand binding function (van Holde, 1985) of the general form \( f(x) = \frac{x^a}{b + x^a} \), using the Kaleidagraph curve fitting function. In such a fit, the value of \( a \) is an indication of the cooperativity ('sigmoidal character') of the binding reaction. A non-cooperative binding reaction has \( a=1 \), whereas cooperative binding reactions have \( a>1 \).

RESULTS

Crosslinking of HMf-DNA complexes

Initial experiments to investigate the structure of HMf-DNA complexes used chemical crosslinking with DMSI in attempts to define the oligomeric form of HMf bound to DNA (Figure 46A). Crosslinking in the presence of linear pUC19 DNA, followed by SDS-PAGE, resulted in the appearance of bands of crosslinked products, the most predominant of which had a mobility consistent with the molecular weight expected of an HMf tetramer (compare lanes 2 and 3). However, the presence of additional bands with mobilities that corresponded to crosslinked higher-order oligomers (HMf hexamers and octamers) confused the interpretation of these results. Subsequently, CD spectroscopy (Chapter 3; Grayling et al., 1995a) established that HMfA and HMfB have conformations in solution that are highly salt type and concentration dependent. As these initial protein-DNA crosslinking experiments were conducted using low salt concentrations (~5 mM
Figure 46.

Crosslinking of HMf and DNA under low-salt conditions. A. Native HMf from *Mt. fervidus* (lane 1), crosslinked in a buffer that contained ~5 mM NaCl (pH 8) with DMSI in the absence (lane 2) or presence (lane 3) of EcoRI-linearized pUC19 DNA, at a protein/DNA mass ratio of 1. The crosslinked products were resolved by SDS-PAGE (Laemmli, 1970) in parallel with size standards (S), and stained with silver (Moirisey, 1981). The location of the predominant monomer, dimer and tetramer forms are indicated. Bands indicated by ‘?’ did not have mobilities consistent with predicted oligomeric forms of native HMf. B. The products of crosslinking rHMfA in the presence of Smal-linearized pRG101 DNA, at increasing protein/DNA mass ratios (0.1, 0.5, 1, 2, 5), in buffer that contained 5 mM NaCl (pH 8), were separated by tricine-SDS-PAGE and stained with CBB. Crosslinking used either DMSI or HCHO. Control samples were crosslinked in the absence of DNA (0), or not crosslinked in the presence (+) or absence (−) of DNA. The oligomeric forms of rHMfA corresponding to a monomer (M), dimer (D), tetramer (T), and octamer (O) are indicated, and the size standards (S) used had MW of 14.4, 21.5, 31.0, and 45 kDa in panel A, and 6.2, 8.2, 10.6, 14.4, 17.0, 21.5, 31.0, 45.0, 66.2, and 97.4 kDa in panel B. The different mobilities of the rHMfA oligomers crosslinked with DMSI or with HCHO reflect the different sizes of these crosslinking reagents.
Figure 46.
NaCl), conditions under which rHMfA and rHMfB were later found to be partially unfolded, there was the possibility of crosslinking artifacts, and therefore these experiments were repeated using a range of different salts, at higher concentrations. If the salt concentration were responsible for the crosslinked oligomers observed in the initial experiments, then these results had to be reproducible, and indeed, Figure 46B shows that later crosslinking experiments with rHMfA in the presence of DNA and 5 mM NaCl generated oligomeric crosslinked products similar to those observed in Figure 46A. This effect was however, most noticeable for DMSI crosslinking. Crosslinking with HCHO under the same conditions generated very few molecules that corresponded in size to rHMfA octamers, and no crosslinked products of higher molecular weight. Under the low-salt conditions, rHMfA was crosslinked by DMSI into tetramers even in the absence of DNA (‘0’ lanes), but this was not observed at higher salt concentrations (see Figure 47). DMSI has a C₆ linker between the two reactive groups, and can crosslink lysine residues up to 11 Å apart (Ji, 1983). Assuming that rHMfA in solution exists in an extended or partially unfolded conformation in 5 mM NaCl, as indicated by the CD analyses (Chapter 3), then internal lysine residues that are not normally available for crosslinking could become exposed and crosslinked, generating tetramers and even octamers in the absence of DNA. In the presence of DNA, they could also be crosslinked to nearby lysines on DNA-bound (rHMfA)₂ dimers by DMSI. This would not, however, be expected to occur with the short HCHO crosslinking reagent. This DMSI phenomenon and difference from HCHO crosslinking should be much more pronounced at higher protein/DNA mass ratios, where the DNA-bound protein dimers should be more closely packed on the DNA, and this is what was observed (Figure 46B). Very similar results were obtained for rHMfB crosslinking under the low-salt (5 mM NaCl) conditions (not shown).
Salt-dependence of rHMfA- and rHMfB-DNA crosslinking. Crosslinking was carried out with Smal-linearized pRG101 DNA, using rHMfA (A, C, E) or rHMfB (B, D, F) with DMSI or HCHO, and in 50 mM TEA-HCl (pH 8) buffer that contained 100 mM NaCl (A, B), 100 mM KCl (C, D), or 100 mM KPi (E, F). Increasing protein/DNA mass ratios of 0.1, 0.5, 1, 2, and 5 were used, and control protein samples that were crosslinked in the absence of DNA (0), or not crosslinked in the presence (+) or absence (−) of DNA were included. Equal amounts of crosslinked protein (3 µg) were loaded in each gel lane. The arrows between panels C and D indicate the bands that correspond to the molecular sizes expected for crosslinked rHMfB monomers, dimers, tetramers, and octamers. The size standards (S) used had MW of 6.2, 8.2, 10.6, 14.4, 17.0, 21.5, 31.0, 45.0, 66.2, and 97.4 kDa.
Figure 47.
The HMf-DNA crosslinking experiments were repeated, using 100 mM NaCl, 100 mM KCl, and 100 mM KPi, for both rHMfA and rHMfB (Figure 47). In 100 mM NaCl (pH 8), both proteins were crosslinked into dimers and into tetramers, in the presence of linear pRGlOl DNA. The amount of the crosslinked tetramer increased with increasing protein/DNA mass ratios for rHMfA, as expected (Figure 47A), but the amount of tetramer apparently decreased for rHMfB (Figure 47B). This was also observed for crosslinking in 100 mM KCl (Figure 47C, D), in which case rHMfB was also crosslinked into octamers. In 100 mM KPi, rHMfA was not crosslinked into tetramers (Figure 47E), whereas rHMfB was crosslinked into tetramers in decreasing amounts with increasing protein/DNA mass ratios (Figure 47F). The lack of crosslinked rHMfA tetramer formation in 100 mM KPi could have resulted from the high concentration of potassium (~200 mM) present, as rHMfA-DNA interactions are more sensitive to the salt concentration than rHMfB-DNA interactions (data not shown).

However, rHMfB also behaved somewhat aberrantly in the 100 mM KPi buffer, forming two populations of crosslinked dimers and tetramers (doublet bands; Figure 47F) that were also observed, albeit to a lesser extent, when rHMfB was crosslinked in 100 mM KCl (Figure 47D). This was not observed for rHMfA, and probably reflects crosslinks involving the two additional lysines present in rHMfB that are not found in rHMfA.

Apparently, crosslinking via alternate sets of lysine residues, in 100 mM KCl and KPi, but not in 100 mM NaCl, generated two populations of rHMfB crosslinked products that, when unfolded in tricine-SDS-PAGE, had different mobilities.

rHMfB was also crosslinked into octamers by both DMSI and HCHO, most noticeably in 100 mM KCl (Figure 47D). As rHMfB binding to DNA is very cooperative (Sandman et al., 1994b), it is possible that rHMfB molecules bound to DNA may direct the aggregation of DNA-wrapped, protein tetramers. Aggregation might result from
either a protein-protein interaction between DNA-bound tetramers, or from the different DNA binding properties of \((rHMfB)_2\) dimers as compared to \((rHMfA)_2\) dimers (Chapter 4; Sandman et al., 1994b). Such closely-packed \((rHMfB)_4\) tetramers could then be crosslinked into octamers, possibly facilitated by the additional C-terminal lysine residue present in rHMfB relative to rHMfA. The decreased intensity of the crosslinked rHMfB tetramer band observed with increasing protein/DNA mass ratio (Figure 47B, D, F) is puzzling, and occurred at all the salt conditions tested. Possibly this reflects reduced dye staining due to more crosslinking of lysine residues, since CBB binds preferentially to lysine and arginine residues (Congdon et al., 1993), but this seems unlikely since the same effect was not observed for rHMfA (Figure 47A, C, E). Perhaps rHMfB forms larger aggregates of tetramers at a few DNA binding sites, which increasingly prevent access of the CBB dye to binding sites as these aggregates become larger with increasing protein/DNA mass ratio.

**MNase digestion of *Mt. fervidus* nucleoprotein**

The experiments described in Chapter 4 demonstrated that the HMf proteins bound to DNA in vitro protect multiples of a ~60 bp DNA fragment from MNase digestion. By analogy with the eukaryal nucleosomes, this result should be reproduced using HMf-based complexes formed in vivo. MNase protection experiments were therefore carried out on the nucleoprotein structures present in lysates of *Mt. fervidus* cells that were either crosslinked in vivo with HCHO (fixed) or not crosslinked (unfixed) before cell lysis. As shown in Figures 48A and 48B, MNase digestion of the nucleoprotein complexes from both fixed and unfixed cells yielded protected ~60 bp DNA fragments, and fragments that were multiples of ~60 bp, although the multiples of ~60 bp were more prominent in the unfixed preparations (Figure 48B). These are clearly seen in
Figure 48.

MNase digestion of fixed and unfixed *Mt. fervidus* nucleoprotein. Protein-DNA complexes in clarified cell lysates from cells that were either crosslinked with HCHO (A) or not crosslinked (B), were digested with MNase for increasing times, and aliquots were electrophoresed through 4% NuSieve-GTG agarose gels. A. Samples were digested for 0, 1, 2, 4, 6, 8, 10, 15, 30, and 45 min, using 4 U MNase/g (wet weight) lysed cells. The control sample (C) contained clarified cell lysate, but no added MNase, to detect any residual nuclease activity in the *Mt. fervidus* lysate. B. Samples were digested for 0, 1, 2, 4, 6, 8, 10, 15 and 30 min, using 1 U MNase/g (wet weight) lysed cells. The numbers to the left of the images in panels A and B indicate the lengths (in bp) of pUC19/MspI marker restriction fragments (S), and to the right of the images, the approximate lengths of the protected DNA are indicated. The cartoons suggest the structures of the HMf-DNA complexes that are proposed to have protected the DNA fragments indicated.
Figure 48.
contrast-enhanced, negative images (Figure 49) in which small amounts of a 30 bp protected fragment (as described in Chapter 3) are also detected. Figure 49 also clearly demonstrates that the protected bands are actually doublets, and that each represents two distinct populations of DNA fragments that differ in size by 10 to 20 bp. These doublet bands can also be seen in both gels in the experiments documented in Figure 48. With increasing MNase digestion, there was a progressive loss of the bands containing the larger multiples of ~60 bp in both the fixed and unfixed nucleoprotein preparations, with the upper band of each doublet disappearing first (Figures 48, 49). After 15 minutes of MNase digestion, the only substantial band that remained corresponded to ~60 bp DNA fragments (Figure 49).

The basis of the doublet banding pattern is not clear. HMfB tetramers might bind and protect slightly more DNA from MNase digestion than HMfA tetramers, perhaps as a result of the more compact structures that rHMfB apparently forms (Sandman et al., 1994b), and this could generate the doublet bands observed. In principle, HMfA and HMfB could form six different tetramers in vivo (Chapter 2), and if each protected a slightly different length of DNA, then a somewhat heterogeneous population of MNase digestion products would be expected. This population could be dominated by two forms that give rise to the predominant doublet banding pattern observed. Alternatively, the upper band of each doublet could result from additional proteins being complexed with HMf, similar to histone H1 that, when attached to the nucleosome octamer, results in a larger MNase digestion product (Noll and Kornberg, 1977; Noll, 1978).

The detection of fewer multiples of ~60 bp in digests of fixed versus unfixed *Mt. fervidus* nucleoprotein could reflect HCHO exposure inhibiting subsequent MNase activity (see Figure 50), but it is also possible that in vivo, HMf tetramers are rarely closely adjacent. The change in salt conditions that necessarily occur following cell lysis
Figure 49.

Contrast-enhanced, negative image of a timecourse of MNase digestion of unfixed *Mt. fervidus* nucleoprotein. Samples were digested with 1.5 U of MNase/g (wet weight) lysed cells for 0, 1, 2, 4, 6, 8, 10, 15, 30, and 45 min, followed by electrophoresis through a 4% agarose gel, EtBr staining, and imaging with contrast enhancement using an inverted pixel intensity lookup table, and the NIH Image software. The same control (C) and size standard (S) samples were used as in Figure 48. The numbers to the right of the image indicate the approximate length (in bp) of the DNA fragments that were protected from MNase digestion.
(~1 M K+ was diluted to ≤ 200 mM K+) could also have resulted in sliding or redistribution of the DNA-bound HMf complexes in unfixed but not in fixed preparations, giving rise to the differences observed (Figures 48, 49).

MNase has a some preference for digestion at A or T residues, as compared to G or C residues (Cockell et al., 1983; Drew, 1984), and therefore control experiments were conducted using fixed and unfixed *Mt. fervidus* nucleoprotein that was then deproteinized before digestion (Figure 50). Using a similar MNase concentration as in Figures 48 and 49, the unfixed, deproteinized DNA was digested completely in 4 min (Figure 50A). With the fixed, deproteinized DNA, the enzyme concentration was reduced and the experiment was lengthened to ensure that digestion proceeded more slowly, but also in this case, there was no indication of MNase preferentially generating ~60 bp DNA (or any other length) DNA fragments. The protection of ~60 bp and multiples of ~60 bp in the nucleoprotein complexes almost certainly therefore resulted from HMf binding to *Mt. fervidus* DNA *in vivo*, since these are very similar results to those observed for complexes assembled *in vitro* using native HMf, rHMfA, or rHMfB (Chapter 4). A direct demonstration of the HMf in these complexes would require Western blotting using anti-HMf polyclonal antibodies, however isolating these protein-DNA complexes intact proved to be very difficult, and was not pursued further. A similar experiment has been completed successfully for H Mt-containing pME2001 complexes isolated from *Mb. thermoautotrophicum* strain Marburg, although these complexes were not MNase-digested (S. Pereira, personal communication).

**Direct investigation of HMf-DNA complexes by size-exclusion HPLC**

The data presented so far resulted from experiments that investigated the protein or DNA components of HMf-DNA complexes after their separation. Approaches were
**Figure 50.**

MNase digestion of unfixed and fixed deproteinized *Mt. fervidus* genomic DNA. Nucleoprotein samples were either not crosslinked (A), or crosslinked with HCHO (B), and then deproteinized, and digested with MNase for 0, 1, 2, 4, 6, 8, 10, 15, 30, and 45 min, followed by electrophoresis as described in Figures 48, 49. In (A) the samples were digested with 0.5 U of MNase/g (wet weight) lysed cells, and in (B) the with 0.3 U of MNase/g (wet weight) lysed cells. The same controls (C) and size standards (S) were used as in Figure 48.
therefore sought to examine the DNA and protein components of intact complexes simultaneously, to elucidate their combined structure. Direct approaches to investigate nucleosomes used sedimentation analyses, X-ray diffraction, and crystallography, however these were not easily applied to the study of HMf-DNA complexes. HMfA and HMfB do not absorb UV light at 260 or 280 nm, which was used to follow the sedimentation of histones, and the isolation of stable and intact HMf-60 bp DNA complexes from *Mt. fervidus* in amounts sufficient for X-ray diffraction or crystallography analysis has not been possible. It was therefore decided to use a size-exclusion HPLC approach to analyze HMf-60 bp DNA complexes generated *in vitro*, that contained ¹⁴C-labeled rHMfA or rHMfB synthesized and purified from *E. coli*. The initial experiments focused on rHMfA-DNA complexes, for correlation with the nuclease protection data generated using rHMfA (Chapter 4).

Linearized pRG101 was HCHO-crosslinked with ¹⁴C-labeled rHMfA, digested with MNase, and the digestion products were analyzed by size-exclusion HPLC. The elution profiles of non-crosslinked, ¹⁴C-labeled rHMfA dimers, and the crosslinked MNase digestion products are compared in Figure 51. A sharp peak was observed in the MNase-digested sample, followed by a second peak with a longer retention time that extended slowly to the baseline. This probably reflected labeled material that was aberrantly retained in the radioisotope flow detector, however the molecular sizes of both crosslinked, ¹⁴C-labeled products were clearly larger (longer retention times) than that of the non-crosslinked rHMfA dimers. The DNA and protein components of the MNase-digested sample were examined individually, but simultaneously, by UV absorbance of the DNA at 260 nm and by isotopic detection of the ¹⁴C-labeled rHMf protein (Figure 52A). The elution profile of the DNA component was compared to an EtBr-stained electrophoretic separation of a sample of the same MNase-digested material (Figure 52B).
Figure 51.

Size-exclusion HPLC resolution of crosslinked rHMfA-DNA complexes. The elution profiles of MNase-digested, HCHO-crosslinked $^{14}$C-labeled rHMfA-pRG101 complexes and non-crosslinked $^{14}$C-labeled (rHMfA)$_2$ dimers are compared. The broken and solid lines indicate the $^{14}$C cpm measured for the labeled, crosslinked protein-DNA complexes and labeled protein alone, respectively. The trailing of $^{14}$C-labeled material from the second peak in the crosslinked sample was probably the result of the sample becoming inappropriately trapped in the detector flow cell (see text). The arrow indicates the peak identified as (rHMfA)$_4$-60 bp DNA complexes in Figure 52.
Figure 51.
Figure 52.

Identification of the material in the crosslinked rHMfA-DNA complexes resolved by size-exclusion HPLC. A. The elution profile of the $^{14}$C-labeled rHMfA-DNA complexes (broken line), as described in Figure 51, superimposed on the $OD_{260}$ profile of the same sample (solid line). The dotted vertical line indicates the retention time of material that eluted with sizes above the exclusion limit of the column (~150 kDa). The peaks predicted to correspond to (rHMfA)$_4$-60 bp DNA complexes and (rHMfA)$_2$-30 bp DNA complexes are indicated as T and D, respectively. The shaded area under the peaks of the $OD_{260}$ profile corresponds to the HMf-DNA complexes analyzed by electrophoresis in panel B. The peak labeled N probably contained free nucleotides and/or small oligonucleotides resulting from the MNase digestion, since its retention time corresponded to ~$V_t$ for the column. B. Agarose gel electrophoretic separation of the MNase-digested $^{14}$C-labeled rHMfA-DNA complexes. A portion of the sample used in the HPLC analysis was electrophoresed through a 4% agarose gel and stained with EtBr, to generate the negative image that was quantitated by densitometry, as shown below the gel image. The positions of the ~60 bp and ~30 bp MNase protected DNA fragments are indicated.
Figure 52.
As the same size profiles were observed for the MNase-digested complexes by both the OD\textsubscript{260} HPLC profile (shaded area under the curve in Figure 52A) and by densitometry of the EtBr-stained gel image (Figure 52B), the larger peak in the OD\textsubscript{260} HPLC profile almost certainly corresponded to the \( \sim \)60 bp DNA fragments. This larger peak also corresponded to the material with the shortest retention time in the \(^{14}\)C-isotope profile of the HPLC analysis. This peak was therefore predicted to contain (rHMfA)\textsubscript{4} tetramers bound to \( \sim \)60 bp DNA fragments, and to confirm this, the apparent molecular weight of these complexes was determined by comparison with size standards (see Material and Methods). The standard curves so constructed were, however, found to underestimate the known molecular weights of rHMfA dimers and rHMfA tetramers crosslinked \textit{in vitro} by DMSI (data not shown). Values of 13.0 and 23.3 kDa were predicted from these standard curves for rHMfA dimers and tetramers known to have MW values of 14.7 and 29.5 kDa, respectively. The MW of the complexes in the first eluting peak was estimated as 53 kDa by comparison with these calibration curves. This peak could have contained \(^{14}\)C-rHMfA-DNA complexes that were comprised of either a (rHMfA)\textsubscript{2} dimer + 30 bp DNA (MW=34.5 kDa), a (rHMfA)\textsubscript{2} dimer + 60 bp DNA (MW=54.3 kDa), a (rHMfA)\textsubscript{4} tetramer + 30 bp DNA (MW=49.3 kDa), or a (rHMfA)\textsubscript{4} tetramer + 60 bp DNA (MW=69.1 kDa). The 53 kDa size estimated for the complexes in this peak was therefore closest to that calculated for a (rHMfA)\textsubscript{2} dimer bound to a 60 bp DNA fragment, but as this is probably an underestimate, this peak most likely contained a (rHMfA)\textsubscript{4} tetramer bound to \( \sim \)60 bp of double-stranded DNA.

**Properties of the rHMfA and rHMfB DNA binding reactions**

Previous DNA binding studies had examined differences in the supercoiling and compaction properties of rHMfA and rHMfB using topological and EMSA assays,
respectively (Sandman et al., 1994b). These studies used incompletely processed rHMfA preparations, since the rHMfA N-terminal processing problem had not been recognized at that time (Sandman et al., 1995). To expand on these studies, and to determine if incomplete N-terminal processing of rHMfA might affect DNA binding, agarose gel EMSA experiments were conducted. In addition, DNA binding of rHMfA and rHMfB was investigated using restriction fragments derived from the cloned DNAs described in Chapter 4, that were each predicted to contain a single binding site for these proteins.

**Multiple site binding**

The agarose EMSA, developed by Sandman et al. (1990), was used with Smal-linearized pRG101 DNA to investigate the DNA binding and compaction properties of rHMfA purified from *E. coli* strains KS1124 [rHMfA(1124)], KS1138 [rHMfA(1138)], and KS1183 [rHMfA(1183)], and rHMfB purified from *E. coli* strain KS1076 (described in Chapter 2). In this EMSA, the binding of HMf protein results in DNA compaction, and therefore the mobility of linear DNA fragments is increased with increasing protein/DNA mass ratios (Figure 53). From the mobilities of the DNA fragments in the gels shown in Figure 53, fractional saturation (*Y*) values for the binding of each protein were calculated, and are plotted in Figure 54A. The curves shown represent least squares fits of Hill-type ligand binding equations of the form \( f(x) = \frac{x^a}{b + x^a} \) superimposed on the data points (see Materials and Methods). The values of *a* in these equations indicate the extent of cooperativity of the protein binding in these titrations, with the *a* value of 1 determined for the rHMfA proteins indicating non-cooperative binding, and the *a* value of \(~1.6\) for rHMfB indicating cooperative protein binding. Cooperative binding of rHMfB to linear pBR322 was documented previously, and the saturation level for rHMfB binding to DNA was shown to occur at higher protein/DNA mass ratios than for rHMfA.
Figure 53.

Agarose gel EMSA of rHMfA and rHMfB using Smal-digested pRG101 DNA. Increasing amounts of rHMfA purified from *E. coli* strain KS1183 (A), KS1124 (C), or KS1138 (D), or rHMfB purified from strain KS1076 (B), were incubated with 30 ng of DNA for 20 min at 25 °C, and the products were then electrophoresed through 0.8% agarose gels at 1 V/cm. The protein/DNA mass ratios used were 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1.0, 1.5, and 2.0 for rHMfA (A, C, D), and 0.1, 0.3, 0.5, 0.7, 0.9, 1.1, 1.3, 1.5, 1.7, 2.0, and 3.0 for rHMfB (B). The mobilities of protein-bound DNA fragments were related to the mobility of DNA samples not incubated with rHMfA or rHMfB (0). Digestion of pRG101 with Smal releases the pUC19 and pME2001 components as two linear DNA fragments with sizes 2686 bp and 4439 bp, respectively. Size standards (S) were generated by *Bst*EII digestion of λ DNA.
Figure 53.
Figure 54.

DNA binding curves for preparations of rHMfA and rHMfB. A. The mobilities of the DNA fragments shown in Figure 53 were used to calculate the fraction of protein bound to DNA (fractional saturation values) at each protein/DNA mass ratio, and these data points were fitted by Hill-type ligand binding equations to generate the curves shown (see text). B. The same data as in panel A, except that the data points generated using the rHMfA protein preparations purified from *E. coli* strains KS1124 and KS1138 were multiplied by (1/0.15), to correct for the fraction of rHMfA present in these protein preparations that was incompletely N-terminally processed (see text). In both panels, the rHMfA and rHMfB protein preparations used were from *E. coli* strains KS1183 (circles), KS1124 (diamonds), KS1138 (triangles), and KS1076 (squares).
Figure 54.  

fractional saturation ($Y$)  

protein/DNA mass ratio

A  

fractional saturation ($Y$)  

protein/DNA mass ratio

B
(Sandman et al., 1994b). The large differences observed in the saturation curves for the different rHMfA protein preparations, derived from the KS1124, KS1138, and KS1076 E. coli strains, was however, unexpected.

Preparations of rHMfA from E. coli strains KS1124 and KS1138 were demonstrated previously to contain only 10-20% correctly processed rHMfA, with the remainder being a mixture of formylmethionylated and methionylated rHMfA, whereas rHMfA preparations from KS1183 grown on TMP and Tdr contained >85% completely processed rHMfA (Chapter 2; Sandman et al., 1995). If only the completely processed (native) rHMfA protein was 'active' in a manner detectable by the agarose EMSA, then multiplication of the data points for the rHMfA(1124) and rHMfA(1138) titrations by a constant to correct for the inactive fractions of rHMfA in these protein preparations should result in binding curves identical to that of rHMfA(1183), and Figure 54B shows that this is indeed the case. By assuming a 15% average value for the content of completely processed protein in the rHMfA(1124) and rHMfA(1138) preparations, multiplication of these data points by \( \sqrt{0.15} \) superimposed their titration curves on the rHMfA(1183) curve. As a scalar multiplication resulted in the superposition of these curves, the differences in the different rHMfA binding curves most likely resulted from the loss of DNA binding activity, and not from changes in the DNA binding properties of these protein preparations. Clearly, these results demonstrate that in quantitative studies of the DNA binding properties of rHMfA, care must be taken to ensure that complete N-terminal processing of the recombinant protein occurs during its synthesis in E. coli. These results also implicate the N-terminus of the rHMfA molecule in DNA binding, either directly, or in the formation of the (tetrameric ?) protein-DNA complexes that result in the observable mobility increase during electrophoresis through agarose gels.
Single site binding

The cloned DNA fragments that were protected from MNase digestion by rHMfA were likely to represent single binding sites for this protein (Chapter 4), and were therefore expected to bind rHMfA as isolated DNA molecules. A polyacrylamide gel EMSA was developed to test the binding of rHMfA and rHMfB to both ~60 bp and ~30 bp cloned DNA fragments (see Materials and Methods). Restriction sites that flanked the 30 and 70 bp inserts in clones #26 and #36 were used to generate 39 bp, 45 bp and 86 bp fragments that were used in the EMSA. Binding of rHMfA and rHMfB to the 86 bp fragment was assayed at protein/DNA molar ratios from 2 to 50 (Figure 55 A, B). The 50% saturation point for protein binding, in both cases, occurred between molar ratios of 8 and 12, and at a molar ratio of 12 and above, both proteins generated an additional band of lesser electrophoretic mobility (doublet bands, Figure 55A, B). The complexes formed with rHMfB apparently were more stable than those formed by rHMfA, as they migrated as tighter bands during the gel electrophoresis.

The structures of the DNA-protein complexes formed with the 86 bp DNA fragment were predicted to be DNA-bound HMf tetramers, based on the MNase protection and crosslinking data. The additional band of greater apparent MW may therefore have contained an additional (HMf)₂ dimer or dimers recruited from solution by aggregation onto already formed tetramer complexes, to form hexamers or (more likely) octamer structures. Alternatively, the faster-migrating but shifted band could have contained DNA plus a bound dimer, with the slower-migrating band representing DNA bound by a tetramer which formed only at the higher protein/DNA molar ratios. Given that at the higher protein/DNA molar ratios some of the complexes formed remained in the gel wells (Figure 55B), the former alternative involving aggregation seems more likely. Experiments to resolve this issue using the ¹⁴C-labeled proteins were unsuccessful.
**Figure 55.**

EMSA of DNA binding by rHMfA and rHMfB using 8% T, 1.3% C polyacrylamide gels polymerized and run in 1X TBE buffer (see text). A. rHMfA, and B. rHMfB was incubated with an 86 bp restriction fragment (25 ng), originally cloned as a rHMfA-protected MNase fragment (Chapter 4), at increasing protein/DNA molar ratios of 2, 5, 8, 12, 16, 20, 50, followed by electrophoresis of the resulting protein-DNA complexes. C. rHMfA (rA) or rHMfB (rB) was incubated with either a 39 bp or 45 bp restriction fragment, also cloned as rHMfA-protected DNA fragments, at a protein/DNA molar ratio of 20, before electrophoresis. The control reactions in each gel contained either no added protein (0), or bovine serum albumin at a protein/DNA molar ratio of 20 (C). The mobility shift in lane C of panel B resulted from an accidental spill-over of the rHMfB-containing sample from the adjacent gel well. Spill-over also occurred from the size standards into lane 0 of the same gel. The arrows in panel B indicate the two shifted bands (see text). The size standards (S) used were restriction fragments obtained by MspI-digestion of pUC19 (26, 34, 67, 110, 147, 190, 242, 331, 404, 489, and 501 bp).
Figure 55.
Labeled rHMfA or rHMfB protein was allowed to bind to a known amount of the 86 bp DNA fragment, at protein/DNA molar ratios of 20 to 30, and the bands containing the complexes that formed were excised from the gel and their protein content determined by liquid scintillation counting. The specific activities of the proteins (~1160 dpm/μg) were, however, too low for reliable protein quantitation.

Binding to the 39 bp and 45 bp DNA fragments was assayed at rHMfA or rHMfB protein/DNA molar ratios of 20. In both cases, mobility shifts were observed, although rHMfB appeared to bind only ~50% as well as rHMfA to these two DNA fragments (Figure 55C). Control (C) lanes contained bovine serum albumin (Figure 55B, C), and the partial shift observed in the control lane of Figure 55B was the result of a torn gel well that allowed spillover from the adjacent rHMfB-containing well.

The protein-DNA complexes formed with the 39 bp and 45 bp DNA fragments probably contained a (rHMfA)_2 or (rHMfB)_2 dimer, as the DNA molecules in these complexes were likely to be too small for (HMf)_4 tetramers to form, based on the model proposed in Chapter 1. In this model, adjacent DNA-bound (HMf)_2 dimers associate, resulting in DNA wrapping, and the model assumes that an (HMf)_4 tetramer can only form by wrapping ~60 bp of DNA. Alternatively, however, a tetramer could form following the binding of an HMf dimer by recruitment of a second dimer from solution, and then the subsequent wrapping of the DNA molecule would be around a pre-formed (HMf)_4 tetramer.

The most likely structure(s) of HMf-DNA complexes

A reasoned argument for the structure of the minimal HMf-DNA complex can be made, based on the features known for the HMf proteins and the DNA to which they bind.
Calculations based on hydrodynamic theory

Based on hydrodynamic principles, an estimate of the equivalent spherical radius of an HMf oligomer can be made, and from the known average translation along the helical DNA axis per base pair (the rise), the number of bp required to wrap around a sphere with this radius can be calculated. Hydrodynamic theory derives the hydrated volume of a macromolecular solute as

\[ V_h = \left( \frac{M}{N_0} \right) (\bar{V}_2 + \delta_1 \bar{V}_1) \]  \hspace{1cm} (8)

where \( M \) is the molecular weight, \( N_0 \) is Avagadro’s number, \( \bar{V}_2 \) is the partial specific volume of the hydrated solute, \( \delta_1 \) is the fractional hydration of the solute, and \( \bar{V}_1 \) is the specific volume of water (taken as 1 cm\(^3\) g\(^{-1}\)) (Cantor and Schimmel, 1980). If the hydration component is eliminated in the HMf-DNA case, which is reasonable as DNA binding probably excludes most water molecules from the protein (solute) surface, then equation (8) simplifies to

\[ V = \left( \frac{M}{N_0} \right) \bar{V}_2 \]  \hspace{1cm} (9)

Since this volume refers to a spherical unit, the radius of the equivalent HMf protein sphere is given by

\[ r = \left[ \left( \frac{3}{4\pi} \right) \left( \frac{M}{N_0} \right) \bar{V}_2 \right]^{\frac{1}{3}} \]  \hspace{1cm} (10)

Calculations for an HMf tetramer

Assuming that the (HMf)\(_4\) tetramer has a partial specific volume of 0.73 cm\(^3\) g\(^{-1}\) (a typical value for globular proteins), and an HMf monomer has an average MW of 7500 Da, then equation (10) gives a value of \( r = 20.6 \) Å for an HMf tetramer and therefore a circumference for the equivalent sphere of ~129 Å. From the MNase digestion data, the
protected DNA fragment size corresponding to a minimal HMf-DNA complex is ~60 bp which, given an average rise of 3.4 Å/bp for B-DNA (Hagerman, 1988), would have a length of ~204 Å. A 60 bp DNA fragment could therefore wrap (204/129) or ~1.6 times around a spherical HMf tetramer. This is consistent with previous experimental data which indicated that the DNA molecule is wrapped ~1.5 times around the protein core in a HMf-DNA complex (Musgrave et al., 1991).

Calculations based on the structure of the rHMfB monomer

Starich et al. (1995) have determined the solution structure of the symmetrical (rHMfB)2 dimer by NMR and shown that it has approximate dimensions of 50 x 35 x 25 Å. If the tetramer formed by the interaction of two of these dimers is an approximately spherical structure, then a minimum radius for the structure would be ~25 Å. Based on this value, a 60 bp DNA fragment could wrap ~1.3 times around the circumference of a (HMf)4 structure.

Interpretation of calculations

Both of the above calculations assume a spherical DNA wrap, which is probably not the case. By analogy to the eukaryal nucleosome, it is likely that the DNA is wrapped in a helical fashion around a protein complex which is probably non-spherical (Arents et al., 1991; 1993). This will result in longer or shorter contour lengths for the wrapped DNA, depending on the shape of the protein complex and the path of the DNA molecule. Both calculations are nevertheless in reasonable agreement with the experimentally determined estimate that the DNA molecule encircles the HMf core ~1.5 times (Musgrave et al., 1991), assuming that the core is an (HMf)4 tetramer. This value is also in good agreement with the number of wraps determined for DNA-protein complexes containing...
the eukaryal (H3•H4)2 tetramer (Dong and van Holde, 1991; Puerta et al., 1993), which is probably the structure most analogous to the HMf-DNA complex (see Discussion below).

Based on the hydrodynamic calculations, an octameric protein core (r~25.9 Å) would also be possible, although in this case, 1.5 wraps of DNA would require ~72 bp of DNA which is significantly longer than the mean ~60 bp observed for MNase protected DNA fragments (Chapter 4). The hydrodynamic calculations do, however, appear to exclude the possibility of ~60 bp being wrapped around an HMf dimer (r~16.3 Å), as this would require more than two complete wraps and extreme bending of the DNA molecule. Similarly, the known dimensions and symmetry of the rHMfB dimer structure seem to exclude an octameric core, since even an aggregated, non-symmetrical structure composed of two tetramers would possess an equivalent spherical radius much greater than 25 Å.

The hydrodynamic calculations, structural dimensions of the rHMfB dimer, and experimental data are therefore most consistent with an HMf-DNA complex containing a ~60 bp DNA fragment wrapped ~1.5 times around an HMf tetramer.

DISCUSSION

The crosslinking, size-exclusion HPLC data, and theoretical arguments indicate that the most probable form of HMf when bound to DNA is a tetramer, but these data infer nothing about the binding reaction itself. Micrococcal nuclease studies demonstrated that HMf protects ~60 bp of DNA, both in vitro (Chapter 4) and in vivo in *M. fervidus*, the minimum size DNA fragment protected by HMf is ~60 bp, and size-exclusion HPLC indicated that this ~60 bp DNA fragment is most likely protected by an HMf tetramer. The
minimum HMf-DNA complex in vitro and in vivo is therefore probably a structure composed of ~60 bp of DNA wrapped around an HMf tetramer. This could be the NLS observed by electron microscopy (Sandman et al., 1990; 1994b), although this may not be the case if aggregates form, or if multiple complexes position at adjacent, 'preferred' binding sites to form the structures visibly identified as single NLS.

By analogy to the eukaryal histones, and as shown in transcription experiments (Thomm et al., 1992), HMfA and HMfB binding must influence local, and possibly also global, gene expression in vivo (Grayling et al., 1994; Sandman et al., 1994b). If, as seems likely, HMfA and HMfB dimers can bind to the same sites but with different affinities, or bind preferentially to different sites in vivo, then this would provide a mechanism for differential regulation of gene expression in Mt. fervidus cells. The EMSA titration data indicate that HMfA and HMfB homodimers have significantly different DNA binding properties and, in agreement with Sandman et al. (1994b), they demonstrate that rHMfB binding is more cooperative than rHMfA binding, and that DNA saturation occurs at much higher protein/DNA mass ratios for rHMfB than for rHMfA. The single site EMSAs show clearly that rHMfA and rHMfB can bind to the same DNA sequences, but the affinities of this binding can be very different for the two proteins (Figure 55C). In principle, six tetramers [(AA)(AA), (AA)(AB), (AA)(BB), (AB)(AB), (AB)(BB), (BB)(BB)] could exist bound to DNA in vivo. If, as the data suggest, each tetramer might also have different DNA binding and wrapping properties (Sandman et al., 1994b), then in concert with the growth phase-dependent regulation of HMfA and HMfB synthesis (Chapter 2), this could provide Mt. fervidus with a significant capacity to fine-tune local and global genome structure and gene expression.

The crosslinking data suggest that rHMfB has a tendency to form aggregated structures when bound to DNA, which probably include octamers. Although such
crosslinking experiments using bifunctional reagents are somewhat prone to 'aggregation artifacts' (Ji, 1983; Wong, 1991; Teo et al., 1995), the data obtained are consistent with previous EM studies which indicated that the size, as well as the number, of NLS increased as the protein/DNA mass ratio increased (Sandman et al., 1994b).

Furthermore, rHMfB clearly binds to DNA cooperatively and forms more stable protein-DNA complexes than rHMfA (Figures 54, 55). It is therefore possible that this cooperative binding by rHMfB is mediated through protein-protein ‘aggregation’ which brings adjacent, DNA-bound tetramers sufficiently close together to be crosslinked into octamers. Alternatively, changes in the local DNA supercoiling or DNA flexibility which might result from the binding of an initial (rHMfB)\(_2\) dimer could improve the thermodynamics for secondary rHMfB binding, which also would lead to aggregation.

Whether the observed octamer structures are actually formed as discrete structures bound to DNA, or more likely result from crosslinking of adjacent tetramers, cannot be determined unequivocally from the data so far accumulated.

If the unit HMf-DNA complex is a NLS structure with a ~60 bp DNA fragment bound and wrapped around an HMf tetramer, it would strongly resemble the eukaryal (H3•H4)\(_2\) tetramer, bound and wrapped by ~70 bp of DNA (Dong and van Holde, 1991; Hayes et al., 1991; Puerta et al., 1993). The conservation of amino acid sequences in the HMf proteins and eukaryal histones, and the structure determined for rHMfB force the conclusion that these eukaryal and archaeal proteins are evolutionarily related (Grayling et al., 1994; Starich et al., 1995) and has led to the term archaeal histones. An ancestral histone-DNA complex that existed before the divergence of the Eukarya and the Archaea was probably similar to the (HMf)\(_4\) tetramer and (H3•H4)\(_2\) tetramer-DNA complexes, although significant differences have evolved in gene structure, organization, and expression within each domain since the eukaryal-archaeal divergence (Keeling et al.,
In the *Eukarya*, genome size increased dramatically and additional regulatory functions (Wolffe, 1994a, b) may have evolved for nucleosomes by the addition of histones H2A and H2B to the \((H3+H4)_2\) 'core' tetramer. In the *Archaea*, different DNA binding and wrapping properties appear to have evolved for the individual core tetramer proteins, which in concert with a growth phase-dependent regulation of their synthesis, may have provided sufficient diversity of regulatory structures for expression of their very small (< 2 Mbp) genomes.
CHAPTER VI
GENERAL DISCUSSION

Summary

The genes that encoded the HMfA and HMfB proteins, and their homologs from a variety of methanogenic and non-methanogenic Archaea (Chapter 1), were cloned before initiation of this study and native HMf, HMt and HFo have been purified from Mt. fervidus, Mb. thermoautotrophicum, and Mb. formicicum, respectively. The amino acid sequences determined and/or predicted for these proteins were found to be evolutionarily related to those of the eukaryal core histones H2A, H2B, H3, and H4, and their secondary structures were also predicted to be similar to those of the histones. These archaeal proteins were shown to form structures that resembled nucleosomes (NLS), and to bind preferentially to intrinsically curved DNA sequences. It was therefore hypothesized that the HMf family of proteins were true prokaryotic histones, homologs of the eukaryotic histones, that should be designated as archaeal histones. During the course of this study, the three-dimensional structure of rHMfB was determined by NMR techniques, and found to be very similar to that of the eukaryal histones (Starich et al., 1995), supporting this hypothesis.

Four projects were initiated to investigate the biochemical and conformational properties of HMfA and HMfB and the extent of their homologies with the eukaryal histones. These examined their configuration in solution, their structural response to the solution ionic environment and temperature, their DNA binding preferences, and the structures (NLS) that they formed as complexes with DNA.
(1) Molecular configuration and gene regulation

HMfA and HMfB were found to be dimers in solution under a wide variety of salt, pH and temperature conditions, including those which resembled the conditions in vivo in *M. fervidus* cells. The amounts of the HMfA and HMfB proteins in vivo were shown to be growth-phase regulated (Chapter 2), and since these proteins bind and wrap DNA differently (Sandman *et al.*, 1994b), this suggested a potential for regulating local and global supercoiling, and therefore for transcription in *M. fervidus* cells. The HMf proteins were also shown to form HMfA-HMfB heterodimers, and most likely to form tetramers when bound to DNA (Chapter 5). In combination with the growth-phase regulation of their levels of synthesis, the six possible HMf tetramers, which presumably also have six different DNA binding activities, could provide a way for *M. fervidus* cells to 'fine-tune' local and global supercoiling and gene expression. In contrast, the eukaryal histones, which are octamers when bound to DNA, participate in the regulation of transcription and genome supercoiling through post-translational modification of their lysine residues in N- and C-terminal 'tail' regions that are not present in the archaeal histones.

(2) Structure and stabilities in solution

The HMfA and HMfB proteins are highly thermostable, and their predominantly α-helical secondary structures were found to be highly dependent on the salt concentration, having their most thermally and conformationally stable structures in ~1 M K+. The eukaryal histones also have conformations that are highly salt dependent, also being most stable in high salt conditions (~1 M Na+ or K+; Park and Fasman, 1987; Baxevanis *et al.*, 1991). This dependence of conformation on salt concentration for both
the eukaryal histones and the HMf proteins suggests that hydrophobic interactions are the dominant stabilizing forces in these proteins, and this was confirmed recently by NMR for the structure of rHMfB (Starich et al., 1995). The thermal stabilities of HMfA and HMfB were found to be much greater than those of the eukaryal histones. Based on studies by Karantza et al. (1995) and van Holde (1988), histones H2A, H2B, H3, and H4 have $T_m$ values for thermal denaturation of 60-80 °C in 1 M NaCl (pH 7.5), whereas in 1 M KCl (pH 7.5), rHMfA and rHMfB had $T_m$ values in this range only if 1.5 M gdm-cl was added to facilitate protein unfolding. In the presence of 1 M potassium phosphate (pH 7.5) and in the absence of gdm-cl, rHMfA and rHMfB had $T_m$ values that exceeded 80 °C (Chapter 3).

3) DNA binding preferences

The HMf proteins were found to be positioned by intrinsically curved DNA sequences, consistent with the way that curved DNAs localize nucleosomes (Chapter 4). Eukaryal nucleosomes are positioned by the $(\text{H3}\cdot\text{H4})_2$ tetramer core, which apparently is capable of recognizing DNA positioning signals in the absence of the $(\text{H2A}\cdot\text{H2B})$ dimers (Chapter 1). Positioning of HMf-based NLS is therefore probably an analogous (and possibly homologous) situation, since the HMf proteins appear to exist as tetramers when complexed with DNA. The statistical cloning and sequencing experiments (Chapter 4) did not identify a consensus DNA binding sequence that could form a specific bend responsible for NLS positioning, but did confirm the importance of sequence-directed curvature in NLS formation.
(4) The NLS

Protein-protein crosslinking (Chapter 2, 5), protein-DNA crosslinking (Chapter 5), and micrococcal nuclease digestion studies in vitro (Chapter 4) and in vivo (Chapter 5) showed that a HMf tetramer is the most likely oligomer when bound to DNA, and that such bound tetramers protect ~60 bp of DNA from MNase digestion. Based on a previous study (Musgrave et al., 1991), this 60 bp of DNA is wrapped in a positive sense ~1.5 times around the tetramer. This arrangement is similar in stoichiometry to the complexes formed by the (H3•H4)2 tetramer, which protects ~70 bp of bound DNA from MNase digestion (see Chapter 5). The NLS therefore appears to be very similar to the (H3•H4)2-DNA core component of the nucleosome, the structure which is responsible for the formation and positioning of the complete nucleosome (Chapter 1). The sense of the DNA wrap in the eukaryal (H3•H4)2-DNA structure is, however, negative in contrast to the positive wrap in the NLS. The negative sense of the DNA wrap in eukaryal nucleosomes results from a left-handed protein supercoil formed by the stacking of the adjacent core histone dimers in the octamer. By analogy, the positive wrap of DNA in the NLS formed by HMf is expected to result from a right-handed protein supercoil. Since both protein cores are formed from dimers that contain essentially the same tertiary structures (Starich et al., 1995), the symmetry of the protein dimer arrangement in the HMf-NLS must be inverted relative to that of the eukaryal histone dimers in the nucleosome.

(5) Overall conclusion

The data obtained and presented here support the hypothesis that the HMfA and HMfB proteins, and by inference all the members of the HMf protein family, are archaeal histones, the first examples of true prokaryotic histones. These proteins form tetrameric
complexes with DNA that have strikingly similar properties to those formed by the 
(H3•H4)2 tetramer. The H3 and H4 histones are the most highly conserved of the 
eukaryal histones, and therefore it is likely that the archaeal [(HMf)4] and eukaryal 
[(H3•H4)2] histone-DNA complexes have evolved from a common ancestor, a histone-
DNA complex that existed before the divergence of the Archaea and the Eukarya (see 
below).

**Concerns and future experiments**

The experiments documented in this work were not conducted using K3cDPG, 
the native intracellular salt, because this compound is not commercially available, and is 
difficult to purify in adequate amounts from *Mt. fervidus*, although the solution 
crosslinking experiments were carried out at 83 °C, the optimal growth temperature for 
this organism (Chapter 2). *In vivo* at 83 °C in the presence of 1 M K+ plus ~300 mM 
cDPG, the equilibrium between HMfA and HMfB homo- and hetero-dimers, and the 
DNA binding properties of these proteins could be very different from those reported 
here. As an example, native HMf did not bind DNA in the presence of ~1 M KC1 at 25 
°C (Chapter 5; data not shown), but Stroup and Reeve (1992) demonstrated DNA binding 
in the presence of ~1 M K+ plus ~300 mM cDPG at 81 °C. Future experiments at 83 °C 
but that also incorporate K3cDPG, either purified from *Mt. fervidus* or produced 
synthetically, are needed to address this problem. CD spectroscopy and gel 
electrophoresis could, however, be technically impossible under these conditions.

The intracellular abundance of HMf in *Mt. fervidus* cells also merits further 
consideration, in the context of the DNA binding data obtained. Stroup and Reeve (1992) 
calculated that between 25% and 50% of the *Mt. fervidus* genome could be constrained in 
NLS. However, HMf is now known to bind preferentially to curved DNA sequences,
and therefore the number of NLS formed in vivo could be limited by this localization. Their numbers and locations could also vary dramatically at different stages of growth in *Mt. fervidus* cultures, as a consequence of the growth phase dependent regulation of HMfA and HMfB synthesis.

The statistical cloning data (Chapter 4) suggest that rHMfA dimers (and HMfA and HMfB dimers, by inference) bind preferentially to, or are positioned by, intrinsically curved DNA sequences, however a consensus sequence was not identified. An 'in vitro evolution'-SELEX approach (Beutel and Gold, 1992) should identify strong HMfA and HMfB DNA binding sequences, and might determine if the curved DNA elements in such sequences differ for the different archaeal histones. In addition, the experiments documented here did not consider differences in HMfA or HMfB binding based on differences in DNA topology. Indirect evidence suggests that rHMfB might bind preferentially to negatively supercoiled DNA (Sandman et al., 1994b), and that both rHMfA and rHMfB bind preferentially to supercoiled rather that relaxed DNA molecules (Stroup and Reeve, 1992). Future experiments, using the agarose gel EMSA described in Chapter 5, should compare rHMfA and rHMfB binding to positively and negatively supercoiled, and relaxed topoisomers of the same DNA molecule directly.

Some aspects of the HMf-DNA binding model proposed in Chapter 1 (Musgrave et al., 1991) have yet to be addressed experimentally. It is clear that the HMf histones are dimers in solution (Chapter 2), that probably form tetramers when bound to DNA (Chapter 5), but this could occur by adjacent HMf dimers binding and then interacting to form the tetrameric core of the NLS (as suggested by the model; Chapter 1; Figure 4), or by a bound HMf dimer recruiting a second dimer from solution to form the tetramer. This issue may best be addressed by using site-directed mutagenesis to generate rHMfA and rHMfB proteins deficient in DNA binding and/or wrapping, and chemical
crosslinking and gel assays to characterize the products of the altered protein-DNA interactions.

It has not been established if the HMf-DNA complexes defined in vitro in this work as HMf tetramers wrapped by ~60 bp of DNA correspond to the NLS observed by EM (Sandman et al., 1990; Sandman et al., 1994b). Given the results in Chapter 5, some or all of the EM-visualized NLS could have resulted from aggregation of DNA-bound HMf tetramers. Rather than attempting to correlate in vitro data to EM data to infer the structure of the NLS in vivo, a better approach would be to directly characterize the NLS formed in vivo in *M. fervidus* cells. Following their isolation as MNase-protected, HCHO-crosslinked ~60 bp DNA-protein complexes, their protein components should be identifiable and quantifiable by SDS-PAGE.

Although the NLS formed by HMf seem to be very similar to eukaryal nucleosomes, they differ in that they wrap DNA in a positive rather than a negative sense, and further studies are needed to determine the structural basis for this difference. Since both eukaryal and archaeal histones have very similar three-dimensional structures with the same dimer symmetry, the dimer-dimer interactions that result in tetramers must be different. The exact arrangement of proteins in the NLS can only be determined by X-ray crystallographic analysis of HMf-DNA complexes, but an approximate configuration might be defined by molecular modeling.

**The HMf family of proteins as model systems to study the histone-DNA interaction and protein thermodynamics.**

The HMf histones have structures and biochemical properties similar to the eukaryal core histones, and they can be readily synthesized and purified in large amounts from *E. coli*. Formation of HMf-DNA complexes occurs spontaneously in vitro, and
does not require the complicated salt dialysis steps that are needed for reconstitution of
eukaryal nucleosomes in vitro (Dong et al., 1990). The HMf proteins therefore appear to
provide technically attractive models for investigating histone-DNA interactions. As the
HMf proteins are also small and have simple, largely α-helical structures without co-
factors, disulfide bonds, or covalent modifications, they also appear to be attractive as
models for protein structure/function/stability studies. These proteins can be denatured
by exposures to both low and high temperatures, and both of these unfolding reactions
are completely reversible. The different members of the HMf family possess different
thermal stabilities that correlate well with the optimal growth temperatures of the
organisms from which they are isolated. These proteins therefore provide an opportunity
to address the fundamental biophysical properties responsible for protein folding, the
thermodynamics of high and low temperature protein denaturation, and the relationship of
these features to the biology of the proteins’ natural environments in vivo.

**Evolutionary considerations**

Members of the HMf family have already been isolated from non-methanogenic
*Archaea*, suggesting that these proteins are probably present in a wide range of *Archaea*. Recently (Baxevanis et al., 1995), the histone fold motif has also been identified in several
non-histone bacterial and eukaryal proteins, indicating that this structural element is not
restricted to histones. The DNA packaging problem presumably existed before the
divergence of the *Bacteria* from the *Archaea* and *Eukarya* (Woese et al., 1990; Doolittle,
1995) and therefore DNA binding and compaction features that evolved to solve the
ancestral problem might still be present in all contemporary ‘structural’ DNA binding
proteins. Although there are no easily recognizable features in common in the three-
dimensional structures of histones and the different families of bacterial DNA binding
proteins discussed in Chapter 1 (Figure 2), detailed amino acid sequence alignments do
suggest a very distant, common evolutionary origin for these proteins. The architecture
of the bacterial genome and bacterial mechanisms of gene expression may now have
diverged so far from those in the Archaea and Eukarya that common structural features no
longer exist in their DNA binding proteins. Promoters, RNA polymerases and
transcription factors are conserved in the Archaea and Eukarya but not in the Bacteria
(Keeling et al., 1994) and the homology now documented in the archaeal HMf family and
eukaryal histones is consistent with conservation of gene expression and genome
organization in the Archaea and Eukarya, but not in the Bacteria.
LIST OF REFERENCES


Ninnemann, O., C. Koch, and R. Kahmann. 1992. The E. coli fis promoter is subject to stringent control and autoregulation. EMBO J. 11: 1075-1083.


