Genetic and Epigenetic Variation in Histone H1

by

Craig A. Mizzen

A thesis submitted in conformity with the requirements for the Doctoral degree of the Department of Physiology University of Toronto

© Copyright by Craig A. Mizzen, 2001
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.
Preface

This thesis describes a portion of the work I performed while registered as a PhD candidate under the supervision of Dr. Donald R. McLachlan in the Department of Physiology from September 1985 until withdrawing from the program in December 1993. Major projects in the laboratory during this interval involved characterizing alterations in the structure and function of brain chromatin in patients afflicted by Alzheimer's disease, and investigating a potential role for aluminum in causing these alterations and other pathological features of the disease. Previous publications from the laboratory had provided evidence that changes in the distribution and binding of the linker histone variant H1 occurred which correlated with altered mRNA metabolism in Alzheimer-affected brain chromatin. Because existing methods for resolving the multiple amino acid sequence variants and covalently-modified forms of H1 found in eukaryotes were cumbersome and lacked resolution, I sought to develop chromatographic procedures that would facilitate further investigations of the role of H1 variants in chromatin function and their possible involvement in Alzheimer's disease. Chicken H1 was selected as a model substrate for method development since the greatest amount of information on the structure of H1 variants was available in this species at that time. This thesis describes the development of a chromatographic system that revealed an unprecedented degree of heterogeneity in chicken H1 and analyses that elucidate the basis of this heterogeneity.

Concomittantly with Dr. McLachlan's retirement in the Spring of 1994, I left Toronto to join the laboratory of Dr. C. D. Allis to investigate substantially different aspects of chromatin structure and function as a research associate. These studies have focused on post-translational modifications of core histones in metazoans and phosphorylation of a highly divergent linker histone in Tetrahymena. There is no overlap between this later work and the work described in this thesis.

Even though the work described here was performed prior to December 1993, informed readers will appreciate that all of the findings remain novel today. A manuscript submitted recently that describes the chromatographic method and portions of the data presented in this thesis was
immediately accepted for publication. The same is true for a manuscript describing a capillary electrophoresis method for analysing H1 that I developed during the degree candidacy but have not described in this thesis. The data from molecular analyses to determine the actual structural differences between chicken H1 variants will be submitted for publication in the near future, pending completion of confirmatory analyses which I am not permitted to include here under the guidelines for theses established by the School of Graduate Studies. Publications arising from the work described in this thesis and other work I performed during the degree candidacy are listed in Appendix B.
Abstract

Histone proteins play a fundamental role in DNA-templated processes in eukaryotes. Access of regulatory factors for DNA replication, repair and gene transcription, to binding sites in chromatin is influenced by histone binding to DNA in nucleosomes, and by higher order folding of nucleosomal filaments that is mediated, in part, by histone H1. Non-allelic amino acid sequence variants of histone H1 are co-expressed in most eukaryotes, and tissue-specific patterns of H1 variant expression in metazoans have long suggested that H1 variants may be functionally distinct. However, direct support for this hypothesis is lacking, in part, due to the inability of current chromatographic procedures to adequately resolve H1 variants and post-translationally modified forms. To facilitate investigation of the functional significance of H1 heterogeneity, a novel HPLC method, cation-exchange-hydrophilic-interaction chromatography (CX-HILIC), was developed that resolved 14 major and numerous minor peaks for H1 from pooled chicken blood. This was unexpected because only six genes for H1 variants are known in this species. Electrophoretic analyses revealed that each CX-HILIC peak represented one of the six non-allelic variants of H1 described previously, but that four variants were represented by multiple peaks. Mass spectrometry revealed that H1 recovered from CX-HILIC peaks was more heterogenous than was apparent in electrophoretic analyses and that the molecular mass of most molecules was smaller than predicted by the corresponding gene sequences. This was not attributable to proteolysis during handling since CX-HILIC peaks of H5 processed in parallel contained molecules that matched the expected value or exceeded it due to covalent modification. Analyses of H1 prepared from blood of individual chickens revealed that allelic variation was responsible for much of the CX-HILIC heterogeneity of H1 from pooled blood. Analyses of peptides from pooled erythrocyte H1 and liver H1 suggested that heterogeneity also arises from limited proteolysis of N-termini of H1 in vivo. Amino acid sequencing revealed that the genes encoding two non-allelic variants have been identified incorrectly in the literature. Together, these analyses demonstrate an unexpectedly high degree of heterogeneity in avian H1 and suggest the novel hypothesis that limited proteolysis may regulate H1 function in vivo.
Acknowledgements

I thank Dr. Don McLachlan for providing the laboratory space and resources necessary to conduct the work described here. Don's enthusiasm for science was infectious and will serve to inspire me in future endeavours. At the beginning of my M.Sc., Don wisely arranged for me to train in the laboratory of Dr. Peter Lewis in the Department of Biochemistry. I will always be grateful for the expert advice I received from Peter and his students Guy Guillemette and Dae Cheung. I must also acknowledge Dr. David Pulleyblank and Walter Kahr for helpful discussions that contributed to this work.

My sincere thanks go to Drs. James Ellis, Peter Lewis, John MacDonald, David Pulleyblank and Julie Silver for serving on my examination committee and critiquing the work. I am especially grateful for the thorough and thoughtful appraisal provided by the external examiner, Dr. Jeff Hayes.

Several individuals made crucial scientific contributions to this work. Dr. Andy Alpert enthusiastically provided insightful advice on chromatography and generously donated columns on numerous occasions. Drs. Rachel Loo and Phil Andrews generously agreed to provide their expertise in mass spectrometry and microsequence analyses on a collaborative basis.

As a student colleague, Lyne Lévesque provided camaraderie and much needed advice and assistance with chicken surgery. Remarkably, the experience of working with me did not dissuade Lyne from marrying me. I am thankful for her love and encouragement that has contributed immensely to the completion of this thesis.

Finally, I thank my parents, Ruth and Cliff, for their patience and encouragement during this endeavour. More than anything else, your unwavering support and love made it all possible.
Table of Contents

Preface ................................................................. ii
Abstract ........................................................................ iv
Acknowledgements ..................................................... v
Table of Contents ........................................................ vi
List of Figures .................................................................. ix
List of Tables ................................................................... xi
List of Abbreviations ..................................................... xiii

Chapter 1. Introduction: Chromatin Structure and Function .......... 1
  1.1 Overview of eukaryotic chromatin structure ......................... 2
      1.1.1 Domain structure of core histones ............................. 4
      1.1.2 Nucleosome core particle structure ........................... 5
      1.1.3 Domain structure of linker histones ........................... 9
      1.1.4 Organization of linker histones in chromatin ............... 14
      1.1.5 Supranucleosomal organization of chromatin ............... 17
  1.2 Modulation of chromatin structure in vivo: chromatin transcription .... 20
      1.2.1 General features of transcriptionally active chromatin .......... 21
      1.2.2 Acetylation of core histones .................................. 22
      1.2.3 Other modifications of core histones ......................... 27
      1.2.4 H1 function and modification .................................. 28
      1.2.5 Chromatin remodeling activities .............................. 32
  1.3 Structural variants of histones: detection, occurrence and function ........ 33
      1.3.1 Evolutionary divergence of histone structure ............... 35
      1.3.2 Allelic and non-allelic variants of core histones ........... 37
      1.3.3 Evidence for distinct functions of core histone variants   44
      1.3.4 Allelic and non-allelic variants of linker histones .......... 46
      1.3.5 Evidence for distinct functions of linker histone variants ... 58
  1.4 Scope of the thesis .................................................. 68

Chapter 2. Materials and Methods ........................................ 72
  2.1 Isolation of chicken erythrocyte nuclei ................................ 73
  2.2 Histone extraction ................................................... 73
  2.3 Preparation of H1<sub>total</sub> and H5<sub>total</sub> from pooled blood ........ 73
  2.4 Polyacrylamide gel electrophoresis .................................. 74
  2.5 HPLC ............................................................... 74
  2.6 Circular dichroism spectroscopy ..................................... 76
  2.7 Mass spectrometry ................................................... 76
  2.8 NBS cleavage procedures ............................................ 77
  2.9 α-chymotrypsin digestion procedures and microsequencing ........ 78
  2.10 Preparation of H1 from non-erythroid chicken tissues ............ 79
  2.11 Preparation of H1 and H5 from blood of single chickens .......... 80
  2.12 Assays for post-translational modifications of H1<sub>total</sub> ........... 80

Chapter 3. Resolution of pooled erythrocyte H1<sub>total</sub> by HPLC .......... 83
  3.1 Introduction ........................................................ 84
  3.2 Results and Discussion ............................................. 85
      3.2.1 Preparation and characterization of pooled erythrocyte H1<sub>total</sub> and H5<sub>total</sub> ........ 85
      3.2.2 Resolution by RP-HPLC ....................................... 85
      3.2.3 Resolution by SE-HPLC ....................................... 94
      3.2.4 Resolution by HI-HPLC ...................................... 97
      3.2.5 Resolution by CX-HPLC ...................................... 98
3.2.6 Resolution by CX-HILIC ................................................................. 103
  a) Effect of CH3CN on H1 retention on PolyCAT A ................................ 106
  b) H1 conformation in buffers containing CH3CN and NaClO4 .............. 108
  c) Effect of PolyCAT A support pore diameter ................................... 108
  d) Optimized resolution of H1total by CX-HILIC ..................................... 112

Chapter 4. Characterization of CX-HILIC fractions of H1total from pooled erythrocytes ............................................................. 114
  4.1 Introduction .................................................................................. 115
  4.2 Results and Discussion ................................................................ 115
    4.2.1 PAGE identification of H1total CX-HILIC peaks .......................... 115
    4.2.2 Electrospray MS characterization of H1total CX-HILIC peaks ....... 121
    4.2.3 Ion spray MS characterization of selected H1total CX-HILIC peaks ... 126

Chapter 5. Characterization of CX-HILIC fractions of H5total from pooled erythrocytes ............................................................. 129
  5.1 Introduction .................................................................................. 130
  5.2 Results and Discussion ................................................................ 131
    5.2.1 Chromatographic heterogeneity of H5total .................................. 131
    5.2.2 PAGE identification of H5total CX-HILIC peaks .......................... 131
    5.2.3 Electrospray MS characterization of H5total CX-HILIC peaks ....... 134
    5.2.4 Chromatographic heterogeneity of H5 from a single chicken ...... 138

Chapter 6. Investigations of the origin of the heterogeneity of pooled erythrocyte H1total ................................................................. 141
  6.1 Introduction .................................................................................. 142
  6.2 Results and Discussion ................................................................ 143
    6.2.1 Apparent absence of common post-translational modifications .... 143
    6.2.2 CX-HILIC heterogeneity of chicken liver H1 .............................. 144
    6.2.3 PAGE identification of liver H1 CX-HILIC peaks ....................... 149
    6.2.4 Ion spray MS characterization of liver H1 CX-HILIC peaks ........ 149
    6.2.5 Chromatographic evidence of allelic polymorphism in chicken erythrocyte H1 ................................................... 153
    6.2.6 PAGE identification of erythrocyte H1 CX-HILIC peaks from a single chicken .................................................. 157
    6.2.7 Electrospray MS characterization of erythrocyte H1 CX-HILIC peaks from a single chicken ........................................ 157

Chapter 7. Evidence that the gene for H1a' has been identified incorrectly ................................................................. 162
  7.1 Introduction .................................................................................. 163
  7.2 Results and Discussion ................................................................ 164
    7.2.1 Characterization of NBS peptides from H1total CX-HILIC peaks 10, 11 and 12 ...................................................... 164
    7.2.2 Microsequence analyses of a-chymotrypsin peptides from H1total CX-HILIC peaks 10 and 11 .......................................... 170

Chapter 8. Discussion ........................................................................ 176
  8.1 Resolution of chicken H1 variants by CX-HILIC compared to other methods ................................................................. 177
  8.2 Factors in the separation of H1 variants by CX-HILIC .................... 178
  8.3 Two of the six genes encoding non-allelic variants of chicken H1 have been incorrectly identified .............................................. 180
  8.4 Characteristics of the allelic and non-allelic variants of chicken H1 resolved by CX-HILIC ........................................................ 186
  8.5 If allelic variation in chicken H1 is so extensive, why is it not better known? ................................................................. 190
  8.6 If proteolytic processing of chicken H1 is so extensive, why is it not better known? ................................................................. 192
  8.7 Potential significance of the molecular heterogeneity in chicken linker histones ................................................................. 194
Appendices ................................................................................................................ 197
  A. Supplementary data and tables for interpretation
     of mass spectroscopy data .................................................................................... 198

  B. Publications from work performed during the degree residency ..................... 204

References .................................................................................................................. 206
List of Figures

Chapter 1.

Fig. 1.1 Nucleosomal structure of chromatin ................................................................. 3
Fig. 1.2 Domain structure of the core histones ................................................................. 6
Fig. 1.3 Structure of the nucleosome core particle ............................................................. 7
Fig. 1.4 The core histone N-terminal tails in the core particle ........................................... 8
Fig. 1.5 Domain structure of metazoan linker histones ....................................................... 11
Fig. 1.6 Models for GH1/GH5 DNA binding and nucleosomal binding sites ......................... 12
Fig. 1.7 Rates of histone sequence divergence during evolution .......................................... 36
Fig. 1.8 Resolution of non-allelic variants of histones by AUT-PAGE ................................. 38
Fig. 1.9 Amino acid sequences of non-allelic variants of mouse core histones ...................... 39
Fig. 1.10 Resolution of non-allelic variants of chicken H1 by cation-exchange chromatography ................................................................. 48
Fig. 1.11 Resolution of mouse linker histones by 2-D PAGE ............................................. 53
Fig. 1.12 Amino acid sequences of mouse linker histones .................................................. 54
Fig. 1.13 Detection of allelic polymorphism in H1 by SDS-PAGE ..................................... 56

Chapter 3.

Fig. 3.1 Preparation of H1_total and H5_total by gel filtration chromatography ....................... 86
Fig. 3.2 Resolution of H1_total and H5_total by RP-HPLC on 30 nm and 100 nm diam. pore Chromegabond C18 ................................................................. 95
Fig. 3.3 Resolution of H1_total and H5_total by SE-HPLC .................................................. 99
Fig. 3.4 Resolution of H1_total by HI-HPLC on PolyPROPYL A ....................................... 100
Fig. 3.5 Resolution of H1_total by CX-HPLC on TSK SP-5PW and CM-3SW ..................... 102
Fig. 3.6 Resolution of H1_total on PolyCAT A in 0.40 and 70% CH3CN ............................ 107
Fig. 3.7 Effect of CH3CN on H1 retention by PolyCAT A ................................................. 109
Fig. 3.8 Conformation of H1 in buffers containing CH3CN and NaClO4 ............................ 110
Fig. 3.9 Effect of PolyCAT A support pore diameter on resolution of H1_total .................. 111
Fig. 3.10 Optimized resolution of H1_total by CX-HILIC .................................................. 113
Chapter 4

Fig. 4.1 Preparative scale purification of H1\textsubscript{total} by CX-HILIC ........................................ 117
Fig. 4.2 Predicted amino acid sequences of chicken H1 variants ........................................ 120

Chapter 5.

Fig. 5.1 Chromatographic heterogeneity of pooled erythrocyte H5\textsubscript{total} ......................... 132
Fig. 5.2 Chromatographic heterogeneity of H5 from a single chicken ........................................ 139

Chapter 6.

Fig. 6.1 PAGE analyses of chicken organ linker histone content ............................................. 145
Fig. 6.2 Comparison of the CX-HILIC profiles of pooled blood and liver H1 ......................... 146
Fig. 6.3 CX-HILIC and electrophoretic heterogeneity of chicken liver H1 .............................. 147
Fig. 6.4 Comparison of the CX-HILIC profiles of erythrocyte H1 from pooled blood and from a single chicken .......................................................... 154
Fig. 6.5 Identification of CX-HILIC peaks for single chicken erythrocyte H1 ....................... 155

Chapter 7.

Fig. 7.1 Cleavage of polypeptides at tyrosine by NBS .......................................................... 165
Fig. 7.2 RP-HPLC peptide maps of CX-HILIC peaks 10 and 11 after NBS cleavage ............. 166
Fig. 7.3 RP-HPLC peptide maps of CX-HILIC peaks 10 and 11 after \(\alpha\)-chymotrypsin cleavage .......................................................... 173

Fig. 7.4 Microsequence analyses of the \(\alpha\)-chymotrypsin C-terminal fragments of CX-HILIC peaks 10 and 11 .......................................................... 175

Appendix

Fig. A.1 Typical electrospray mass spectra for a CX-HILIC fraction of pooled blood H1\textsubscript{total} ... 198
Fig. A.2 Typical ion spray mass spectra for a CX-HILIC fraction of pooled blood H1\textsubscript{total} .... 199
Fig. A.3 Typical electrospray mass spectra for a CX-HILIC fraction of pooled blood H5\textsubscript{total} ... 200
List of Tables

Chapter 3
Table 3.1 RP-HPLC columns tested for the resolution of chicken H1 variants .......................... 91
Table 3.2 SEC and HIC HPLC columns tested for the resolution of chicken H1 variants .... 96
Table 3.3 CX-HPLC columns tested for the resolution of chicken H1 variants .................... 104

Chapter 4
Table 4.1 Predicted properties of chicken H1 variants .......................................................... 119
Table 4.2 Electrospray MS of pooled chicken erythrocyte H1 CX-HILIC fractions ................. 123
Table 4.3 Ion spray MS of selected pooled chicken erythrocyte H1 CX-HILIC fractions ..... 127
Table 4.4 Comparison of electrospray and ion spray MS of selected pooled chicken erythrocyte H1 CX-HILIC fractions ................................................................. 128

Chapter 5
Table 5.1 Electrospray MS of pooled chicken erythrocyte H5 CX-HILIC fractions ............... 136

Chapter 6
Table 6.1 Ion spray MS of chicken liver H1 CX-HILIC fractions ........................................ 151
Table 6.2 Comparison of selected chicken liver and erythrocyte H1 CX-HILIC fractions by ion spray MS .............................................................. 152
Table 6.3 Electrospray MS of erythrocyte H1 CX-HILIC fractions isolated from a single chicken .......................................................... 159
Table 6.4 Comparison of the masses of erythrocyte H1 CX-HILIC fractions from a single chicken with those from pooled blood ........................................ 160

Chapter 7
Table 7.1 Ion spray MS of N-bromosuccinimide fragments of pooled chicken erythrocyte H1 CX-HILIC peaks 10, 11 and 12 .................... 167
Table 7.2 Putative identification of pooled erythrocyte H1 CX-HILIC peaks 10, 11 and 12 based on the M, of NBS fragments ................. 168
Table 7.3 Ion spray MS of the C-terminal fragments of pooled erythrocyte H1 CX-HILIC peaks 10 and 11 released by α-chymotrypsin .......... 174

Chapter 8
Table 8.1 Comparisons of electrospray MS data for pooled erythrocyte H1a' and H1b after revising the identity of the corresponding genes ........ 184
Table 8.2 Existing and revised alignments of chicken H1 variant proteins with their respective genes ........................................ 185
Table 8.3 Revised comparison of N-bromosuccinimide fragments of pooled chicken erythrocyte H1 CX-HILIC peaks 10, 11 and 12 ............................... 189

Appendix

Table A1 Decrements in predicted H1 variant M_r values after N-terminal proteolysis .......... 201
Table A2 Decrements in predicted H1 variant M_r after C-terminal proteolysis .................... 203
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-D</td>
<td>two-dimensional</td>
</tr>
<tr>
<td>AU</td>
<td>acetic acid-urea</td>
</tr>
<tr>
<td>AUT</td>
<td>acetic acid-urea-Triton</td>
</tr>
<tr>
<td>CBP</td>
<td>CREB (cAMP response element binding protein) binding protein</td>
</tr>
<tr>
<td>CD</td>
<td>circular dichroism</td>
</tr>
<tr>
<td>CX-HILIC</td>
<td>cation-exchange-hydrophilic-interaction chromatography</td>
</tr>
<tr>
<td>CX-HPLC</td>
<td>cation-exchange HPLC</td>
</tr>
<tr>
<td>dH$_2$O</td>
<td>distilled/deionized water</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylene diaminetetraacetic acid</td>
</tr>
<tr>
<td>ESMS</td>
<td>electrospray ionization mass spectrometry</td>
</tr>
<tr>
<td>GH1</td>
<td>globular domain of histone H1</td>
</tr>
<tr>
<td>GH5</td>
<td>globular domain of histone H5</td>
</tr>
<tr>
<td>HAT</td>
<td>histone acetyltransferase</td>
</tr>
<tr>
<td>HI-HPLC</td>
<td>hydrophobic-interaction HPLC</td>
</tr>
<tr>
<td>HPLC</td>
<td>high performance liquid chromatography</td>
</tr>
<tr>
<td>i.m.</td>
<td>intramuscular</td>
</tr>
<tr>
<td>ISMS</td>
<td>ion spray ionization mass spectrometry</td>
</tr>
<tr>
<td>LC-MS/MS</td>
<td>liquid chromatography-tandem mass spectrometry</td>
</tr>
<tr>
<td>$M_c$</td>
<td>molecular mass</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>NBS</td>
<td>N-bromosuccinimide</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCA</td>
<td>perchloric acid</td>
</tr>
<tr>
<td>PCAF</td>
<td>p300/CBP associated factor</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethylsulfonyl fluoride</td>
</tr>
<tr>
<td>RP-HPLC</td>
<td>reverse phase HPLC</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>SE-HPLC</td>
<td>size-exclusion HPLC</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>TCA</td>
<td>trichloroacetic acid</td>
</tr>
<tr>
<td>TFA</td>
<td>trifluoroacetic acid</td>
</tr>
</tbody>
</table>
Chapter 1. Introduction: Chromatin Structure and Function
1.1 Overview of eukaryotic chromatin structure

In interphase eukaryotic cells, genomic DNA and processes that utilize genomic DNA such as gene transcription and repair and replication of DNA are confined to the nucleus (Alberts, 1989). Genomic DNA is highly folded in eukaryotes since the length of their genomes, if linearized, greatly exceeds the diameter of nuclei. For example, linear DNA molecules corresponding to the 23 chromosomes in humans are estimated to range from 1.7 to 8.5 cm in length while human nuclei are approximately 5-8 μm in diameter (Alberts, 1989). Several levels of folding of DNA are thought to contribute to this degree of compaction (Fig. 1.1). The initial level is the best understood and results from the wrapping of approximately 2 turns of DNA around an octamer containing two molecules of each of the four core histones (H2A, H2B, H3 and H4) to form the nucleosome, the basic structural unit of eukaryotic chromatin. A molecule of a fifth type of histone, known as H1 or linker histone, binds DNA at or near the dyad axis of the nucleosome where DNA strands entering and exiting the nucleosome cross, and may also bind portions of the “linker” DNA joining adjacent nucleosomes. Further compaction of DNA results from folding of nucleosomal filaments into supranucleosomal structures whose nature is poorly understood. In addition to histones, chromatin contains an equivalent mass of a diverse group of proteins known collectively as “non-histone” proteins such that the stoichiometry of DNA:histones:non-histone proteins in chromatin is roughly 1:1:1 by mass (Alberts, 1989, van Holde, 1989, Wolffe, 1998).

A combination of biochemical and biophysical approaches led to the establishment of the nucleosomal structure of chromatin and this early work has been reviewed elsewhere (van Holde, 1989). Digestion of chromatin with micrococcal nuclease was a key technique in the development of the nucleosomal model and gave rise to terminology that is adopted here. Extensive micrococcal nuclease digestion of chromatin releases the minimal repeating subunit called the nucleosome core particle containing 146 bp of DNA and two molecules each of the core histones H2A, H2B, H3 and H4. Mononucleosomes, (also called chromatosomes) containing 165-200 bp of DNA and histone H1 in addition to eight molecules of the core histones can be isolated from less extensive digests (Allan et al., 1980, Simpson, 1978). Limited digestion of chromatin with micrococcal nuclease or other nuclease releases an array of fragments ranging from mononucleosomes to
Figure 1.1. Nucleosomal structure of chromatin.

(A) Schematic model of a nucleosome in which two superhelical turns of double stranded DNA are wrapped about a protein complex formed by an octamer of the core histones (two each of H2A, H2B, H3 and H4). A molecule of the linker histone, H1, is usually associated with nucleosomes in vivo but the location of H1 relative to other nucleosomal components is currently disputed. In the model shown here, the globular domain of H1 (GH1) binds the outer surface of the nucleosome at the point where the DNA strands meet as they leave the particle and thus makes contact with three gyres of DNA (Allan et al., 1980). Other models of H1 binding are discussed in Chapter 1.1.4.

(B) Schematic model of the folding of a decondensed "10 nm" nucleosomal filament (bottom right) into a "30 nm" nucleosomal fiber (top). As discussed in the text, the existence of a regularly folded structure, such as the 30 nm fiber depicted here, in vivo is disputed. However, abundant evidence indicates that transcriptionally active segments are less folded than bulk, non-transcribing chromatin and are thought to be in an extended "10 nm" conformation as shown here. (Figure reproduced from Wolffe, 1998 with permission)
nucleosome oligomers whose maximum length depends on the conditions employed. Differences in the susceptibility of specific DNA sequences in chromation to digestion by micrococcal and other nucleases that correlate with gene transcription, DNA replication and repair, and chromosome condensation have long suggested that the structure of chromatin is altered to accommodate these processes (Alberts, 1989, van Holde, 1989, Wolffe, 1998). Elucidating the nature of these alterations and the mechanisms that bring them about is a central pursuit in current biological research.

Since the basic composition of the nucleosome was established in the late 1970’s, the existence of nucleosomal structure has been confirmed in all eukaryotes examined (reviewed in van Holde, 1989). While this suggests that many mechanisms acting to modulate chromatin structure and activity are also likely to be universal, it must be recognized that differences in genome size and complexity, and the composition of histone and non-histone proteins, exist between evolutionarily distant species that are likely to impact on specific details of nuclear function. Recent research has revealed that core histones play a central role in transcriptional regulation that is likely to be similar in all eukaryotes owing to the high degree of evolutionary conservation of core histone sequences. In contrast, linker histone sequences have diverged more extensively during evolution and less is known about the role(s) of these proteins in chromatin function. Below, after reviewing important features of chromatin structure and recent findings on mechanisms that regulate chromatin transcription, evidence that structural variants of core and linker histones have evolved to perform specialized roles in chromatin will be discussed.

1.1.1 Domain structure of core histones

The composition, solution conformation and associative behaviour of the core histones have been intensively studied and are reviewed elsewhere (Isenberg, 1979; Beaudette et al., 1981; van Holde, 1989; Karantza et al., 1995; Karantza et al., 1996). Examples of core histone amino acid sequences are presented in Figure 1.9. It suffices to say here that the four core histones, H2A, H2B, H3 and H4, are small, basic proteins that share a common domain structure shown schematically in Fig. 1.2A. Crystallographic analyses of chicken core histone octamers devoid of
DNA (Arents et al., 1991; Arents and Moudrianakis, 1993, 1995) and nucleosome core particles reconstituted from cloned DNA and recombinant histones (Luger et al., 1997) revealed the existence of a “histone-fold” domain within each type of core histone formed by the bundling of three α-helices (α1, α2 and α3) separated by short loops (L1 and L2) (Fig. 1.2A). Interactions between these motifs contribute to the assembly and stability of the histone octamer and determine the exposure of residues that interact with DNA wrapped around the octamer in the nucleosome (see below). In contrast, the N-terminal portions of the core histones could not be accounted for in the crystal structures, indicating they are disordered under the conditions employed (dashed lines in Fig. 1.2A). As discussed in section 1.2, these N-terminal tails are targeted by enzyme systems including acetyltransferases, deacetylases, kinases and phosphatases that regulate histone tail function in transcription and other processes. The histone fold domains of H3 and H4 occupy the very C-terminal portions of these molecules, but C-terminal extensions beyond the histone fold are present in histones H2A and H2B. The C-terminal extension of a specific histone H2A variant can be phosphorylated in response to DNA damage (see section 1.3).

1.1.2 Nucleosome core particle structure

Details of the organization of core histones and DNA in nucleosome core particles revealed by X-ray crystallography have been continuously refined since the initial description (Finch et al., 1977) and information from many early studies has been reviewed elsewhere (van Holde, 1989; Wolfe, 1998). Discussion here will be limited to the highest resolution structure available that was recently described by Luger et al., (1997). In agreement with earlier models, they found that the overall structure resembles a disc due to the wrapping of 146 bp of DNA around the histone octamer in 1.65 turns of a flat, left-handed superhelix (Fig. 1.3). The overall dimensions of the particle are consistent with the 10 nm nucleosomal filament visualized previously in electron microscopy (e.g. Thoma et al., 1979; reviewed in van Holde, 1989). Within the particle, the histones are organized such that an H3-H4 tetramer is flanked by two H2A-H2B dimers. The histone fold regions of each H3-H4 pair (half-tetramer) and each H2A-H2B pair interact in a “handshake” to form a crescent-shaped heterodimer that interacts with 27-28 bp of DNA
Figure 1.2. Domain structure of the core histones and histone-fold:DNA interactions.

(A) Schematic model of core histone secondary structure illustrating the common central histone-fold region. Helices (rectangles) and loops are labeled according to Luger et al., 1997). The regions interacting with DNA are indicated by solid black lines (segments equivalent to the three shown for H4 are present in each core histone but are not shown for clarity, see B and C). The portions of the N-terminal tails which were not visualized in the crystal structure are shown as dashed lines (not to scale).

(B and C) Interactions of the histone-fold regions of the H2A-H2B dimer (B) and the H3-H4 dimer (C) with nucleosomal DNA. Colors and structural elements are as in (A). (Figure reproduced from Dutnall and Ramakrishnan, 1997 with permission)
Figure 1.3.  Structure of the nucleosome core particle.

(A) The 146 bp palindromic DNA sequence used in crystallization. The dyad is indicated by the arrow and turns of the DNA helix are numbered. The approximate positions of the histones with respect to DNA is shown by the colored rectangles (colors as in Fig. 1.2). DNA phosphates contacted by histones are indicated by circles using color to indicate which histone. The colored bases indicate close proximity to an arginine inserted into the minor groove (the histone is indicated by the color).

(B) The core particle as viewed along the superhelical axis of the DNA. The histones are colored as in (A).

(C) View of the core particle perpendicular to the superhelical axis of the DNA (same as (B) but viewed from side).

(Figure reproduced from Dutnall and Ramakrishnan, 1997 with permission)
Figure 1.4. The core histone N-terminal tails in the core particle.

The histone octamer of the core particle viewed along the superhelical axis of the DNA (same view as in Fig. 1.3B but without DNA). Numbers indicate the first (and for H2A C-terminus, the last) amino acid residue observed in the crystal structure. The regions of the tails and extensions of the histone-fold that contact DNA are shaded darkly. The remainder of the histone tails are shown as dashed lines (length to scale) in arbitrary positions. Documented sites of acetylation are shown by triangles. Sites of ubiquitination are indicated (U). (Figure reproduced from Luger and Richmond, 1998 with permission)
These interactions involve equivalent structures in the histone fold of each core histone (indicated by the solid bars under α1, L1/α2 and α2/L2 in Fig. 1.2A). As shown in overall views of the structure in Fig. 1.3, the H3-H4 tetramer contacts the central turns of the DNA while the two H2A-H2B dimers contact more peripheral DNA adjacent to where DNA strands are thought to enter and exit nucleosomes of nucleosomal fibers in chromatin. Approximately 121 bp (of 146 bp total in the core particle) are organized by the histone fold regions of the histone octamer.

Additional contacts with DNA are made by elements outside of the histone folds of H2A, H2B and H3. Portions of the N-terminal domains of these histones (shaded segments in Fig. 1.4) contact DNA as they exit the particle (Figs. 1.2A and 1.4). In seven of the eight cases, these portions were observed to pass through the minor groove of the DNA to exit the particle. The portions of the tails that are external to the particle and were not visualized are indicated by dashed lines with some known sites of acetylation indicated by arrows in Fig. 1.4. These external portions of the tails can potentially interact with nucleosomal or linker DNA of the same particle to restrict transcription factor access or may interact with components of adjacent nucleosomes to contribute to higher order folding of nucleosomal filaments (Luger and Richmond, 1998). As discussed in section 1.2, acetylation and other modifications targeted to these tails are thought to modulate such interactions.

1.1.3 Domain structure of linker histones

Linker histones have diverged more than core histones during evolution and the structure of H1 in lower eukaryotes such as yeast and protozoans differs substantially from that in other eukaryotes (van Holde, 1989; Wolffe et al., 1997; see section 1.3). The majority of information available on H1 structure and function derives from studies of linker histones from metazoans and discussion here will focus primarily on these “typical” H1 proteins with features of “atypical” H1 from lower eukaryotes introduced when pertinent.
Most tissues in metazoans co-express multiple genes encoding H1 proteins with different amino acid sequences (see section 1.3) but which all share a tripartite domain structure in which a central globular domain, rich in nonpolar amino acids is flanked by flexible, basic N- and C-terminal domains rich in lysine, alanine and proline (Fig. 1.5A; see Figs. 1.12 and 4.2 for examples of H1 sequences) (van Holde, 1989; Wolffe, 1998). The globular domain is the only portion of the molecule with regular structure in physiological solution (Hartmann et al., 1977) and this represents the most highly conserved portion of the H1 molecules from different metazoan species and between different H1 variants within species (Cole, 1984, 1987; van Holde, 1989; see section 1.3). This portion alone was found to protect linker DNA from digestion with micrococcal nuclease (Allan et al., 1980), leading to a well known model in which the globular domain binds the outer surface of the nucleosome at the dyad axis of the particle, making contact with three DNA strands: the two strands entering and exiting the nucleosome and the central strand present at the dyad axis (Fig. 1.1). As discussed below and in section 1.14, this long-standing model has recently been called into question.

The crystal structure of the globular domain of H5 (GH5), an erythrocyte-specific linker histone found in birds, reptiles and fish, was solved recently (Ramakrishnan et al., 1993: reviewed in Ramakrishnan, 1994). The GH5 protein consists of a bundle of three α-helices connected to a three-stranded antiparallel β-sheet (Fig. 1.5B). The structure shares similarity with the DNA-binding domain of the bacterial catabolite gene activator protein (CAP), a prototypical helix-turn-helix DNA-binding protein but is even more similar to the DNA-binding domain of the HNF-3 (hepatocyte nuclear factor 3) family of proteins belonging to the “winged-helix” group of helix-turn-helix transcription factors that includes the fork head protein of Drosophila (Clark et al., 1993, reviewed in Patikoglou and Burley, 1997). The structure of the DNA-binding domain of HNF-3 is shown in Fig. 1.5B for comparison. The crystal structure of GH5 agrees well with the solution structure proposed from an earlier NMR study (Clore et al., 1987) and a second NMR study has
Figure 1.5. Domain structure of metazoan linker histones.

(A) Schematic model of a metazoan linker histone molecule showing the typical tripartite domain structure. The globular core domain (G), composed of approx. 80 amino acids is flanked by a short (30-40 residues), basic N-terminal domain and a longer (typically 100-110 residues), basic C-terminal domain enriched in lysine, alanine and proline.

(B) Crystal structure of the globular domain of H5 (left) compared to that of HNF-3 (right), a winged-helix variant of the helix-turn-helix class of transcription factors. Three α-helices are bundled together and flanked by a three-stranded anti-parallel β-sheet in GH5. By analogy to the mode of DNA binding by HNF-3, basic residues on the surface of helix 3 of GH5 are thought to interact with DNA phosphates (see Fig. 1.6). (Figure reproduced from Ramakrishnan, 1994 with permission)
Figure 1.6. Models for GH1/GH5 DNA binding and nucleosomal binding sites.

(A) Model for the interaction of GH5 with DNA. Helix 3 of GH5 binds in the major groove of the DNA where conserved basic residues on the surface of helix 3 can interact with DNA phosphates. A cluster of basic residues (K40, R42 and K52) on the opposite face of GH5 are thought to form a second DNA binding site enabling GH5 to bind two gyres of nucleosomal DNA as shown in the bulk model in (B) and described in the text. (Figure reproduced from Ramakrishnan, 1984 with permission)

(B) Alternative schematic models for the location of GH1/GH5 in the nucleosome. GH1/GH5 is shown as a sphere, duplex DNA is shown as a ribbon and the core histone octamer has been omitted for clarity. The ‘5S model’ proposes an asymmetric binding site of GH1 with respect to the nucleosomal dyad and inside a DNA gyre from studies performed on nucleosomes reconstituted on Xenopus 5S RNA gene DNA (Pruss et al., 1996). The current ‘bulk model’ proposes the binding of GH1/GH5 near the nucleosomal dyad at a site that is asymmetric with respect to the central DNA gyre from a variety of studies as discussed in the text. In the lower portion of the figure, models showing alternate dispositions of the DNA protected from micrococcal nuclease digestion by H1 in the bulk model are shown. (Figure reproduced from Thomas, 1999 with permission)
established that the structure of GH1 is nearly identical to that of GH5 except that GH1 contains an additional short α-helical element at the N-terminus of this domain (Cerf et al., 1993).

Site-directed mutagenesis has identified two DNA-binding sites in GH5 (Goytisolo, et al., 1996). Site I corresponds to α-helix 3 of GH5 and in Fig. 1.6A, a model of GH5 binding to DNA is shown in which the α-helix 3 of GH5 contacts the major groove, based on the structure of the homologous portion of the HNF-3:DNA co-crystal. Site II consists of a cluster of four basic residues on the opposite face of the domain relative to site I. The juxtaposition of these sites would enable one GH5 molecule to contact two DNA strands and agrees well with previous observations that H5 preferentially binds nucleosomal DNA rather than linear duplex DNA (Hayes and Wolffe, 1993) and analyses of the cooperative binding of DNA by GH5 that indicated the presence of two binding sites (Thomas et al., 1992; Draves et al., 1992). The data do not support the original model of Allan et al., (1980) in which GH5/GH1 binding at the dyad axis was assumed to involve contact with three DNA strands. Furthermore, recent data discussed in section 1.14 suggest that GH1/GH5 may actually bind at an asymmetric site rather than at the dyad, making the location of GH5/GH1 in chromatin currently a controversial topic among chromatin researchers.

The N- and C-terminal domains of metazoan H1s are required for ionic strength-induced condensation of chromatin in vitro (Thoma et al., 1983), and are multiply phosphorylated at mitosis (reviewed by Bradbury, 1992; Roth and Allis, 1992). Thus, these domains appear to play important roles in linker histone function but little is known of their mechanism of action. Although these domains are relatively extended and unstructured in solution, data from spectroscopic studies of linker histones under conditions selected to stabilize helices suggest that the C-terminal domain may assume an α-helical conformation that is interrupted at points by proline residues (Clark et al., 1988). Prolines are spaced with characteristic regularity in the C-terminal domain of H1 and it has been suggested that this punctuated helical structure may allow the C-terminal domain to "track" one of the major grooves of DNA.
Although efforts to identify an H1 homolog in yeast by biochemical means were unsuccessful, data from the yeast genome sequencing project enabled identification of a presumptive homolog in *S. cerevisiae* based on sequence homologies with GH1/GH5 (Landsman, 1996; Ushinsky et al., 1997). The HHO1 gene encodes a protein possessing two domains with significant homology to GH1/GH5 separated by a short segment enriched in lysine, alanine and proline that resembles the N- and C-terminal domains of metazoan H1s in composition. Thus, the Hho1 protein is predicted to resemble a dumb-bell with two globular domains (approx. 75 residues each) separated by a flexible basic segment approx. 42 residues in length (reviewed in Wolffe et al., 1997; Thomas, 1999). Hho1p is localized to yeast nuclei *in vivo* (Uchinsky et al., 1997), and binding of recombinant Hho1p to reconstituted dinucleosomes protects linker DNA from micrococcal nuclease digestion *in vitro* (Patterton et al., 1998), suggesting that the location and function of Hho1p in yeast chromatin may resemble that of canonical linker histones in higher eukaryotes. However, the significance of this role remains unknown. Hho1p is not essential for viability, cell growth or mating, and its deletion does not result in marked changes in basal transcription or telomeric silencing (Uchinsky et al., 1997; Patterton et al., 1998).

Macronuclei of the ciliated protozoan *Tetrahymena* contain a polypeptide which is soluble in dilute perchloric acid like metazoan linker histones, resembles the C-terminal domain of metazoan H1 in amino acid composition and some sequence elements, but which appears to lack a globular domain (Wu et al., 1986; reviewed in Wolffe et al., 1997). Despite the possibility that fundamental differences may exist in the manner in which this molecule and metazoan linker histones interact with chromatin, functional analyses have revealed a gene-specific role for this H1 in transcriptional regulation in *Tetrahymena* that may have parallels in higher eukaryotes (see section 1.2.4).

### 1.1.4 Organization of linker histones in chromatin

Under appropriate conditions, micrococcal nuclease digestion of chromatin containing H1 or GH1/GH5 releases mononucleosomes containing approx. 165 bp of DNA whereas equivalent digestion of chromatin that has been depleted of H1 (e.g. by salt extraction) releases core particles
containing 146 bp of DNA (Allan et al., 1980; reviewed in van Holde, 1989). Based on the early model of GH1 binding shown in Fig. 1.1, this protection of 20 bp of DNA was assumed to result from the symmetrical protection of 10 bp of DNA in the strands entering and exiting the nucleosome due to contacts made with GH1 bound at the dyad axis. This model was supported by the observation that DNA in dinucleosomes containing H5 was protected from DNase I digestion relative to DNA in core particles with protection at the dyad axis being particularly apparent (Staynov and Crane-Robinson, 1988). However, recent analyses employing positioned nucleosomes have found that the nuclease protection afforded by H1 and GH1 was asymmetric, i.e. that only one DNA strand was protected upon linker histone binding (e.g. Wong et al., 1997; An et al., 1998a), suggesting that GH1 was not bound at the dyad as depicted in Fig. 1.1. Two models for asymmetric GH1 binding that differ substantially in the location of GH1 have been proposed recently and are discussed below.

Using a DNA fragment containing a Xenopus 5S RNA gene to reconstitute positioned mononucleosomes (referred to here as 5S nucleosomes). Hayes and Wolffe (1993) found that preferential binding of H1/H5 to 5S nucleosomes (relative to free DNA) required the presence of linker DNA on either side of the nucleosome core (i.e. symmetric distribution of the linker DNA was not required for binding). H1/H5 binding was found to confer protection of an additional 20 bp of linker DNA but mapping analyses revealed that this protection was asymmetrically distributed with about 15 bp and 5 bp of DNA protected at the opposite edges of the nucleosome, respectively. Together with their finding that H1/H5 binding did not alter the pattern of nucleosomal DNA cleavage by DNase I or hydroxyl radical, the data suggested the possibility for an asymmetric binding site for GH1/GH5. Subsequent work by this group using histone-DNA cross-linking and site-directed DNA cleavage by linker histone-Fe(II)EDTA conjugates revealed a single major contact between the globular domain of H5 and 5S nucleosomal DNA that was more than 65 bp removed from the nucleosomal dyad (Hayes et al., 1994; Pruss et al., 1996; Hayes, 1996). These latter results imply that binding site of the globular domain is most likely to be inside the DNA gyres where the major groove faces toward the octamer as shown in Fig. 1.6B. The model is consistent with earlier data on protein-protein interactions between linker histones...
H2A and other features of nucleosomal structure as reviewed in Pruss et al., (1995). However, several discrepancies exist between the model and established data and concerns have been raised regarding the reliability of some of the techniques employed. Perhaps most puzzling is that the model does not account for the protection of an additional 20 bp of DNA from micrococcal nuclease digestion upon linker histone binding. This discrepancy may be explained by data from protein-DNA cross-linking studies indicating that differences exist in the contacts of H2A and H3 with DNA in the presence and absence of linker histones (Hayes et al., 1994; Lee and Hayes, 1998; Guschin et al., 1998), suggesting that linker histone binding may cause changes in the folding of the histone octamer. At present it is difficult to assess whether such changes could account for the additional 20 bp of DNA protected from micrococcal nuclease digestion.

The 5S nucleosome model has been criticized on technical grounds also. The model depends on the assumption that the 5S nucleosome occupies a single dominant translational position with respect to the ends of the DNA fragment as determined by mapping using micrococcal nuclease. However, given the well known sequence-specificity of this enzyme, the precision and reliability of this technique have been questioned (An et al., 1998b). Moreover, multiple translational positions were observed for nucleosomes reconstituted onto the same DNA fragment when the same (An et al., 1998b) or different (Panetta et al., 1998) mapping techniques were employed.

An additional concern with the 5S model stems from the fact that only one GH5-DNA contact was observed. As discussed in section 1.13, some evidence suggests that GH5/GH1 possess 2 DNA binding sites and can bind 2 strands simultaneously. A recent study of the binding of GH5 to bulk (random sequence) mononucleosomes using an engineered globular domain of H5 bearing unique cysteine residues on the surface of the globular domain to facilitate incorporation of a photoactivatable cross-linking reagent, identified two GH5-DNA contacts, one involving helix-3 with the major groove within 5 bp of one DNA termini, and a second contact involving residues in site II of GH5 with the central turn of DNA close to the dyad (Fig. 1.6B) (Zhou et al., 1998). Thus in this model, GH1/GH5 bind near the dyad but binding is asymmetric nonetheless. Even
though this work employed recombinant protein corresponding to only the globular domain of H5, the authors concluded that the mode of GH5 binding positions the N- and C-terminal tails on the outside of the particle and enables the C-terminal tail to extend towards linker DNA. However, given the evidence that the C-terminal domain of linker histones, in particular, interacts strongly with DNA (see section 1.13), it seems likely that these results are biased by the absence of the H5 C-terminal domain. The studies performed on 5S reconstitutes employed full length linker histones (either H1, H5 or H1°). Whether this fundamental difference between these studies is wholly responsible for the different results remains to be resolved.

At present, it is not clear whether the asymmetric protection of linker DNA observed in the studies of positioned nucleosomes cited above also occurs in bulk nucleosomes so further analyses will be required to distinguish whether one of the possibilities depicted in Fig. 1.6B occurs preferentially in vivo. Hopefully, crystal structures of mono- or dinucleosomes containing linker histones will be obtained soon which can shed light not only on these issues but the location of the globular domain itself. Given the evidence (discussed below) that H1 plays a role in supranucleosomal folding of chromatin, precise knowledge of how H1 binds nucleosomes is likely to be central to understanding the role(s) of this molecule in the regulation of chromatin activity.

1.1.5 Supranucleosomal organization of chromatin

Early investigations of chromatin morphology by transmission electron microscopy using isolated nucleosomal fragments or chromatin “spread” from nuclei lysed at low ionic strength revealed that under these conditions chromatin appeared as a beaded string in which nucleosomes with an approx. diameter of 10 nm were joined by extended linker DNA (e.g. Olins and Olins, 1974; reviewed in van Holde, 1989; van Holde and Zlatanova, 1995). Observations that these “10 nm filaments” condensed upon exposure to higher ionic strength to form 25-30 nm diameter fibers, which appeared to be regular in structure in some instances (e.g. Fig. 1.1B; Thoma et al., 1979), fostered the notion that a regular pattern of folding may account for supranucleosomal structure in vivo. Thus, various models in which nucleosomal filaments are folded in regular helical (Finch and Klug, 1976; Thoma et al., 1979; McGhee et al., 1980; Butler, 1984) or ziz-zag (Worcel et al.,
fashion to form an approx. 30 nm diam. fiber have been proposed to account for the bulk form of chromatin in vivo. However, it has come to be generally agreed that the spreading and staining techniques used in much of this early work distort chromatin, casting doubt on the validity of observations based on such material (reviewed in Widom, 1989; van Holde and Zlatanova, 1995), and the details of these models will not be discussed here. Throughout the remainder of this thesis, the terminology “extended or decondensed” will be used to refer to the structure of transcriptionally active/competent chromatin and which appears to resemble that of chromatin prepared at low ionic strength. The term “condensed chromatin fiber” will be used in place of “30 nm fiber” to refer to more highly folded chromatin (van Holde and Zlatanova, 1995).

Although the presence of regular folding like that in the models cited above is presently the subject of much controversy (reviewed in van Holde and Zlatanova, 1995, 1996; Zlatanova and van Holde, 1996), it is apparent that nucleosomal filaments must be highly folded in vivo to compact genomic DNA sufficiently for it to be contained within nuclei. Analyses of sectioned nuclei by cryo-electron microscopy suggest that nucleosomal filaments are folded in some fashion to form “thick” chromatin fibers of variable diameter that have an irregular or kinky conformation rather than a regular cylindrical appearance (Horowitz et al., 1994; Woodcock, 1994). Despite uncertainty regarding the nature of higher order chromatin folding, several lines of evidence indicate that H1 contributes to this folding. Recent studies of unfixed chromatin fibers by cryo-electron microscopy and scanning force microscopy suggest, in contrast to earlier results from samples fixed for transmission EM, that extended nucleosomal filaments are folded in an irregular ziz-zag fashion at low ionic strength and that this folding is H1-dependent (Leuba et al., 1994; Bednar et al., 1998). Analyses of the sedimentation of nucleosomal arrays containing H1 or depleted of H1, and arrays containing intact core histones or arrays which have been treated with trypsin to remove the N-terminal tails of the core histones, have revealed that both the N-terminal tails of the core histones and H1 contribute to ionic strength-induced compaction of chromatin in vitro (Clark and Kimura, 1990; Hansen and Wolffe, 1992; Garcia-Ramirez et al., 1992, 1995; Fletcher and Hansen, 1995; Schwartz et al., 1996). Higher ionic strengths are required to achieve
similar degrees of compaction of H1-depleted arrays compared to arrays containing H1, suggesting that neutralization of linker DNA charge by H1 facilitates chromatin folding. These results are consistent with earlier findings from transmission electron microscopy that the presence of H1 was required for ionic strength-induced condensation of 10 nm "beads-on-a-string" nucleosomal filaments into 30 nm fibers (e.g. Fig. 1.1B; Thoma and Koller, 1977; Thoma et al., 1979). The notion that charge neutralization by H1 and the N-terminal tails of the core histones facilitates bending/compaction of linker DNA in condensed chromatin is consistent with proposed roles for H1 phosphorylation and core histone acetylation in contributing to decondensation of transcriptionally active/competent chromatin. A role for H1 in promoting higher order folding of chromatin is further suggested by the apparent depletion of H1 in isolated fractions of nuclease-sensitive/transcriptionally active chromatin (see section 1.2).

A role for H1-H1 interactions in promoting condensed fiber structure was implicit in many of the EM-based models cited above in which H1 was proposed to occupy an axial position in the interior of folded fibers but this remains controversial. Early analyses suggested that the globular domains of H1 and H5 did not self-associate in solution under a variety of conditions (Thomas and Khabaza, 1980; Draves et al., 1992). However, a recent study demonstrated a tendency for the globular domain of H5 to assemble in solution and form a specific structure (Maman et al., 1994). Curiously, GH1 showed little tendency to self-associate under the same conditions, suggesting the possibility that the higher order structure of chromatin containing these linker histones may differ. A subsequent study found that both GH5 and H5 interacted whether free in solution or bound to DNA (Carter and van Holde, 1998). Although data demonstrating that the globular domain of linker histones can interact in solution is often considered to support a role for such interactions in chromatin condensation, there is little direct evidence to suggest that this is the case. Neutron-scattering studies of condensed chromatin fibers containing deuterated H1 place the molecule in the interior of the fiber, but also suggest that that globular domains of adjacent H1 molecules are too distant for direct interaction (Graziano et al., 1994).
Early studies utilizing chemical cross-linking of chromatin to demonstrate interactions between the N- and C-terminal domains of H1 suggested possible roles for interactions of these domains in chromatin condensation (Ring and Cole, 1983; Nikolaev et al., 1981, 1983; Kotthaus and Stratling, 1984; Lennard and Thomas, 1985). However, since different types of cross-linked products (e.g. N-N, C-C and N-C) were found to be predominant in different studies, it is not apparent whether a particular interaction predominates in vivo. Such seemingly contradictory data is reminiscent of the current controversy regarding the nucleosomal binding site and orientation of H1 discussed above (section 1.15) and suggests that a reinvestigation of this issue might benefit from the improved methods of chromatin preparation and analysis available currently.

Condensed chromatin fibers are further folded in vivo and considerable evidence suggests that within nuclei, chromatin is organized into loops in which DNA elements referred to as MARs (matrix-attachment regions) at the base of loops are bound to the nuclear matrix (reviewed by Davie, 1995, 1997). Loops containing expressed genes appear to be less folded than inactive loops as the former are more sensitive to DNase I digestion than the latter (Davie, 1995, 1997). In interphase nuclei, individual chromosomes appear to occupy defined “territories” in which transcriptionally active regions appear to be spatially organized with respect to subnuclear compartments involved in mRNA processing. A review of this literature is beyond the scope of this thesis but the area has recently been reviewed elsewhere (Singer and Green, 1997; Misteli and Spector, 1998; Lamond and Earnshaw, 1998; Belmont et al., 1999).

1.2 Modulation of chromatin structure in vivo: chromatin transcription

Transcription of protein-coding genes in eukaryotes requires the assembly of a large preinitiation complex (PIC) containing more than 50 separate polypeptides on promoter DNA upstream of the site of initiation of mRNA synthesis. Investigations, most often utilizing naked DNA templates, have defined requirements for the general transcription factors (GTFs) TFIID, TFIIB, TFIID, TFIIE, TFIIF and TFIIF, and RNA polymerase II, to reconstitute transcription in vitro (reviewed in Orphanides, et al., 1996; Roeder, 1996). TFIID, a multimeric protein complex consisting of the TATA box-binding protein (TBP) and a collection of TBP-associated factors
(TAF11s), plays a critical role in this process since it is the only PIC component capable of binding core promoters specifically (reviewed in Burley and Roeder, 1996). Control of promoter recognition by TFIID is thought to represent an important mechanism in transcriptional regulation. Upon binding their cognate DNA targets, many transcriptional activators interact directly, or via intervening proteins referred to as coactivators or adaptors, with TAF11s to enhance the rate of promoter binding by TFIID or stabilize TFIID-promoter complexes (reviewed in Goodrich and Tjian, 1994; Kingston and Green, 1994; Struhl, 1996; Burley and Roeder, 1996).

A major goal of current research in molecular biology is to understand how transcriptional regulation is accomplished in the context of chromatin. As discussed above, DNA elements involved in transcriptional regulation (e.g. enhancers and promoters) and the protein-coding segments of genes exist in vivo as a highly constrained chromatin fiber. Analyses of transcription from chromatin templates assembled in vitro have revealed that nucleosomes are not an absolute barrier to transcription but that the ability of transcriptional activators to bind their cognate DNA sites is influenced by the positioning of nucleosomes with respect to these sites (reviewed in Fedor, 1992; Kornberg and Lorch, 1992; Felsenfeld, 1992; Owen-Hughes and Workman, 1994). Thus, an intimate relationship must exist between the transcription apparatus and the chromatin in which it functions. Below I review recent research which indicates that modulation of chromatin structure through the post-translational modification of histones plays a central role in transcriptional regulation. The histones and their modification are likely to play similar roles in other DNA-templated processes such as replication and repair but since these have been much less researched, the discussion here will focus primarily on chromatin transcription.

1.2.1 General features of transcriptionally active chromatin

As mentioned in Chapter 1.1, early investigations of chromatin structure in isolated nuclei revealed that transcriptionally active chromatin was distinguished from bulk chromatin in that it displayed heightened susceptibility to digestion by nucleases and various schemes have been devised to separate active chromatin from bulk chromatin. Fractionation of the chromatin
fragments released by mild digestion of nuclei with micrococcal nuclease according to solubility in 0.15 M NaCl has been employed to enrich for active chromatin (e.g. Komaiko and Felsenfeld, 1985; Ridsdale and Davie, 1987; Ridsdale et al., 1988; reviewed in van Holde 1989). The soluble fraction obtained by these procedures is enriched in transcribed genes and acetylated forms of core histones and ubiquitinated H2A and H2B relative to bulk chromatin. In most cases, the content of H1 in such fractions is much lower than that of bulk chromatin. Such evidence, although indirect, taken together with the increased nuclease sensitivity of this chromatin, has long suggested that core histone hyperacetylation and H1 depletion are associated with decondensation of chromatin fibers to permit transcription. As discussed below, the recent identification of enzymes responsible for core histone acetylation, and the development of direct methods to isolate chromatin fragments according to the presence of modified histones, have permitted direct links to made between histone modification and transcriptional activity.

1.2.2 Acetylation of core histones

Two types of reversible acetylation affecting the ε-amino groups of lysines in histones are recognized that are distinct from the co-translational α-amino acetylation that occurs for many eukaryotic proteins. Acetylation of newly synthesized cytoplasmic histones H3 and H4 at specific sites, thought to facilitate the incorporation of these proteins into replicating chromatin, is referred to as deposition-related acetylation and enzymes responsible for this have historically been referred to as “B-type” HATs. Diacetylation of newly synthesized H4 at lysines 5 and 12 occurs in a wide variety of organisms but sites of deposition-related acetylation of H3 are more variable (Turner and O’Neill, 1995; Sobel et al., 1995). In contrast, acetylation related to transcription occurs on all four core histones. Nuclear enzymes responsible for acetylation of chromatin-associated histones have historically been referred to as A-types HATs (reviewed in Mizzen and Allis, 1998; Grant and Berger, 1999). Both classes of enzymes catalyse the transfer of acetate from acetylCoA to the ε-amino group of conserved lysine residues in the amino terminal tails of the core histones (e.g. see
triangles in Fig 1.4), neutralizing the positive charge associated with the residue. The positive charge is restored upon deacetylation of the residue by histone deacetylases (HDACs).

A novel screening strategy enabled the recent cloning of the HAT1 gene that encodes the catalytic subunit of the major cytosolic HAT in yeast (Kleff et al., 1995). Characterization of the HAT-B activity in yeast extracts revealed that the HAT1p catalytic subunit functions in a complex with another subunit, HAT2p (Parthun et al., 1996). HAT2p is responsible for the high affinity binding of H4 by the HAT1p/HAT2p complex which acetylates free (non-nucleosomal) H4 exclusively in vitro. Enzymes responsible for deposition-related H3 acetylation in yeast have not been identified. It is puzzling that disruption of the HAT1 gene does not confer an obvious phenotype (Kleff et al., 1995; Parthun et al., 1996) given the widespread conservation of H4 diacetylation (Sobel et al., 1995) and the presumed role of deposition-related acetylation of H3 and H4 in facilitating chromatin assembly (reviewed in Roth and Allis, 1996).

A human homologue of HAT2p, referred to as p48, has been identified as a subunit of a multimeric complex, CAF-1, involved in DNA replication-dependent chromatin assembly (Verreault et al., 1996). Significantly, p48 has also been shown to be a subunit of the HD1 human deacetylase complex (Taunton et al., 1996), suggesting a role for p48 in the coordinate regulation of chromatin assembly and subsequent deacetylation of H3 and H4 following incorporation into nucleosomes (reviewed in Roth and Allis, 1996; Grant and Berger, 1999). At present, enzymes responsible for deposition-related acetylation of H4 in species other than yeast have not been unequivocally identified but corresponding activities have been identified in maize and human cells (Eberharter et al., 1996; Verrault et al., 1997).

Although only one B-type HAT catalytic subunit has been identified to date, multiple A-type activities have been identified following the initial identification of the p55 protein of Tetrahymena as a nuclear HAT (Brownell et al., 1996). The identities and characteristics of these various enzymes have been reviewed elsewhere (Mizzen and Allis, 1998; Workman and Kingston, 1998; Kouzarides, 1999; Grant and Berger, 1999) and a comprehensive review of this rapidly
growing literature is beyond the scope of this thesis. Salient to the discussion here is the fact that in many cases, these proteins were known to have roles in transcriptional activation prior to the discovery that they possessed HAT activity. For example, the yeast homolog of Tetrahymena p55, GCN5, which also possesses HAT activity, had previously been identified in genetic screens in yeast as a coactivator necessary for activation of gene expression by certain DNA-binding transcriptional activators (Georgakopoulos and Thireos, 1992). Similarly, well known coactivators in metazoans such as the TAFii250 subunit of TFIID and CBP/p300 were found to possess intrinsic HAT activity (Mizzen et al., 1996; Ogryzko et al., 1996). Together with findings that other proteins with known or suspected roles in targeting the transcriptional apparatus to promoters were also HATs, these data strengthened the notion that histone acetylation is an integral step in gene activation. As discussed below, functional analyses of HAT activity have borne out this expectation.

The requirements for HAT activity for coactivator function have been best characterized for GCN5 in yeast. Mutations which decreased the HAT activity of GCN5 in vitro, compromised the ability of the mutated GCN5 to complement transcriptional defects in GCN5 deletion strains (Kuo et al., 1998; Wang et al., 1998). That this was attributable to alterations in the HAT activity of the mutated GCN5 rather than other functions of GCN5 was demonstrated directly using antibodies specific for acetylated core histones in chromatin immunoprecipitation analyses. These revealed that the HAT activity of GCN5 was required for histone acetylation at a GCN5-dependent promoter in vivo (Kuo et al., 1998). Similar evidence has been provided for the HAT activity of CBP. Mutations which reduced the HAT activity of CBP in vitro also reduced transcriptional activation by a Gal4-CBP fusion transfected into mammalian cells (Martinez-Balbas et al., 1998). Mutational analyses have also provided evidence that specific activators require the HAT activity of specific coactivators. CBP and p300 function as coactivators for a variety of transcriptional activators, including nuclear hormone receptors and CREB (reviewed in Shikama et al., 1997). In addition, these proteins serve as platforms for the assembly of multimeric complexes that contain other proteins with HAT activity such as PCAF (Yang et al., 1996) and the nuclear receptor coactivators ACTR and SRC-1 (Chen et al., 1997; Spencer et al., 1997). Mutational analyses have
revealed that the activators Myo D and the retinoic acid receptor require the HAT activity of PCAF but not that of CBP/p300 whereas activation by CREB requires the HAT activity of CBP/p300 and not that of PCAF (Puri et al., 1997; Korzus et al., 1998).

Taken together, the reports cited above underscore the importance of histone acetylation for chromatin transcription. Yet, mechanisms by which acetylation facilitates transcription are poorly understood. An extensive literature, which will only be summarized here, provides evidence for four different mechanisms by which acetylation may facilitate chromatin transcription. (1) Alterations of nucleosome conformation which may increase the accessibility of nucleosomal DNA to transcription factors have been described for core particles reconstituted with acetylated histones (Norton et al., 1989; Bauer et al., 1994). (2) Numerous studies have provided evidence that acetylation attenuates direct inhibition of transcription factor binding by the amino-terminal tails in nucleosomes reconstituted on DNA fragments containing their cognate sequences (e.g. Lee et al., 1993; Vettese-Dadey et al., 1994, 1996) and acetylation has also been reported to facilitate TBP binding to nucleosomes containing a TATA box (Imbalzano et al., 1994; Godde et al., 1995). (3) Sedimentation analyses of ionic strength-induced condensation of oligonucleosomes prepared with acetylated histones or histones treated with trypsin to remove the N-terminal tails (presumed to mimic acetylation) have revealed that acetylated/trypsinized oligomers remain in an extended conformation compared to nonacetylated/nontrypsinized controls (Allan et al., 1982; Ridsdale et al., 1990; Garcia-Ramirez et al., 1992; Garcia-Ramirez et al., 1995). This suggests that acetylation may increase the access of the transcriptional apparatus to promoters by promoting decondensation of chromatin fibers. Increased transcription by RNA polymerase III of 5S RNA genes incorporated into nucleosomal arrays was recently shown to correlate with the extended conformation of acetylated versus nonacetylated arrays (Tse et al., 1998). (4) Acetylation may modulate interactions of regulatory factors with the histone amino terminal tails. Transcriptional silencing by the SIR3 and Tup1 proteins in yeast depends on interactions with the amino terminal tail of H4 and genetic evidence are consistent with the possibility that acetylation may modulate these interactions (Johnson et al., 1990, 1992; Edmondson et al., 1996). The interaction of the Drosophila nucleosome remodeling complex NURF with nucleosomes is impaired by histone N
terminal tail proteolysis (Georgel et al., 1997), suggesting a role for acetylation in modulating the recruitment of remodeling activities (remodeling activities are discussed in section 1.2.5). Finally, the recent discovery that the bromodomain of PCAF interacts specifically with acetylated lysine residues in H3 and H4 tail peptides (Dhalluin et al., 1999) suggests that acetylation may contribute to recruitment of a variety of regulatory factors since bromodomains are found in large number of transcriptional regulators, including several HATs other than PCAF (reviewed in Winston and Allis, 1999).

Taken together, the data suggest that more than one, and possibly all, of the mechanisms discussed above contribute to facilitation of chromatin transcription by core histone acetylation. In addition, it should be recognized that several of the enzymes originally described as HATs by virtue of their ability to acetylate histones in vitro, appear to have additional targets, including transcriptional activators, in vivo. Thus, an emerging view is that acetylation represents a major regulatory pathway in nuclei whose diversity of targets and effects appears to be similar to that of phosphorylation cascades (reviewed by Berger, 1999; Kouzarides, 2000).

The importance of core histone acetylation in transcriptional regulation was further underscored by the contemporaneous identification of histone deacetylases. Human histone deacetylase 1 (HDAC1) isolated by affinity chromatography on a matrix containing a form of the HDAC inhibitor trapoxin (Taunton et al., 1996) was found to share significant homology to the yeast protein RPD3, a known transcriptional regulator. Subsequent demonstration that RPD3 containing complexes isolated from yeast also possessed deacetylase activity (Carmen et al., 1996), preceded the rapid identification of other histone deacetylases in a variety of species based on homology to RPD3 or by biochemical means. These enzymes are capable of removing the acetyl groups that HATs add to the ε-amino groups of lysines in the N-terminal tails of core histones, restoring the positive charge at these residues to modulate interactions of the N-terminal tails with components of chromatin. Currently, three classes of HDACs are recognized based on sequence homology. In general, these proteins are components of multi-subunit complexes that...
are recruited to chromatin loci through physical interactions with negative regulators of transcription. Analyses of the requirements for the activity of HDACs in vivo has revealed roles as co-repressors of specific genes and in chromatin silencing (reviewed by Ayer, 1999; Ng and Bird, 2000). Furthermore, reports that some HDAC complexes interact with known cell cycle regulators and oncoproteins suggest potential roles in cancer (reviewed by Davie et al., 1999; Magnaghi-Jaulin et al., 1999; Cress and Seto, 2000).

1.2.3 Other modifications of core histones

Metabolic labeling has revealed that much or all of histone H3 is phosphorylated at serine 10 in mammalian cells at mitosis (reviewed by Bradbury, 1992) and a phospho-Ser10-H3 specific antisera has revealed that H3 phosphorylation correlates with mitotic chromosome condensation in organisms ranging from yeast to humans (Hendzel et al., 1997; Wei and Allis, 1998). Evidence that this phosphorylation is required for proper chromosome condensation and segregation has come from experiments in Tetrahymena in which mutagenesis of this phosphorylation site led to extensive chromosome loss (Wei et al., 1999). Recently, it has been shown that H3 is also phosphorylated at serine 28 in mammalian cells but the functional consequences of this phosphorylation are not known.

Ser10 phosphorylation of a small subset of H3 molecules correlates with the activation of expression of “early response” genes including the transcription factors c-fos and c-jun in response to growth factor stimulation of quiescent cells (Mahadevan et al., 1991). Direct evidence that H3 phosphorylation contributes to activation of such loci comes from the demonstration that phosphorylated H3 becomes associated with c-fos and c-myc genes upon activation in chromatin immunoprecipitation assays using the phospho-Ser10-H3 specific antisera (Chadee et al., 1999). Although little is known of the role of H3 phosphorylation in gene activation, it seems likely that one or more of the possible mechanisms discussed above for core histone acetylation may apply.

Forms of H2A and H2B that are ubiquitinated in their C-terminal regions have been found to be enriched in isolated fractions of active chromatin. In contrast to the ubiquitination of other
proteins that marks them for degradation by the proteasome (reviewed Varshavsky, 1997), ubiquitination does not appear to be related to turnover of H2A and H2B and the addition of this 76 residue acidic protein is thought to have a role in altering chromatin structure to facilitate transcription although few details of the mechanism are known (reviewed by Davie, 1997).

1.2.4 H1 function and modification

In general, linker histones repress transcription from both naked DNA and chromatin templates in vitro (Hannon et al., 1984; Croston et al., 1991; Laybourn and Kadonaga, 1991; O’Neill et al., 1995; Ura et al., 1997), supporting the notion that they are general repressors. Taken together, the experimental evidence suggests that at least three interrelated mechanisms contribute to H1-mediated inhibition of transcription initiation and/or elongation: direct inhibition due to occlusion of regulatory DNA sequences upon binding of H1, inhibition due to decreased mobility of nucleosomes containing H1, and inhibition due to higher order folding of chromatin. Evidence that linker histones can compete with transcription factors for binding sites comes primarily from analyses of transcription from naked DNA templates (e.g. Croston et al., 1991) and studies demonstrating preferential binding of H1 to consensus sites for transcription factors (Ristiniemi and Oikarinen, 1989), AT-rich DNA found in flanking sequences of genes and matrix-attachment regions (Sevall, 1988; Jerzmanowski and Cole, 1990; Pauli et al., 1989; Izaurralde et al., 1989; Zhao et al., 1993) and DNA with altered conformation (Hendrickson and Cole, 1984; Varga-Weisz et al., 1994) relative to bulk DNA support the notion that H1 binding may directly hinder the access of regulatory factors to binding sites in chromatin in vivo.

H1 binding has been reported to decrease the mobility of nucleosomes on chromatin templates in vitro (Pennings et al., 1994; Ura et al., 1997) and it has been suggested that this contributes to repression of transcription in vivo. H1 binding and repression of RNA pol III transcription from Xenopus 5S RNA genes incorporated into a dinucleosomal template was found to be independent of whether core histones were acetylated or not, suggesting that decreases in the mobility of nucleosomes due to H1 binding are dominant over facilitatory effects of core histone acetylation (Ura et al., 1997). An opposing view is supported by a study in which core histone
acetylation or amino tail proteolysis were found to alleviate H1-mediated repression of the binding of the transcription factor USF to mononucleosomes (Juan et al., 1994). H1 binding was not found to influence nucleosome mobility in this latter system, leading to the proposal that H1 binding inhibits factor binding by reducing transient dynamic dissociation of nucleosomal DNA from the core histone octamer (Juan et al., 1997).

Indirect evidence that H1 represses transcription through the facilitation of higher order chromatin folding comes from numerous reports that nuclease sensitive chromatin fractions enriched in transcribing genes are found to be depleted of H1 when isolated (see section 1.2.1). Although well known, it is possible that the H1-depletion reported may be artifactual in some cases, stemming from the cooperative binding of chromatin by H1 which promotes redistribution of H1 from short to longer chromatin fragments at the ionic strengths used in these studies (Lasters et al., 1981; Caron and Thomas, 1981; Thomas and Rees, 1983). Studies using alternate approaches have reported that active and inactive chromatin contain similar amounts of H1 (Colavito-Shepanski and Gorovsky, 1983; Weintraub, 1984; Hill et al., 1989; Nacheva et al., 1989; Ericsson et al., 1990; Dedon et al., 1991), but that the interaction of H1 with DNA (Nacheva et al., 1989) or the stability of supranucleosomal structures (Weintraub, 1984) in active chromatin is altered compared to bulk chromatin. These latter findings are consistent with recent evidence that phosphorylation of H1 plays a role in transcriptional regulation (discussed below).

Although mechanisms by which H1 may mediate higher order folding of chromatin are unclear (see section 1.1.5) direct evidence that linker histones are involved in chromatin condensation in vivo has been obtained in Tetrahymena. Macronuclear H1 knockout strains are viable but have macronuclei that are enlarged two-fold compared to wild type cells. Conversely, micronuclear linker histone knockout strains have enlarged micronuclei (Shen et al., 1995). Subsequent analyses revealed that although transcription by pol I and pol III were unaffected, both positive and negative changes in the transcription of specific genes by pol II occurred in response to the loss of macronuclear H1 (Shen and Gorovsky, 1996). These results suggest linker histones can act as positive or negative gene-specific transcriptional regulators and further evidence of selective
effects of H1 on transcription have been obtained in analyses of embryogenesis in Xenopus (Wolffe et al., 1997; Steinbach et al., 1997). Similarly, preferential repression of pol I-transcribed rRNA genes and the pol II-transcribed ACT1 and URA3 genes, but not the pol II-transcribed Ty gene, occurred when a sea urchin H1 was overexpressed in yeast (Linder and Thoma, 1994). However, disruption of the endogenous HHOI gene did not lead to any overt changes in transcription nor did it lead to detectable changes in chromatin organization (Patterson et al., 1998). This discrepancy may relate to the unique structure of the HhoI protein itself or to specific features of chromatin structure in yeast which possess an unusually short repeat length (approximately 160 bp) and hence, minimal linker DNA. Evidence that H1 is involved in transcriptional regulation in higher eukaryotes comes from analyses of mutants of the DT40 chicken B cell line. Cells engineered to express only one copy a single H1 gene contained approximately one half the normal levels of H1 mRNA and protein but their growth rate or global chromatin structure was not altered compared to the wild-type cell line (Takami and Nakayama, 1997). However, 2-D gel analyses revealed that the pattern of protein synthesis was altered in the mutant cells, suggesting gene-specific effects for H1 in agreement with the results obtained in Tetrahymena.

The nearly universal correlation of linker histone hyperphosphorylation with mitosis has suggested to many that H1 hyperphosphorylation directly promotes higher degrees of chromatin compaction and by extension, repression of transcription (reviewed in Bradbury, 1992). However, recent experiments demonstrating that H1 is not required for chromosome condensation in vitro (Ohsumi et al., 1993), and that chromosome condensation induced by phosphatase inhibitors does not require H1 hyperphosphorylation (Guo et al., 1995) suggest that mitotic H1 phosphorylation is distinct from mechanisms with more direct roles in causing chromosome condensation (e.g. condensins, see Hirano et al 1997; Koshland and Strunnikov, 1996). Moreover, other studies have shown that condensed chromatin structures in interphase nuclei are frequently enriched in dephosphorylated forms of linker histones (reviewed in Roth and Allis, 1992). Using antibodies specific for phosphorylated and dephosphorylated H1, it has been shown that in Tetrahymena, dephosphorylated H1 localizes to condensed chromatin within macronuclear chromatin bodies while phosphorylated H1 localizes to euchromatin peripheral to the chromatin bodies. Parallel staining
with other antisera revealed that the distributions of TBP and hv1 (an H2A variant enriched in transcriptionally active chromatin) resembled that of phosphorylated H1 while the major H2A protein was distributed equivalently throughout euchromatin and chromatin bodies (Lu et al., 1995). These data suggest that phosphorylated H1 is enriched in transcriptionally competent chromatin while dephosphorylated H1 is enriched in condensed, presumably transcriptionally inactive chromatin. Phosphorylated H1-specific antisera has also been used to show that transformation of cells by a variety of means is accompanied by increased levels of phosphorylated H1 and decondensation of bulk chromatin structure compared to parental cells (Taylor et al., 1995; Chadee et al., 1995; Herrera et al., 1996). Similarly, phosphorylation of H1 is required for glucocorticoid receptor-mediated disruption of mouse mammary tumor virus (MMTV) promoter chromatin in vivo (Lee and Archer, 1998), arguing collectively that linker histone phosphorylation plays a role in decondensing chromatin which may be an important mechanism for altering gene expression. Direct evidence that H1 phosphorylation plays a role in transcriptional regulation comes from Tetrahymena strains mutated at five phosphorylation sites in macronuclear H1 to mimic constitutive phosphorylation or constitutive dephosphorylation (Dou et al., 1999). As was found for the macronuclear H1 knockout strain discussed above, global gene expression was not detectably altered in either strain although the transcription of specific genes was activated or repressed compared to wild type cells. Interestingly, the pattern of gene expression in cells expressing the "constitutively phosphorylated" H1 resembled that of H1 knockout cells, supporting the notion that the conformation of chromatin containing phosphorylated H1 may resemble that of H1-depleted chromatin.

While it has generally been found that H1 represses transcription in vitro, more recent experiments suggest that H1 depletion and H1 phosphorylation can have both positive and negative effects on transcription in vivo. This discrepancy underscores the need for further information regarding the mechanism(s) by which H1 regulates transcription and how this may be modulated by phosphorylation and possibly other modifications (e.g. poly(ADP ribosylation), reviewed in Boulikas, 1993). Multisite phosphorylation significantly decreases the affinity of H1 for chromatin (Hill et al., 1991), potentially enhancing access for trans-acting factors involved in either chromatin decondensation or compaction (Roth and Allis, 1992). However, evidence suggests that the
relationship between H1 phosphorylation and chromatin function is likely to be complex. Studies of the sedimentation and in vitro replication of SV40 chromatin reconstituted with differentially phosphorylated forms of H1 prepared from G0-, S-, or M-phase cells suggest that phosphorylation at different sites in H1 is associated with specific chromatin conformations (Halmer and Gruss, 1996). Combined with evidence for functional differences among the multiple forms of H1 found in most eukaryotes (see section 1.3.5), this suggests the possibility that the function of H1 in chromatin is highly varied.

1.2.5 Chromatin remodeling activities

Recently, protein complexes that appear to be capable of “remodeling” chromatin structure to facilitate transcription have been described. The best studied of these is yeast SWI/SNF, a large multi-protein complex named for the SWI2/SNF2 subunit (switching mating type 2/sucrose non-fermenting 2) which contains a domain with homology to known DNA helicases and which possesses DNA-dependent ATPase activity (Kingston et al., 1996; Pazin and Kadonaga, 1997). Several other complexes have been described, including human SWI/SNF, yeast RSC, mammalian NRD/NuRD and Drosophila NURF, CHRAC and ACF, all of which contain a subunit that shares homology with yeast SWI2/SNF2 (reviewed in Kadonaga, 1998; Gregory and Horz, 1998; Varga-Weisz and Becker, 1998). In yeast, deletion of SWI/SNF complex subunits causes defects in the transcription of a subset of genes which can be suppressed by reduced synthesis of histones H2A and H2B (Hirschhorn et al., 1992) or partially suppressed by single point mutations within the histone fold domains of H3 and H4 (Kruger et al., 1995). These data suggest that such complexes antagonize nucleosome-mediated repression of transcription and several remodeling complexes have been shown to increase transcription factor binding or restriction enzyme access to nucleosomal DNA in ATP-dependent fashion in vitro (e.g. Cote et al., 1994; Cairns et al., 1996). Although the mechanism of action of any of these complexes is not known at present, it is thought, in general, that they use energy from ATP hydrolysis to effect alterations in nucleosome structure that do not involve complete dissociation of histones from DNA (reviewed in Workman and Kingston, 1998). The enhanced accessibility of nucleosomal DNA to digestion by micrococcal nuclease or DNase I that represents the basis of common assays for these activities in vitro suggests
that these complexes are able to disrupt or weaken histone-DNA contacts (Kadonaga, 1998; Workman and Kingston, 1998). One possible mechanism of how such complexes may facilitate transcription is suggested by the finding that the yeast SWI/SNF, and the Drosophila NURF and CHRAC complexes all catalyse nucleosome sliding along template DNA in vitro (Whitehouse et al., 1999; Hamiche et al., 1999; Langst et al., 1999). At present it is not known, what role, if any, that such complexes may have in antagonizing H1-mediated repression of transcription.

1.3 Structural variants of histones: detection, occurrence and function

Another means for variegating chromatin structure is the incorporation of variant histone molecules that differ in amino acid sequence compared to other molecules of that histone class. Prior to the widespread use of molecular genetic technology, surveys of histone expression using specialized electrophoretic and chromatographic techniques revealed that multiple structural variants with limited but characteristic differences in amino acid sequence, are found for all histones except H4 in tissues of many eukaryotes. Depending on the species and tissue analysed, 5-7, 2-5, 2-3 and 2-3 forms of H1, H2A, H2B and H3, respectively, have been resolved which are not accounted for by post-translational modification (reviewed in Zweidler, 1984; Cole, 1984, 1987; Lennox and Cohen, 1984; von Holt et al., 1984; van Holde, 1989; Wolffe, 1998). The distinct electrophoretic and/or chromatographic properties of these forms enabled their expression in different species, in different tissues of a species, and in individual members of a species, to be compared. Typically, it was found that the pattern of expression of these forms was equivalent in all individuals within a species, indicating that they represented non-allelic variants resulting from the coexpression of genes shared by all members of a species, rather than allelic variants attributable to the expression of distinct alleles in different individuals (allelic variants of histones have been resolved in some instances - see sections 1.3.2 and 1.3.4). Coupled with findings that the characteristic amino acid sequence differences and metabolic properties of some non-allelic variants are conserved among many species, these observations led to the hypothesis that non-allelic variants have acquired specialized functions in chromatin during evolution.
Subsequently, molecular genetic analyses revealed that multiple genes are present in most eukaryotes for each of the five histones. The organization and regulation of histone genes has been intensively studied in a variety of organisms. Although a review of this literature is beyond the scope of this thesis (see Stein et al., 1984; Nakayama, 1993; Wang et al., 1997; Doenecke et al., 1997 for reviews), some general notions supported by these studies are summarized here to facilitate discussion of the structure and function of histone variants. Genes encoding non-allelic histone variants appear to have arisen by duplication of ancestral genes with subsequent accumulation of alterations in DNA and protein sequence during evolution. Not all changes in the nucleotide sequence of the coding region of a gene lead to an altered protein due to the degeneracy of the third base in codons and codon redundancy. However, in other cases, point mutations or other alterations that affect both alleles of a gene copy lead to expression of a non-allelic variant that differs in protein sequence from other members of that histone class and that corresponds to one of the distinct electrophoretic or chromatographic forms discussed below.

Histone genes can be classified into one of two groups based on their pattern of expression. The majority are replication-dependent genes whose expression is regulated in the cell cycle. Complex regulatory pathways controlling the transcription of these genes, and the processing and turnover of their transcripts, restrict the synthesis of replication-dependent histones to S-phase when newly synthesized histones are required to replicate chromatin prior to cell division (reviewed in Osley, 1991; Heintz, 1991). Replication-independent histone genes comprise a less abundant set of genes that are expressed throughout the cell cycle (Wu and Bonner, 1981). These are often called basal or replacement histone genes since the histones they encode are presumed to replace replication-dependent forms deposited in replicating chromatin during the previous S-phase. Several of the non-allelic variants of core and linker histones that have been proposed to have specialized functions are replacement forms.

The notion that variegation of histone structure could be functionally related to differences in chromatin structure and activity seems rational and has proven to be popular, particularly in relation to the non-allelic variants of H1 (Cole, 1984, 1987; Lennox and Cohen, 1984). However,
direct evidence supporting this hypothesis is limited and the majority of support comes from indirect measures of variant function such as their expression in developing organisms or their abilities to interact with chromatin and DNA in vitro. Given this situation, I will only briefly summarize the voluminous literature describing electrophoretic and chromatographic variants in order to illustrate the techniques employed and typical results that have led to the notion that histone variants may have specific functions. Whenever possible, evidence from biochemical assays or experiments using molecular genetic approaches (e.g. gene knockouts) to test histone variant function in vivo will be described.

1.3.1 Evolutionary divergence of histone structure

The primary structures of all five classes of histones have diverged more slowly than those of most nonhistone proteins during evolution and are highly conserved. In general, between species and across plant and animal phyla (reviewed in Isenberg, 1978; van Holde, 1989). In conjunction with the apparent universality of nucleosome structure, this suggests that each histone class performs the same general function in different eukaryotic species (van Holde, 1989). However, constraints on these functions may differ among the five classes of histones since discernable differences exist in the rates and extents by which their sequences have changed during evolution. Based on changes in the amino acid sequences of histones in different species, compared to the changes in other types of proteins (using information available circa 1978), estimates of the degree of protein sequence divergence over evolutionary time shown in Fig. 1.7 suggested that histones H3 and H4 are among the most invariant proteins known with an estimated rate of change of less than 1% in 400 million years. Histone H2A and H2B sequences have been less conserved and have diverged more rapidly, with an estimated rate of change of approximately 6% over the same interval. Relative to the core histones, linker histone sequences are much more variable and have diverged even more rapidly with an estimated rate of change of 12% in 100 million years. Comparisons of all the histone sequences currently available for each class of histone in the histone sequence database, available on the World Wide Web at http://genome.nhgri.nih.gov/histones/ (Makalowska et al., 1999), confirm the general trends shown in Fig. 1.7 but reveal that H3 has diverged discernably more than H4 (see below). When
Figure 1.7. Rates of histone sequence divergence during evolution.

This is a schematic representation based on protein sequence data available circa 1978 and is used here to demonstrate that the amino acid sequence of H1 is much more variable than are the sequences of the core histones. As described in the text, an updated version would probably put a separate line for H3 intermediate to that of H4 and the line for H2A and H2B. (Figure reproduced from van Holde, 1988 with permission)
the sequences of individual histones from different species are compared, changes in sequence are most often localized to the N- and C-terminal domains flanking the histone fold domain of histones H2A, H2B and H3 and similarly in the N- and C-terminal domains flanking the globular core of metazoan linker histones (van Holde, 1989; Makalowska et al., 1999).

### 1.3.2 Allelic and non-allelic variants of core histones

Non-allelic variants of H2A, H2B and H3 were first detected using electrophoresis in acid-urea-Triton X-100 (AUT) gels (Zweidler and Cohen, 1972; Franklin and Zweidler, 1977; Zweidler, 1978). This technique is able to separate variants differing in neutral amino acid substitutions based on alterations in hydrophobicity, and possibly other properties, in addition to resolving forms that differ sufficiently in size or net charge (reviewed by Zweidler, 1984). Both AUT and acid-urea (AU) gels are capable of separating many charge-modified forms of histones, i.e. molecules that are acetylated or phosphorylated to different extents in addition to variants differing in substitutions of positively-charged residues. Thus, appropriate experiments are required to distinguish electrophoretic heterogeneity attributable to post-translational modifications from that due to amino acid sequence differences. Many non-allelic variants described to date that are resolved on AUT gels comigrate on AU gels, suggesting substitutions of neutral amino acids are frequently involved, and in many cases this has been confirmed by protein sequence analyses (Zweidler, 1984). Surveys of the histones expressed in various mouse tissues and in liver of developing mice published previously by Zweidler (1984) are shown in Fig. 1.8 to demonstrate the resolution of non-allelic variants of histones that can be attained with this system. Note that two or more bands were resolved for each class of histone in this instance. Franklin and Zweidler (1977) proposed a nomenclature wherein numerals appended to the histone class designation, e.g. H2A.1, H2A.2, H2B.1, H2B.2, H3.1, H3.2 are used to designate non-allelic variants of H2A, H2B and H3. In most cases, this nomenclature has also been applied to the genes encoding these variants upon identification.

As mentioned above, H4 sequences have been highly conserved during evolution and this is reflected in the fact that non-allelic variants of H4 have not been described in any species. The
Figure 1.8. Resolution of non-allelic variants of histones by AUT-PAGE.

(A) Differences in the expression of non-allelic variants of core histones among different tissues of the adult mouse as detected by AUT-PAGE are shown in the upper panel. Small numbers denote variants (e.g. H2B.1, H2B.2). "U" indicates ubiquitinated forms. Note the differences in the relative abundance of variants of H2A, H3, and H2B in different tissues. In the lower panel, changes in the PCA soluble linker histone and HMG proteins are shown on a separate gel. Note the difference in the relative abundance of H1O in different tissues. Proteins stained with Amido Black.

(B) Changes in the expression of non-allelic variants of histones in developing mouse liver as revealed by AUT-PAGE. Total histones were isolated from livers of mice from one week before birth ("-1") to 30 weeks after birth. "T" = mouse thymus standard. Note the changes in relative abundance of the variants of H2A, H3, H1, and H2B with age. Proteins stained with Amido Black. (Figure reproduced from van Holde, 1988 with permission)
| H2A.1  | --SGRGKQGG KARAKAKTRS SRAGLQFPVG RVHRLLR-KG NYSERVGAGA PVYLAAVLEY LTAEILELAG NAARDNKKTR |
| H2A.2  | --********* *********S** *********-- **A*********** **M*********** *********-- *********-- |
| H2A.3  | --********* *********-- *********-- *********-- *********-- *********-- *********-- *********-- |
| H2A.X  | --*********T** *********G** *********-- H*A** *********-- *********-- *********-- *********-- |
| H2A.Z  | AGGKA**DS* **KT**VS** Q********* I**H*KRTS TSHG****T* A**S**I*** ***V***** **SK*L*VK* |

| H2B.1  | PEPTKSAPAP KKGSKKAVTK AQKKGKDKRK RSRKESYSVY VYKLKQVHP DTGISKSMAG IMNSFVNDIF ERIAGEASRL |
| H2B.2  | --VR*V** *********-- *********-- *********-- *********-- *********-- *********-- *********-- |
| TH2B   | *********-- *********-- *********-- *********-- *********-- *********-- *********-- *********-- |

| H3.1   | ARTKQTARKS TGGKAPRKQL ATKAARKSAP ATGGVKKPHR YRPGLTVLRE IRRYQKSTEL LIRKLPQRL VREIAQDFKT |
| H3.2   | *********-- *********-- *********-- *********-- *********-- *********-- *********-- *********-- |
| H3.3   | --*********-- *********-- *********-- *********-- *********-- *********-- *********-- *********-- |
| CENP-A | --GP*RK P--QT**RRP SSP*PGP*RQ SSSVGSQLT* R*QKMFW*K* *KTL***D* *F*K*SMV ***CEK*SR |

---DLRFQSSA VMALQEACEA YLVGLFEDTN LCAIHAKRT IMPKDQLAR RIRGERA--- 135
---******A* IG*****S** *********-- *********-- *********-- *********-- *********-- 135

---*******A* IG*****S** *********-- *********-- *********-- *********-- *********-- 135

GV*FWA*AQ* LL*****A** F*IH*****AY *LSL**G** LF*******T* ***FEGLP 133
Figure 1.9. Amino acid sequences of non-allelic variants of mouse core histones.

The amino acid sequences of non-allelic variants of mouse core histones H2A, H2B and H3 predicted from nucleotide sequences contained in the non-redundant set of the histone sequence database as of May 2000 are shown in one letter code after alignment using ClustalW 1.8 (Higgins et al., 1996). The sequences are identified according to the nomenclature proposed by Franklin and Zweidler (1977) with the total number of residues listed at the end of each sequence. Stars indicate residues identical to the corresponding position in the first variant listed for each histone class. Hyphens indicate gaps introduced to maximize homology. The histone fold domain within each histone class is indicated by the horizontal line above the sequences. Note that discrepancies exist between the nucleotide sequence data and protein sequence data for some entries. All residues are listed here as identified in the nucleotide sequence database. Underlining indicates residues identified differently in a compilation published previously (Fig. 4-15 of van Holde, 1989). The specific discrepancies are: H2A.3 = A10V, A66G, K99R, H124E, A126V, G128S; H2B.1 = T4A; H3.3 = S31A, A87S. The sequences used correspond to the following genbank accession numbers: H2A.1 = gi109975; H2A.2 = gi121970; H2A.3 = gi121978; H2A.X = gi2118979; H2A.Z = gi7949045; H2B.1 = 51303; H2B.2 = gi122028; TH2B = gi1518338; H3.1 = gi90623; H3.2 = gi51319; H3.3 = gi90624; CENP-A = gi3929334.
same 102 amino acid sequence has been found in H4 from all species of mammals examined (Makalowska et al., 1999). Note however that divergence is apparent at the nucleotide sequence level. More than ten H4 genes with different nucleotide sequences, but which encode identical proteins, have been characterized in humans (Doenecke et al., 1997). Heterogeneity of H4 on AU and AUT gels is usually attributable to acetylation or phosphorylation (the fastest migrating form of H4 in Fig. 1.8 is likely to be unmodified H4 while the slower band is likely to represent monoacetylated H4).

Sequences retrieved from the histone sequence database corresponding to non-allelic variants of H2A, H2B and H3 in mice are shown in Fig. 1.9. Note that, with the exception of the variants H2A.X and H2A.Z, in general, only a limited number of amino acid substitutions distinguish the variants within a histone class and that these occur at similar frequencies within and outside of the histone fold domain. As noted by underlining in Fig. 1.9, some discrepancies (listed in the legend to Fig. 1.9) exist between entries in the database and results from amino acid analyses and sequencing of histone peptides reported by Zweidler and others (Franklin and Zweidler, 1977; Urban et al., 1979; Zweidler, 1980; Hayashi et al., 1980; reviewed in Zweidler, 1984). In general, fewer substitutions are recorded for sequences in the database so it is possible that the actual differences between variants may be greater than depicted in Fig. 1.9.

Histone H2A from many eukaryotic cells exhibits a high degree of heterogeneity on both AU and AUT gels due to the presence of molecules bearing charge-altering modifications that include phosphorylation, acetylation, ubiquitination, and combinations thereof (see section 1.2; reviewed in Zweidler, 1984). In addition, more than five non-allelic variants of H2A, whose unmodified forms are resolved on AUT gels, have been described in mammals. These include the "somatic variants" H2A.1, H2A.2, H2A.3, H2A.4 (formerly H2A.S), H2A.X and H2A.Z, that are expressed in a variety of tissues (Bonner et al., 1980; West and Bonner, 1980; Zweidler et al., 1984; van Holde, 1989) and putative spermatocyte-specific variants that have not been well characterized to date (Trostle-Weige et al., 1982). As shown in Fig. 1.9, the sequences of H2A.X and H2A.Z have diverged more than H2A.1-H2A.3, and due to an extension and a deletion in the
C-terminal of H2A.X and H2A.Z, respectively, these forms can be resolved from each other and from the other H2A variants on both AU and SDS gels due to the differences in size and composition (West and Bonner, 1980). In mice, H2A.1 and H2A.2 appear to be replication-dependent forms whereas H2A.3, H2A.X and H2A.Z are considered to be replacement forms (Zweidler, 1984; Doenecke et al., 1997). Note that in some cases the distinction between replication-dependent and independent genes may be difficult to make. In proliferating human cells, the H2A.X gene gives rise to two transcripts, a 1600 base polyadenylated one and a 575 base transcript that lacks polyA. In the presence of inhibitors of DNA synthesis, levels of the shorter transcript diminish while the longer transcript is still expressed (Mannironi et al., 1989).

In many eukaryotes, fewer non-allelic variants have been described for H3 than for H2A. However, identification of forms of H3 resolved on AU and AUT gels is similarly complicated by acetylation, phosphorylation and combined acetylation and phosphorylation that give rise to forms of H3 with altered mobilities. Three non-allelic somatic variants, H3.1, H3.2 and H3.3, whose unmodified forms are resolved on AUT gels have been described in mammals. Proteins equivalent to H3.2 and H3.3 are expressed in a wide range of eukaryotes and the sequences of these variants have been absolutely conserved in fish, birds and mammals (Zweidler, 1984). H3.3 is a replacement variant whereas expression of H3.2 is replication-dependent (Wu and Bonner, 1981). H3.1 is a replication-dependent form that has been suggested to occur in mammals but not in other vertebrates (Zweidler, 1984). In addition to these somatic H3 variants, a spermatocyte-specific variant, H3.S, reported to be expressed in mice and rats, but not other mammals, has also been described (Zweidler, 1984).

Although greater differences are likely to exist between H2B sequences of different species than for H3 sequences (Fig. 1.7), in general, fewer non-allelic variants of H2B have been described in many species compared to H3. In the absence of post-translational modification, H2B from many organisms runs as a single band in AUT gels. Two somatic variants, H2B.1 and H2B.2 are expressed in mice (Zweidler, 1984). A spermatocyte-specific variant H2B.S thought to differ from H2B.1 by a deletion at the N-terminus and several conservative amino acid
substitutions has also been described (Zweidler, 1984) but the sequence of this protein has not
been established. Homologs of H2B.1 and H2B.S do not appear to be expressed in rats or other
mammals (Zweidler, 1984), suggesting these are unique to the mouse lineage. A different
spermatocyte-specific variant, TH2B, has been described in other mammals (Shires et al., 1975;
Meistrich et al., 1985).

Given that AUT gels, in some instances, can resolve non-allelic variants of core histones
that differ by a single neutral amino acid substitution (e.g. mouse H3.1 and H3.2, Figs. 1.8 and
1.9), it is not surprising that the resolution of electrophoretic forms of core histones displaying
characteristics of allelic polymorphisms has also been described. However, there appears to be
only one instance in which the nature of the polymorphism was unequivocally established. Three
allelic variants of sea urchin (Parechinus angulosus) sperm H2B, designated H2B₁, H2B₂ and
H2B₃, were resolved on AUT gels and shown to be expressed differentially among individual sea
urchins collected at different sites (Strickland et al., 1981). All individuals contained H2B₁ and in
addition either H2B₂ or H2B₃, or a mixture of both. Amino acid sequencing of these variants
revealed that a repeating pentapeptide, PTKRS, in the N-terminus of H2B₁ was mutated to
PRKGS in H2B₂ and H2B₃, and that the latter two were distinguished from each other by the
number of times the mutant pentapeptide was reiterated.

Zweidler (1984) reported that electrophoretic polymorphism of the spermatocyte-specific
variants H1.S and H2B.S in a colony of mice established from a wild population was attributable
to allelic variation since some individuals expressed equal amounts of the two electrophoretic forms
(heterozygotes) whereas other individuals expressed only one of the two forms (homozygotes).
Evidence of electrophoretic polymorphism of H2A.1 that appears to be specific to the ICR line of
BALB/c mice, and of electrophoretic polymorphism of H2A.2 that occurs in different species
including mice, pigs and horses are due to allelic variation has also been described (Zweidler,
1984). Allelic variation was confirmed to be the cause of the electrophoretic polymorphism of
H2A.1 in BALB/c ICR mice by analysing the distribution of the two electrophoretic forms,
H2A.1a and H2A.1b, among progeny of crosses and backcrosses between BALB/c ICR mice and
another inbred line, C57BL/6, shown not to possess the H2A.1b allele (Zweidler, 1984). The molecular basis of the polymorphism has not been reported for any of these cases but may involve substitutions of neutral or ionizable residues since the allomorphs were detected according to altered mobility in AUT gels.

1.3.3 Evidence for distinct functions of core histone variants

Direct evidence that non-allelic variants of core histones perform specialized functions is limited and at present, the variants can be grouped according to the strength of the evidence suggesting they have distinct functions. Convincing evidence is available for variants displaying more pronounced differences in amino acid sequence such as H2A.X, H2A.Z, macro-H2A. and CENP-A, that have made it possible to generate antisera permitting the specific detection and fractionation of chromatin containing these variants.

Considerable evidence suggests a specific role for homologs of mammalian H2A.Z (Hatch and Bonner, 1988), including H2A.vD of Drosophila (van Daal et al., 1988), H2A.F/Z in chickens (Harvey et al., 1983) and hv1 in Tetrahymena (White et al., 1988). This protein is the product of a single copy gene that contains an intron and whose mRNA is polyadenylated in all the cases above. H2A.Z is enriched in transcriptionally active chromatin in mammals (Gabrielli et al., 1981), as is hv1 in Tetrahymena (Allis et al., 1982; Stargell et al., 1993). The importance of this variant is underscored by the finding that H2A.vD is essential for development in Drosophila (van Daal and Elgin, 1992). The homolog pht1 in the fission yeast S. pombe is not essential for viability but deleted strains grow slowly, with altered colony morphology, increased resistance to heat shock and decreased fidelity of minichromosome segregation, suggesting a role in chromosome segregation in addition to transcriptional regulation (Carr et al., 1994).

The variant H2A.X has a C-terminal extension that is conserved among species. H2A.X expression can be detected throughout the cell cycle but the H2A.X gene is unlike other replacement-type histone genes in that it lacks introns and gives rise to 2 mRNA species, a shorter, non-polyadenylated, replication-dependent form containing an 3’ inverted repeat typical of
replication-dependent histone mRNAs in addition to a longer polyadenylated mRNA which also contains the inverted repeat that is expressed in the absence of DNA replication (Mannironi, et al., 1989; Mannironi et al., 1994). Due to the C-terminal extension (Fig 1.12), H2A.X is the longest of the H2A proteins in all species examined. The sequence of this extension varies somewhat among species but the serine equivalent to position 139 in mammals is absolutely conserved (Mannironi et al., 1989). Recently it been shown that this serine is rapidly phosphorylated to form γ-H2A.X in mammalian cells following treatments that introduce DNA double-strand breaks (Rogaku et al., 1998). Ser139 phosphorylation in response to ionizing radiation is detected in diverse species including Xenopus, Drosophila and yeast (Rogaku et al., 1999), demonstrating that both H2A.X and the pathway responsible for this modification have been conserved during evolution. Using a laser microbeam to induce DNA damage, Bonners’ group has shown that small foci of γ-H2A.X form at sites of damage and spread to occupy megabase chromatin domains (Rogaku et al., 1999). The close spatial correlation observed between sites irradiated and initial γ-H2A.X foci suggest that Ser139 phosphorylation may be an early event in DNA repair processes. Formation of γ-H2A.X also appears to be an early event in apoptosis (Rogaku et al., 2000).

MacroH2A was initially discovered as a component of rat liver chromatin. This H2A variant contains a large C-terminal extension that has motifs resembling DNA-binding and leucine zipper domains found in transcriptional regulators and thus may have arisen from the fusion of an H2A gene with a transcriptional regulator (Pehrson and Fried, 1992). The function of macroH2A is unknown but presence of the C-terminal DNA-binding motif suggests the possibility of sequence-specific binding that could potentially contribute to nucleosome positioning while the presence of the leucine zipper motif suggests a possible role in mediating interactions between nucleosomes containing macroH2A and transcriptional regulators or with other macroH2A containing nucleosomes. MacroH2A is surprisingly abundant in rat liver, it is estimated that there is one molecule of macroH2A for every 30 nucleosomes (Pehrson and Fried, 1992).
CENP-A is an H3 variant specifically localized in centromeric heterochromatin (Palmer et al., 1991; Kingwell and Rattner, 1987). The histone fold of human CENP-A, which is very similar to that of mammalian H3, is required for targeting to centromere, while the function of the more divergent N-terminal tail is unknown (Sullivan et al., 1994). CSE4, the apparent homolog in S. cerevisiae is essential and the finding that non-lethal mutations in CSE4 lead to non-disjunction of chromosomes and mitotic arrest suggest it is required for normal separation of chromosomes (Stoler et al., 1995). The histone fold of CSE4 is conserved with that of CENP-A, but N-terminal domains of the two molecules are quite different. These differences may explain observations that CENP-A will not rescue cse4 mutants and that CSE4 does not localize to human centromeres. Presumably, the N-termini of these variants interact with proteins that are distinct in yeast and humans.

No direct evidence for specialized roles is available for core histone variants displaying lesser degrees of amino acid sequence heterogeneity such as H2A.1, H2A.2, H2A.3, H2B.1, H2B.2, H3.1, H3.2 and H3.3. Differences in the content of these core histone variants among tissues and during development, like those shown in Fig. 1.8, have long suggested that these proteins play a role in determining, or are required to maintain, differences in transcriptional or replicative activity among cell types. These variants could have particular roles if regulatory factors recognized features attributable to the amino acid sequence differences of these variants, or if their incorporation into chromatin leads to nucleosomes with altered properties. However, antisera specific for these variants, like those generated for CENP-A and H2A.X, are not available and it is not known whether any of these forms are distributed in chromatin in specific fashion. Moreover, it may not be feasible to generate antisera capable of distinguishing these variants in situ, and the possibility that they are functionally distinct may be better examined using molecular genetic approaches (e.g. gene knockout, expression of epitope tagged forms).

1.3.4 Allelic and non-allelic variants of linker histones

Linker histones have diverged to a greater extent than core histones during evolution (Fig. 1.7) and chromatographic and electrophoretic analyses have revealed that non-allelic variants of
linker histones exist in most eukaryotes. Here, the ability of various methods to resolve non-allelic variants of H1 are reviewed and typical results achieved with two of the most popular techniques are presented. Unfortunately, workers in this field have not adopted a systematic nomenclature to classify H1 variants. Typically, variants are indicated by numbers or letters that reflect their relative chromatographic retention or electrophoretic migration. Because variation in experimental procedures between laboratories and the divergence of H1 among different species can complicate identification of homologous variants in different species, no attempt will be made here to correlate results obtained in different species and the nomenclature used to distinguish H1 variants in the original reports will be employed throughout this thesis.

Species-specific and tissue-specific heterogeneity of H1 was first discovered using Bio-Rex 70 chromatography, prior to the advent of modern electrophoretic methods (Luck et al., 1958; Kinkade and Cole, 1966; Bustin and Cole, 1968; Kinkade, 1969; reviewed by Cole, 1984, 1987). Optimal resolution of H1 proteins on this cation-exchanger (also known as Amberlite IRC-50) requires elution with very shallow gradients of the chaotropic salt guanidinium hydrochloride (GuHCl) in sodium phosphate buffer (typically 0.1 M) at neutral pH (typically pH 6.8) (see Cole, 1992 for a detailed description of the method). Because such shallow gradients are required, many reported H1 fractionations on Bio-Rex 70 required a week, or more, to complete. Typically, four to six components are apparent in chromatograms of H1 from metazoans. In the case of H1 extracted from pooled chicken erythrocytes (i.e. similar to that employed here in Chapter 3), typical chromatograms in which five components were partially resolved in Bio-Rex 70 fractionations lasting 6 or 21 days are shown in Fig. 1.10A & B, respectively (Kinkade, 1969; Dupressoir and Sautierre, 1984). Amino acid sequencing of individual Bio-Rex 70 components of bovine and rabbit H1 by the Cole lab and others revealed that in these species, different H1 variants differed to a greater extent than did core histones from different species (reviewed by Cole, 1977, 1984, 1987). This marked intra-species sequence heterogeneity, coupled with an abundant literature documenting tissue-specific differences in the number and relative abundance of Bio-Rex 70 components among different tissues within an organism, has long suggested that different H1 variants may be functionally distinct. An example of such tissue-specific heterogeneity is shown in
Figure 1.10  Resolution of non-allelic variants of chicken H1 by cation-exchange chromatography

(A) Crude H1+H5 from chicken liver, spleen, and erythrocytes prepared by selective extraction in 5% TCA was chromatographed on a column (2.3 x 26 cm) of Bio-Rex 70 using a linear gradient of 7-14% GuHCl in 0.1 M phosphate buffer pH 6.8. Columns were eluted at room temperature at a flow rate of 12 ml/h for the entire 1700 ml gradient volume, corresponding to a total run time of 5.9 days. Note that H5 is retained more strongly than H1 and is not eluted here. (Figure reproduced from Kinkade, 1969 with permission)

(B) Preparative separation of chicken erythrocyte H1 variants. One gram of the H1 + H5 mixture extracted from chicken erythrocyte chromatin with 5% HClO4 was applied to a 2.5 x 200 cm column of Bio-Rex70 and eluted with a 4000 ml linear gradient from 9 to 11% GuHCl in 0.1M phosphate buffer pH 6.8. A flow rate of 8 ml/h was used, corresponding to a run time (H1 resolving part only) of 20.8 days. The five "purified subfractions" H1A-H1E recovered are indicated by the shaded areas. H5 was eluted by a step to 40% GuHCl in the fraction marked 11. (Figure reproduced from Dupressoir and Sautierre, 1984 with permission)

(C) Resolution of chicken erythrocyte H1 variants on a Mono S FPLC column using Bio-Rex 70-type eluents. H1 and H5 prepared by acid extraction were resolved using an initial round of chromatography on a Mono S column (5.0 mm x 5.0 cm) using a steep gradient of GuHCl. One mg of the H1 protein recovered was reapplied to the column and variants resolved using a gradient from 4.5 to 5% GuHCl in 50 mM phosphate buffer pH 6.8 at a flow rate of 1 ml/min, corresponding to a total run time of 26 min. Figure reproduced from Shannon and Wells, 1987 with permission)
Fig. 1.10A where differences in the relative abundance of the components resolved by Bio-Rex 70 chromatography are apparent when the profiles of chicken liver, spleen and erythrocyte H1 are compared.

Other cation-exchangers have been used to purify H1 proteins but in general, these procedures do not resolve H1 subtypes as well as Bio-Rex 70 chromatography. Chromatography on CM-Sephadex C-25 using a gradient of sodium chloride readily separated chicken erythrocyte H1 proteins from the erythrocyte-specific variant H5 but did not resolve the H1 subtypes to any degree under the conditions employed (Garcia-Ramirez et al., 1990). Using similar materials but employing a much shallower gradient, others were able to resolve chicken H1 into two peaks (Harborne and Allan, 1986). Similarly, chromatography of perchloric acid-extracted crude steer thymus H1 on Mono S resin using a sodium chloride gradient separated H1 from HMG proteins, but provided only partial resolution of the H1 subtypes in this sample (Rice and Cole, 1990). Separation of two peaks of chicken H1 that were well resolved from H5 when salt-extracted histones were eluted from phosphocellulose with a gradient of sodium chloride at pH 9.0 has also been reported (Garg and Reeck, 1998). The efficacy of Bio-Rex 70 chromatography to resolve H1 subtypes may be due in part to the effects of GuHCl on the conformation of H1 proteins since four peaks were partially resolved within 30 minutes when a gradient of GuHCl was used to elute chicken erythrocyte H1 from a Mono S column (Shannon and Wells, 1987), as shown in Fig. 1.10C.

Purification of H1 by hydrophobic-interaction chromatography (HIC) using a phenyl-agarose support, as part of a “non-denaturing” strategy intended to maintain the activity of H1 in repressing transcription in vitro by RNA polymerase II, provided H1 free of contamination by other proteins present in extracts of HeLa cells and calf thymus but did not resolve any of the H1 subtypes of these species (Croston et al., 1991).

Size-exclusion chromatography under acidic conditions has also been used to fractionate histones. Although this technique does not resolve the common H1 subtypes (e.g. H1a, H1b,
H1c, ... etc.), it has been used to separate the erythrocyte-specific variant H5 from the remaining chicken erythrocyte H1 proteins (Von Holt et al., 1989; Dupressoir and Sautierre, 1984; also see Chapter 3). Similarly, chromatography on Bio-Gel P100 in 10 mM HCl has been used to separate H1°, a homologue of H5, from H1 subtypes in extracts of mammalian cells (e.g. Pehrson and Cole, 1980).

More recently, high-performance liquid chromatography (HPLC) techniques have been applied to resolve H1 variants. To date, the majority of reports have described the use of reversed-phase HPLC (RP-HPLC). However, the resolution of H1 variants by RP-HPLC is often lower than that achieved by Bio-Rex 70 chromatography or electrophoretic methods. H1 samples exhibiting electrophoretic heterogeneity eluted from RP-HPLC columns as a single peak (Gurley et al., 1983a, 1983b, 1984; Hallenbeck and Mueller, 1984; Helliger et al. 1988), although two to five peaks have been resolved depending on the sample and the chromatographic method (Certa and von Ehrenstein, 1981; Kurokawa and MacLeod, 1985; Wurtz, 1985; Lindner et al., 1986a, 1986b, 1988; Ohe et al., 1986; Karhu et al., 1988; Tchouatcha-Tchouassom et al., 1989; Lindner et al., 1990; Helliger et al., 1992; Lindner et al., 1992a, 1992b; Giancotti et al., 1993). Although the volatile mobile phases routinely employed in this method facilitate subsequent biomolecular characterization, the limited analyses performed in many of these studies make it difficult to assess whether any H1 variants were purified to homogeneity. Frequently, one or more of the H1 peaks isolated by RP-HPLC were shown to be electrophoretically heterogeneous (Gurley et al., 1983b, 1984; Helliger et al., 1988; Kurokawa and MacLeod, 1985; Lindner et al., 1986b, 1988; Tchouatcha-Tchouassom et al., 1989; Lindner et al., 1990, 1992a, 1997; Giancotti et al., 1993).

Electrophoretic techniques differ somewhat in their ability to resolve the non-allelic variants of H1 of different species. Excluding species such as Tetrahymena and D. melanogaster which appear to express only a single H1 polypeptide (Wu et al., 1986; Croston et al., 1991), and in the absence of post-translational modifications, H1 from most eukaryotes migrates as a set closely-spaced bands in SDS gels. In general, the resolution of non-allelic variants of H1 in SDS gels is not complete: only 2-3 bands are resolved for H1 prepared from tissues such as calf thymus and...
chicken erythrocytes known to express six H1 subtypes based on genetic analyses (Coles et al., 1987) or microsequence analyses of fractions isolated by Bio-Rex 70 chromatography (Cole, 1987). In avian species, the erythrocyte-specific linker histone H5 is sufficiently smaller in size that it is well resolved from the common H1 subtypes on SDS gels (e.g. Fig. 3.1) as are the related H1° proteins found in certain mammalian tissues (Smith and Johns, 1980).

Non-allelic variation of H1 is also usually apparent on AU and AUT gels commonly employed in histone analyses. As discussed above for core histones, charge-altering post-translational modifications of H1 (e.g. phosphorylation, poly(ADP-ribosylation) result in forms with altered mobility in AU and AUT gels (e.g. Balhorn and Chalkley, 1975) and appropriate controls are required to distinguish these from unmodified forms that differ in amino acid sequence. In contrast to the core histones, Triton X-100 does not appear to affect the resolution of non-allelic variants of H1 (Lennox and Cohen, 1989), and most analyses of H1 employ AU gels. The extent to which H1 subtypes are resolved on AU gels varies with the species and tissue under investigation, in addition to the conditions employed for electrophoresis (i.e. acrylamide and urea concentrations). In some cases, the number of bands resolved for H1 on AU gels is less than the number of components apparent in Bio-Rex 70 chromatography. For example, although 5 non-allelic variants of calf thymus H1 are partially resolved by Bio-Rex 70 chromatography, the 2-3 bands resolved in AU gels are so closely spaced that they can appear to be a single broad band (Smerdon and Isenberg, 1976). In contrast, AU-PAGE under optimal conditions separates six well-resolved bands for chicken erythrocyte H1, (e.g. Shannon et al., 1987; Fig. 4.1) whereas lengthy chromatography on Bio-Rex 70 incompletely resolves only 5 peaks (e.g. Fig. 1.10A & B; Kinkade, 1969; Dupressoir and Sautiere, 1984).

Two-dimensional electrophoresis has been employed to analyse both linker and core histones. The extremely basic nature of the histones (pI 10-12) precludes the use of conventional isoelectric focusing (IEF) since the histones migrate into the cathodal electrolyte during focusing. Although conditions can be employed in nonequilibrium pH gradient electrophoresis (NEPHGE) that keep histones on these gels, the resolution of non-allelic histone variants and their modified
forms is less than that achieved in AU and AUT gels (O'Farrell et al., 1977). Thus unlike the IEF-SDS and NEPHGE-SDS procedures frequently employed in cell biology, two-dimensional electrophoresis of histones commonly employs a first dimension AU or AUT gel and a second dimension SDS gel (AU-SDS or AUT-SDS gels). Many variations of the technique have been described, including AU-AUT, SDS-AU and SDS-AUT gels (Davie, 1982). A technique in which the second dimension AU gel includes the cationic detergent cetyltrimethylammonium bromide (CTAB) (AUC gel) has also been described (Bonnet et al., 1980).

Lennox and Cohen (Lennox et al., 1982; Lennox and Cohen, 1983, 1984a, 1984b) have used two-dimensional electrophoresis on AU-SDS gels to systematically analyse the expression and metabolism of non-allelic variants of H1 in mice. Depending on the sample, as many as 8 nonphosphorylated and 15 phosphorylated components of mouse H1 were resolved in these studies. Changes in the expression and turnover of the nonphosphorylated components that were apparent in electropherograms of various tissues taken from mice at different stages of development support the possibility that non-allelic variants of H1 are functionally distinct. An example of the resolution obtained with this technique is shown in Fig. 1.11 where clear differences in the expression of the non-allelic variants H1α and H1α were observed between dividing and quiescent 3T3 cells (Lennox and Cohen, 1984). Although the speed, resolution and sensitivity of this method clearly surpass that of Bio-Rex 70 chromatography, the small size of the samples employed have precluded the use of this technique in structural analyses to date.

H1 genes have been cloned in a wide variety of species and it is thought that the complete set of somatic (i.e. not tissue-specific) H1 variants has been identified in chickens, humans and mice (Nakayama et al., 1993; Albig et al., 1996; Wang et al., 1997). The amino acid sequences of the mouse somatic variants H1.1-H1.5 and the testes-specific H1t, as predicted by the respective genes and aligned for maximum homology, are shown in Fig. 1.12A. The corresponding variants (H1α-H1e) resolved by two dimensional electrophoresis (Lennox and Cohen, 1983) are also shown (Wang et al., 1997). In Fig 1.12B, the sequence of mouse H1.2 is shown aligned with that of the more divergent variant H1α, a replacement form expressed in a variety of differentiated
Figure 1.11. Resolution of mouse linker histones by 2-D PAGE.

H1 proteins prepared from dividing (A) and quiescent (B) cultures of murine 3T3 cells were resolved by two-dimensional electrophoresis (50 cm long AU gel first dimension, 20 cm long SDS gel second dimension) and detected by silver staining. The positions of the unmodified forms of six non-allelic variants, H1a, H1b, H1c, H1d, H1e and H10 are labeled in both A and B. Note the decreased content of H1a and increased content of H10 in non-dividing cells relative to dividing cells. Unlabeled components in A are phosphorylated forms of H1. (Figure reproduced from Lennox, 1984 with permission)
H1.2 (c) SEAPAAPAA APPA**AP*K **AA*K**P* -**P GM*R**A**P*P **TK**A* A**S**T**A** G**Y**V**A**A**A**N** S**S**E**R**G**V**S**L**A**A**L**K**S**L**A**A**A** G**Y**D**V**E**K**N**N**S**R**

H1.3 (d) **A**P**P** APPAA**AP*K **AA*K**P* -**P G**M**R**A**P**A**P** **TK**A* A**S**T**A** G**Y**V**A**A**A**N** S**S**E**R**G**V**S**L**A**A**L**K**S**L**A**A**A** G**Y**D**V**E**K**N**N**S**R**

H1.4 (e) **A**P**P** APPAA**AP*K **AA*K**P* -**P G**M**R**A**P**A**P** **TK**A* A**S**T**A** G**Y**V**A**A**A**N** S**S**E**R**G**V**S**L**A**A**L**K**S**L**A**A**A** G**Y**D**V**E**K**N**N**S**R**

H1.5 (b) **A**P**P** APPAA**AP*K **AA*K**P* -**P G**M**R**A**P**A**P** **TK**A* A**S**T**A** G**Y**V**A**A**A**N** S**S**E**R**G**V**S**L**A**A**L**K**S**L**A**A**A** G**Y**D**V**E**K**N**N**S**R**
Figure 1.12. Amino acid sequences of mouse linker histones.

The amino acid sequences of non-allelic variants of mouse H1 predicted from nucleotide sequences contained in the non-redundant set of the histone sequence database as of May 2000 are shown in one letter code after alignment using ClustalW 1.8 (Higgins et al., 1996). (A) The sequences of the "somatic" and the testes-specific H1 variants are compared. (B) The sequence of the "extreme" variant H1° is compared to that of the shortest somatic variant, H1.2. Genes are identified according to the nomenclature (H1.1-.5) proposed previously (Drabent et al., 1995). The corresponding proteins are identified by the nomenclature (a-e) based on 2-D electrophoresis (Lennox and Cohen, 1983) as described previously (Wang et al., 1997). The total number of predicted residues are listed at the end of each sequence. Stars indicate residues identical to the corresponding position in the first sequence listed. Hyphens indicate gaps introduced to maximize homology. The approximate location of the globular domain is indicated by the horizontal line above the sequences. The sequences correspond to the following Genbank accession numbers: H1.1 = gi1170153; H1.2 = gi121905; H1.3 = gi1170155; H1.4 = gi1170151; H1.5 = gi1170154; H1t = gi631776; H1° = gi2118973.
H1 proteins, prepared by HClO4 extraction of nuclei isolated from erythrocytes of individual Japanese quail (Coturnix coturnix japonica) were resolved on SDS gels (22 cm long, 13.5% acrylamide, 0.36% N,N'-methylenebisacrylamide) and detected by staining with Coomassie brilliant blue R-250. Non-allelic variants H1.1 - H1.5 are designated arbitrarily according to SDS gel mobility. Quail phenotypes H1.3+, H1.3- and H1.3+/.3- are designated arbitrarily according to the intensity of staining of the H1.3 band. Note the apparent absence of a band corresponding to H1.3 in the H1.3- individual. Subsequent analyses on two-dimensional gels revealed the presence of an allelic variant of H1.3 in H1.3- and H1.3+/.3- individuals that comigrates with H1.4 in the gel shown here. (Figure reproduced from Palyga, 1991a with permission)
cell types (Doenecke et al., 1997). Note that while some differences exist in the globular domains of H1.1-H1.5, the majority of differences between these variants involve conservative substitutions within the N- and C-terminal domains (Fig. 1.12A). Insertions or deletions in the C-terminal domain account for much of the difference in the lengths of these proteins. Greater variation is found in H1t where numerous conservative and non-conservative substitutions are found within the globular domain (Fig 1.12A). An even greater degree of divergence is apparent in the sequence of H1° (Fig. 1.12B). Numerous conservative and non-conservative substitutions are found in the globular and flanking

The partial resolution on Bio-Rex 70 of two allelic variants of chicken H5 differing by a single Gln to Arg substitution (Greenaway and Murray, 1971) represents the first instance in which allelic variation was demonstrated in a histone. Subsequently, it was suggested that polymorphisms encountered during the sequencing of variants of calf and rabbit H1 prepared by Bio-Rex 70 were due to allelic variation (Cole, 1987). Aside from these reports, I am not aware of any instance in which allelic variants of linker histones were detected by liquid chromatography methods. However, numerous reports describe the detection of allelic variants of H1 by electrophoretic methods. Stout and Philips (1973) observed that 2 forms of the major H1 protein of maize resolved by AU-PAGE were independently inherited in typical Mendelian fashion in crosses of inbred strains. The difference(s) in amino acid composition responsible for the different AU gel mobilities of the two forms were localized to the carboxyl-terminal domain of the molecules but were not defined unequivocally at the level of the amino acid sequence (Hurley and Stout, 1980). Subsequently, the existence of allelic variants of linker histones with altered mobility in SDS gels, AU gels, or both, was described in a variety of animals including salmon (Zalenskii et al., 1981), rabbits (Palyga, 1990a), Xenopus (Dworkin-Rastl et al., 1994) mice (Zweidler, 1984) several avian species (Gorel’ et al., 1982; Palyga, 1990b, 1991a, 1991b, 1991c, 1998a, 1998b; Palyga et al., 1993, 2000; Gornicka-Michalska et al., 1998; Kowalski et al., 1998). A particularly striking example is shown in Fig. 1.13 where allelic polymorphism affecting the H1 variant H1.3 of Japanese quails was detected when the H1 profiles of individual birds on SDS gels were compared (Palyga, 1991). The molecular basis for polymorphism has not been established in any
of these cases although evidence has been presented that polymorphism affects the N-terminal domain of H5 (Gornicka-Michalska et al., 1998) and the C-terminal domain of duck H1b (Palyga et al., 2000) and duck H1.z (Palyga et al., 1993).

A recent analysis of the H1 gene locus (sw3) in the trypanosome Leishmania major revealed that this gene is polymorphic in different strains of this protozoan, giving rise to proteins that differ by deletion of nine residues in the central portion of the protein (Belli et al., 1999). Like the macronuclear H1 from Tetrahymena (Wu et al., 1986), Leishmania H1 lacks the conserved sequence corresponding to the globular core of metazoan H1s. The functional significance of the polymorphism is not known.

1.3.5 Evidence for distinct functions of linker histone variants

Differences in the abundance of forms of H1 resolved by Bio-Rex 70 chromatography or electrophoretic methods that correlate with the differentiated phenotype or metabolic activity of cells, like those shown in Figs. 1.10A and 1.11, have been interpreted by many researchers as evidence that non-allelic variants of H1 fulfill specialized roles in regulating transcription, and possibly other activities, of chromatin. This literature has been reviewed elsewhere (Hohmann, 1983; Cole, 1984, 1987; Lennox and Cohen, 1984, 1988). Here, apart from citing key reports in this older literature, I will focus on studies attempting to address this issue in more direct ways.

The ability to prepare chromatographic fractions representing individual purified H1 variants or defined mixtures of specific variants has facilitated analyses to determine if these fractions impart distinct properties to the products formed upon reconstitution with DNA or chromatin in vitro. Cole and co-workers have systematically studied the physical properties of H1:DNA complexes and chromatin reconstituted with chromatographic fractions of H1 prepared by Bio Rex-70 chromatography. Welch and Cole (1979) reported that four fractions of rabbit H1 differed by as much as 2-fold in their ability to distort the circular dichroism (CD) spectrum of linear T7 bacteriophage DNA and the degree to which they decreased the intrinsic viscosity of solutions of this DNA. A subsequent study revealed that fractions of bovine thymus H1 differed
significantly in their ability to distort the CD spectra of both linear (T7) or superhelical (PM2 bacteriophage) DNA (Liao and Cole, 1981a). The changes in the CD spectra observed (formation of ψ-type spectra) were proposed to reflect the condensation of DNA into ordered aggregates.

Analyses of the salt dependence of the H1:DNA interactions revealed that changes in spectra were not observed at 0 or 300 mM NaCl for any of the H1 fractions, while significant degrees of condensation of both linear and superhelical DNA that differed for each H1 fraction were observed at 150-200 mM NaCl, correlating with attainment of maximal α-helical content in each H1 fraction.

Since much of this α-helical structure can be attributed to the folding of the globular domain of H1 (Ramakrishnan et al., 1993), these results suggest that the differences in the condensing power of the H1 fractions may derive in part from differences in the interaction of their respective globular domains with DNA or that differential condensation of DNA by the more variable flanking N- and C-terminal domains of the variants is dependent on the folding of the globular domain. Although H1:DNA complexes are intuitively not expected to be accurate models for the interaction of H1 with DNA in chromatin, similar salt-dependent differences were observed in the ability of these same fractions of H1 to distort the DNA CD spectrum when they were reconstituted with dinucleosomes previously depleated of H1 (Liao and Cole, 1981b).

More recently, three purified rat H1 variants and a fraction containing two variants prepared from extracts of testes using RP-HPLC were shown to differ in their ability to condense H1-depleted nucleosome oligomers (approx. 10-25 nucleosomes) as measured by distortion of the DNA CD spectrum (De Lucia et al., 1994). The testes-specific H1 variant H1t, found to be enriched in soluble chromatin prepared by DNase I digestion, was the least effective in condensing nucleosome oligomers, leading these authors to suggest a role for H1t in maintaining genes required during spermatogenesis in an open conformation to facilitate transcription. Additional evidence that H1 variants differ in their interactions with nucleosomal DNA comes from a recent analysis of the binding affinity of mouse variants H1a, H1b, H1c, H1d/e, H1o and H1t purified by RP-HPLC for nucleosomes reconstituted on a 210 bp fragment of the MMTV promoter (Talasz et
al., 1998). The affinity of H1b for this model nucleosome, as assessed by gel shift assays, was less than one half that of the other variants tested (dissociation constant for H1b was approx. 8-16 nM versus 2-4 nM for the other variants). Additionally, diminished cooperativity in the binding of H1b and H1t was suggested to be the basis for the relative inability of these variants to aggregate a model hexanucleosomal 1.3 Kbp MMTV promoter reconstitute.

Chromatin oligomers reconstituted with mouse H1° were digested by micrococcal nuclease more rapidly than oligomers reconstituted with mouse H1-1 or H1-2 (Biard-Roche et al., 1982). The increased accessibility of linker DNA in H1°-containing reconstitutes correlated with thermal denaturation analyses which suggested that H1° protected less DNA than other variants. This may be related to the smaller size of H1° compared to other H1 variants. These results were somewhat unexpected as it had been shown previously that H1° was enriched (relative to bulk chromatin) in chromatin resistant to digestion when isolated nuclei were treated with micrococcal nuclease (Gorka and Lawrence, 1979). Together these findings support the notion that linker histones contribute to chromatin folding by mechanisms in addition to charge shielding of linker DNA such that H1° is preferentially associated with more condensed chromatin. Despite the similarities between H1° and H5, it was found that oligonucleosomes containing H5 (similar to H1°) displayed greater resistance to thermal denaturation than did chromatin reconstituted with chicken erythrocyte or calf thymus H1 (Allan et al., 1981). It was also concluded that H5 protected slightly more DNA than did H1 at early stages of micrococcal nuclease digestion. These discrepancies suggest may be attributable to differences in the interactions of H5 and H1°, respectively, with nucleosomal or linker DNA or possibly to differences in the methods employed. Nonetheless, they demonstrate that characteristics of reconstituted chromatins are determined in part by the type of linker histone present.

Analyses of salt-dependent chromatin condensation revealed that nucleosome oligomers were condensed to a greater degree (assayed by sucrose gradient centrifugation) when reconstituted at various stoichiometries with chicken H5 compared to chicken erythrocyte H1 (containing six somatic variants) (Hannon et al., 1984). Moreover, initiation by bacterial RNA polymerase was
repressed to a greater degree in oligomers reconstituted with H5 compared to H1, suggesting that the greater degree of condensation produced by H5 hindered template access by the polymerase to a greater extent than the condensation produced by H1. In contrast, a recent study found that a greater percentage of nucleosome oligomers reconstituted with chicken erythrocyte H1 (containing 6 somatic variants) were aggregated and precipitated by 0.15 M NaCl compared to oligomers reconstituted with H5 (Nagaraja et al., 1995). Similarly, trout H1b facilitated aggregation of oligomers in 0.15 M NaCl to a greater degree than trout H1a. The apparent contradictions in these studies for results with H5 versus chicken H1 may be attributable to the analyses employed (e.g. sedimentation versus precipitation) or other experimental differences (e.g. length of the chromatin fragments employed). In addition, Nagaraja et al., (1995) found that longer chromatin fragments were preferentially precipitated when reconstitutes were made 0.15 M NaCl with the size of the chromatin remaining soluble inversely related to the amount of linker histone added. Strikingly, transcriptionally competent oligomers (revealed by hybridization analysis for an erythroid-specific gene) were enriched in the low molecular weight fraction remaining soluble in 0.15 M NaCl regardless of the type of linker histone added. Unfortunately, the histone content of the soluble and precipitated fractions were not compared in this experiment but based on previous observations (see section 1.2.1 and below), it seems likely that the soluble fraction was depleted of H1.

Decreased solubility of oligonucleosomes containing H1 at physiological ionic strength compared to those depleted of H1 has been used in chromatin fractionation schemes to enrich for transcriptionally active chromatin (section 1.2.1). Although numerous reports document that such fractions are depleted of H1, surprisingly few have characterized the H1 proteins that are present in these fractions. Analyses of long chromatin oligomers (containing 1,500-20,000 bp of DNA or approx. 8-110 nucleosomes) recovered from micrococcal nuclease-digested bovine thymus nuclei upon lysis at low ionic strength revealed that oligomers which remained soluble upon addition of 120-300 mM NaCl contained much less H1 than those rendered insoluble in the same experiment (Huang and Cole, 1984). Significantly, it was shown that the size distributions of the soluble and insoluble fractions was equivalent, indicating that chromatin precipitation was not a function of fragment size. SDS gel analyses revealed that the relative abundance of the two bands, H1a/b and
H1c, resolved for bovine H1 differed between the soluble and insoluble fractions such that the relative content of H1c (i.e. the ratio of H1c to H1a/b from densitometric scans of stained gels) was greater in the soluble fraction. H1c corresponds to the chromatographic fraction of H1 shown previously to resist salt-induced condensation when reconstituted with naked DNA or dinucleosomes (Liao and Cole, 1981a,b). Conversely, the relative abundance of H1a/b was greater in oligomers that were precipitated at 120-300 mM NaCl and H1a is the major component of the chromatographic fraction of bovine H1 that promoted salt-induced condensation of H1:DNA and H1:dinucleosome complexes more than other forms of H1. Consistent with previous reports on active chromatin enriched based on solubility (section 1.2.1), it was also found that chromatin which remained soluble at 120 mM NaCl was enriched in acetylated core histones and displayed greater sensitivity to DNase I digestion compared to the insoluble fraction (Huang and Cole, 1984). Because redistribution of H1 to generate a new soluble fraction was not observed when chromatin precipitated at 120 mM NaCl was redissolved and incubated at 0-50 mM NaCl prior to reprecipitation at 120 mM NaCl, the authors concluded that the H1 depletion observed in soluble fractions was attributable to an uneven distribution of H1 in the original chromatin rather than a redistribution of H1 among chromatin fragments during handling. This implies that the distributions of H1c and H1a/b in the original chromatin must have differed such that H1c represented a greater percentage of the H1 present in domains containing lesser amounts of H1 than bulk chromatin and which resist salt-induced condensation. Taken together with the findings that DNA and dinucleosomes reconstituted with H1c resisted salt-induced condensation compared to reconstitutes containing other H1 variants (Liao and Cole, 1981a,b), these data suggest that both the amount and identity of the H1 variant, together with factors such as content of acetylated histones, influence the extent to which regions of chromatin are condensed in vivo. The discrepancies with respect to the size distributions of the soluble and insoluble chromatin fractions generated from “native” chromatin as described by Huang and Cole (1984) and those generated from reconstituted chromatin as described by Nagaraja et al., (1995) may be attributable to the significantly lower stoichiometry of H1 in the soluble fraction from native chromatin or possibly due to differences in the mode of binding by H1 reconstituted with chromatin compared to that of endogenous H1.
Given the range of fragment sizes in the chromatin fraction employed by Huang and Cole (1984) and the fact that the size distribution was unaffected by salt-induced precipitation, it is not possible to infer from their data the size of fragments over which zones of H1 depletion and increased relative abundance of H1c are propagated or whether a mean fragment length corresponding to such zones exists. However, it is important to note that the apparent distributions of some H1 variants may be influenced by the size range of the chromatin fractions employed. Using brain chromatin fractions with size distributions similar to those described for thymus chromatin above, Jin and Cole (1985) found that the distributions of H1c and H1a/b in the aggregatable and aggregation-resistant fractions of brain chromatin mirrored those determined for thymus chromatin, but that the variant H1° was found to be distributed equivalently in the two fractions. Because others have found that H1° was nonrandomly distributed with respect to specific genes (discussed below), the data suggest that mechanisms exist to facilitate the preferential localization of H1 variants at the level of nucleosomal arrays that may correlate to chromatin loops or domains (section 1.1.5) and also within smaller spans representing single genes or elements such as promoters, introns and exons.

Several approaches have been used in attempts to determine directly whether H1 variants are evenly dispersed in chromatin. Due to a suspected role in the regulation of specific genes during cellular differentiation, the distribution of H1° has been analysed by several groups. Exploiting an electrophoretic method that resolved a class of mononucleosomes containing only H1° apart from classes devoid of linker histones and from those containing a mixture of H1 variants, Roche et al., (1985) separated chromatin recovered from micrococcal nuclease digests of nuclei from adult mouse liver. Hybridization analyses with 32P-labeled probes determined that H1° was preferentially associated with nucleosomes containing sequences corresponding to the repressed α-fetoprotein gene whereas nucleosomes containing the expressed albumin gene were found to be preferentially associated with nucleosomes containing H1 variants other than H1°. These findings suggested a role for H1° in transcriptional repression, and in particular for the developmental repression of α-fetoprotein transcription. However, other studies have shown that
H1° may be associated with active chromatin in some cases. Using a similar methodology, Delabar (1985) compared the association of H1° with various types of DNA sequences in nuclei. The content of H1° in chromatin containing non-expressed genes such as globin in liver and albumin in brain was found to be lower than that of bulk chromatin whereas H1° was enriched approximately two-fold in liver nucleosomes containing expressed genes and it was suggested that H1° is preferentially associated with transcriptionally competent chromatin. The basis for the contradictory conclusions of these two studies is not apparent but additional studies suggest that H1° is not completely excluded from active chromatin. Using antisera immobilized on beads, Mendelson et al., (1986) fractionated rat liver oligonucleosomes solubilized with micrococcal nuclease into H1°-enriched and H1°-depleted fractions and the content of specific DNA sequences in these fractions assayed by hybridization with 32P-labeled probes. These authors concluded that similar amounts of H1° were associated with non-transcribed and constitutively transcribed genes but that highly inducible genes such as cytochrome P-450 in liver were depleted of H1°. These observations are supported by the localization of H1° to both condensed perinucleolar chromatin and decondensed chromatin in immunoelectronmicroscopy (Gorka et al., 1993). Thus, while each of these studies present evidence that H1° is nonrandomly distributed in chromatin, the fact that the presence of this variant does not correlate strictly with transcriptional repression or activation underscores the notion that factors in addition to the distribution of H1 variants play a role in transcriptional regulation.

Employing Bio-Rex 70 chromatography in combination with RP-HPLC, Parseghian et al., (1993) resolved H1° and four fractions of H1 from extracts of human placental nuclei. Polyclonal antisera generated against the protein recovered from the fraction designated H1-3 were specific for this fraction in ELISA and Western blotting analyses and preferentially stained chromatin at the periphery of interphase nuclei in immunofluorescence microscopy, suggesting that this variant was distributed nonrandomly within chromatin. Attempts to generate antisera specific for H1-1, H1-2 and H1-4 using intact proteins recovered from these fractions as antigens were unsuccessful. However, these workers subsequently generated antisera specific for H1-1 and H1-2 using synthetic peptides corresponding to the first 32-35 residues of these proteins and antisera specific
for H1-4 was obtained using the amino terminal fragment recovered following N-bromosuccinimide cleavage of this protein (Parseghian et al., 1994). Immunofluorescence analyses of the distribution of these variants in cultured human fibroblasts suggested that the distribution of H1-1 was parallel to that of DNA whereas both H1-2 and H1-4 were distributed nonrandomly, giving punctate staining patterns that were distinct from that described previously for H1-3. Together, these results are consistent with the findings described above based on chromatin immunofractionation that some H1 variants are preferentially associated with distinct regions of the genome.

Recombinant DNA methodology has enabled H1 variant function to be studied in vivo in cells in which the normal stoichiometry of H1 variants has been perturbed through the transfection or disruption of genes for specific variants. Overexpression of avian H5 in rat sarcoma cells led to selective repression of certain genes in an H5 dose-dependent fashion and a transient inhibition of cell proliferation (Sun et al., 1989). Chromatin from cells overexpressing H5 was shown to resist digestion by micrococcal nuclease (Sun et al., 1990). These results are consistent with the notion that H5 condenses chromatin and functions as a repressor, as suggested by the accumulation of H5 during nuclear condensation that accompanies the terminal stages of avian erythropoiesis (Affolter et al., 1987). However, since these results were not compared to cells in which other linker histones were overexpressed, it was not possible to attribute these effects to specific features of H5 as opposed to overexpression of linker histones in general. Brown et al., (1996) compared the effects of overexpression mouse H1c and H1° in 3T3 cells stably transformed with these genes under the control of a mouse metallothionein promoter. Upon induction, the levels of these variants were increased 3-5 fold over that in wildtype cells. This was only partially compensated for by decreased expression of other H1 variants such that the overall levels of linker histones, expressed as the mole ratio of the sum of all linker histones to H2B rose from 0.8 to 1.1-1.3. Despite this increase in linker histone stoichiometry, no changes in viability or morphology were observed for cells overexpressing H1c or H1°. However, cultures overexpressing H1° exhibited transient inhibition of G1 and S-phase progression when released from synchronization at different points in the cell cycle. Overexpression of H1c to comparable levels had no effect on cell cycle
progression. Quantitation of mRNAs corresponding to several cell cycle-regulated and housekeeping genes revealed that expression of all these genes was significantly reduced in cells overexpressing H1\(^{\alpha}\). On the other hand, overexpression of H1c resulted in either no significant change or dramatic increases in the levels of these transcripts. These data provide strong evidence that functional differences exist between these two H1 variants.

Subsequently, these authors attempted to identify the features of H1c and H1\(^{\alpha}\) responsible for the difference in their effects on transcription by expressing a series of chimeric molecules containing N-terminal, C-terminal and globular domains from either H1c or H1\(^{\alpha}\) (Brown et al 1997). These analyses revealed that the differential effect of overexpression of H1c and H1\(^{\alpha}\) on transcription of the genes analysed was due primarily to differences in their globular domains. This provides further evidence for functional differences between H1c and H1\(^{\alpha}\), but given that the globular domain is much more highly conserved in H1a-e than it is between these variants and H1\(^{\alpha}\) (see Fig. 1.12), this raises the question of whether any differences can be observed if the effects of overexpression of H1a-e are compared. Such an analysis has not been reported, but recent experiments discussed below involving targeted disruption of H1 genes in mice and chickens have provided evidence for both differences and redundancy in the function of H1 variants.

As discussed above, H1\(^{\alpha}\) is unique among H1 variants in mammals because of its greater degree of sequence divergence and the fact that its expression is not cell cycle-regulated. There is much indirect evidence that H1\(^{\alpha}\) performs distinct functions in modulating chromatin transcription during cellular differentiation (reviewed in Zlatanova and Doenecke, 1994). Thus it was surprising that disruption of the single copy H1\(^{\alpha}\) gene did not affect mouse development and mice completely lacking H1\(^{\alpha}\) grew and reproduced normally with no anatomic or histologic abnormalities (Sirotkin et al., 1995). Quantitation of the H1 variants expressed in the livers of wildtype and H1\(^{\alpha}\)/- mice revealed that the loss of H1\(^{\alpha}\), which normally represents approx. 30% of total H1 in liver, was compensated for by equivalent increases in the expression of variants H1a-e. Thus, the relative abundance of H1a-e or the stoichiometry of total H1 in liver chromatin were not significantly altered in H1\(^{\alpha}\)/- mice. Although unexpected, this suggests that one or more of the variants H1a-e
can compensate for the loss of H1°.

A similar finding has recently been reported for H1.1. Disruption of the single copy H1.1 gene had no effect on mouse development and even though H1.1 is the most abundant somatic variant in testes of wild type animals, spermatogenesis and reproduction were normal in mice completely lacking H1.1 (Rabini et al., 2000). Quantitation of the levels of H1 proteins and their mRNAs in testes revealed that the loss of H1.1 was compensated for by increased expression of H1.2, H1.3 and H1.4 with no significant changes in the expression of H1t, H1.5 or H1°. It is not known presently whether the selective nature of this compensation is biologically significant.

The results of the H1.1 and H1° knockouts in mice suggest that at least some H1 variants share degrees of functional redundancy. Additionally, they reveal the existence of a mechanism regulating total H1 levels in chromatin but do not provide any information regarding the nature of the mechanism. However, results from systems in which multiple H1 variant genes have been disrupted suggest that there are limits to both the degree of total H1 deficiency that can be compensated for and the extent to which H1 variants are functionally redundant. Analyses of mutants generated from the DT40 chicken B cell line revealed that a single copy of an H1 gene, out of the 12 copies present in diploid cells (there are 6 single copy non-allelic variants per haploid chicken genome) was sufficient for cell proliferation (Takami et al., 1997). However, the levels of H1 mRNA and protein in these cells were approximately one half that observed in wild type cells and cells containing two copies of the same H1 variant gene (encoding H1d). Global changes in chromatin structure were not detected when micrococcal nuclease digests of nuclei from cells containing only one H1 gene were compared to those from wild type cells, but comparisons of 2D electropherograms of total cellular protein revealed differences in the abundance of numerous proteins. Subsequent comparisons of mutant DT40 lines in which the genes for all six H1 variants had been individually disrupted revealed that specific changes in the abundance of various proteins resolved in 2D electropherograms of total cellular protein accompanied the loss of each particular variant (Takami et al., 2000). Although increased expression of some proteins upon the loss of a particular H1 variant is evidence that H1 acts as a repressor, the finding that the expression of other
proteins was repressed in the same cells suggests that H1 may facilitate transcription at other loci. These findings are similar to the gene-specific effects observed upon the loss of H1 or mutation of H1 phosphorylation sites in Tetrahymena (Shen and Gorovsky, 1996; Dou et al., 1999). Together these findings support the notion that H1 can act as a positive or negative regulator of transcription in gene-specific fashion and that differences in the structure, and possibly post-translational modification, of H1 variants may be involved in determining the specificity or magnitude of these effects.

1.4 Scope of the thesis

A large body of evidence supports the hypothesis that non-allelic variants of H1 fulfill specialized roles in regulating chromatin condensation and transcription. When the work described in this thesis was initiated, nearly all of the evidence supporting this hypothesis was indirect, based on tissue-specific and developmental stage-specific patterns of H1 variant expression detected by chromatographic and electrophoretic analyses. Furthermore, little was known about the potential roles that phosphorylation or other post-translational modifications of H1 may play in regulating chromatin transcription. Thus, I sought to test the hypothesis that particular variants or modified forms of H1 are associated with transcriptionally active chromatin. In principle, this hypothesis could be tested directly by determining the abundance of specific variants and modified forms of H1 in isolated fractions of nuclease-sensitive/transcriptionally active chromatin compared to bulk chromatin provided that the conditions used did not promote redistribution of H1. Two factors influencing the potential success of this approach are the resolution and sensitivity of the method used to resolve H1 variants. An ideal method would be capable of separating all amino acid sequence variants and post-translationally modified forms of H1 and possess sensitivity that was sufficient to permit analyses of small samples generated in chromatin fractionations. Furthermore, if the same method could be used to prepare forms of H1 for molecular characterization and for use in chromatin reconstitutions, this would permit direct tests of the importance of specific features of variants or modified forms to be made. Unfortunately, disadvantages associated with common methods for H1 analyses made them unsuitable for one or both of these tasks. SDS and AU gel electrophoresis are well suited for analyses of small samples, but the resolution of variants and
modified forms of H1 on SDS gels is incomplete and the enhanced resolution attainable on AU gels requires cumbersome large format gels and lengthy electrophoresis times. Moreover, neither technique is well suited for the preparation of proteins for molecular analyses or use in chromatin reconstitutions. The resolution and sensitivity of Bio-Rex 70 chromatography are inadequate and the lengthy chromatography times employed increase the likelihood of proteolysis and loss of modifications, hindering molecular characterization. Although RP-HPLC is more sensitive, the resolution of variants and modified forms of H1 is usually incomplete. Moreover, the acidic conditions typically employed may limit the suitability of conventional RP-HPLC for the preparation of proteins for use in chromatin reconstitution and analyses of acid-labile putative modifications of H1 (e.g. N-linked phosphorylation). Given these limitations of existing methods, I sought to develop a relatively rapid and sensitive HPLC method to separate non-allelic variants and modified forms of H1 with the goal of using this method to determine the occurrence and distribution of various forms of H1 in chromatin.

Several factors made chicken erythrocyte H1 the best choice for a model sample to use in developing the HPLC method. Chicken erythrocyte chromatin has been intensively studied and simple, rapid procedures existed for preparing nuclei, chromatin and histones from chicken blood with minimal likelihood of proteolysis. Furthermore, numerous reports have been published describing the resolution of non-allelic variants of chicken H1 by various techniques, facilitating evaluation of any observations made here. Most importantly, more information regarding the structure of H1 variants was available for the chicken compared to any other higher eukaryote. The sequences of six genes, each encoding a distinct non-allelic variant, and estimated to represent the total H1 complement in chickens, were published before this work was initiated (Coles et al., 1987). Additionally, the six protein bands resolved for chicken H1 on AU gels had been aligned with these six gene sequences (Shannon and Wells, 1987), suggesting that chromatographic fractions of chicken H1 recovered from experimental trials could be unequivocally identified based solely on AU gel mobility. Inadequate resolution of variants on AU gels and incomplete sequence data precluded the application of this simplified process of identification to other species. Because little or no transcription is detectable in mature chicken erythrocytes (Gasaryan, 1982), it was
intended that work to investigate potential links with transcriptional regulation would be performed with chicken liver or cultured cells once the forms of chicken erythrocyte H1 separated by the optimized HPLC technique had been characterized.

All modes of HPLC compatible with protein chromatography were tested and optimized conditions for the resolution of chicken H1 variants were identified in most cases. The results of these studies are summarized in Chapter 3. This work led to the development of a novel chromatographic method, referred to as cation-exchange-hydrophilic-interaction chromatography (CX-HILIC), also described in Chapter 3, that resolved 14 major and numerous minor components for chicken erythrocyte H1 obtained from a sample of pooled blood. This degree of chromatographic heterogeneity was unexpected given that a total of six H1 genes had been identified and that only 3 and 6 bands were resolved for this sample of H1 on SDS and AU gels, respectively. Because of the unprecedented nature of this finding, the focus of the work shifted from testing the original hypothesis to elucidating the physical basis for the heterogeneity of chicken H1 in CX-HILIC and these efforts represent the remainder of this thesis. Characterization of the major CX-HILIC peaks of pooled erythrocyte H1 by AU gel electrophoresis (Chapter 4) revealed that most peaks represented one of the six H1 variants described previously, purified to apparent homogeneity, but that four variants (and possibly more) were represented by multiple peaks. With only one notable exception, mass spectrometry revealed that the molecular mass of each CX-HILIC form of H1 was smaller than expected based on the gene sequences previously identified to correspond to the six variants (Chapter 4). These discrepancies were not attributable to proteolysis during handling since the molecular mass of the major form of H5 purified by CX-HILIC and processed in parallel was identical to the expected value (Chapter 5). Although H5 was also surprisingly heterogeneous in CX-HILIC, the finding that the CX-HILIC profiles of H5 prepared from pooled blood and H5 from a single chicken were very similar suggested that much of the heterogeneity was attributable to post-translational modifications, as suggested by the detection of forms of H5 with molecular masses that were larger than predicted by the gene sequence (Chapter 5). Chromatographic analyses and mass spectrometry of H1 variants prepared from chicken liver and from blood of a single chicken revealed that allelic variation, rather than
tissue-specific expression or proteolysis of H1 variants, was responsible for most of the CX-HILIC heterogeneity of H1 from pooled blood (Chapter 6). However, the detection of subsets of H1 variants that appeared to be missing the first four amino acid residues in a subset of CX-HILIC peaks suggested that limited proteolytic processing of H1 does occur, possibly in a regulated fashion, in vivo (Chapter 6). Finally, mass spectrometry and microsequencing of peptides generated to confirm the correlation between the H1 variant proteins and the six H1 genes published previously revealed that the genes encoding two variants were identified incorrectly in the original report (Chapter 7).
Chapter 2. Materials and Methods
2.1 Isolation of chicken erythrocyte nuclei

Pooled chicken blood, comprised of approximately 10 ml of blood from each of 100 adult animals, was obtained at a local poultry packer. Blood was collected into an equal volume of ice-cold 154 mM NaCl, 20 mM Na-EDTA pH 8.0, 1.0 mM phenylmethylsulfonyl fluoride (PMSF) pooled with other samples and kept on ice for transport to the laboratory. Nuclei were prepared by detergent lysis of erythrocytes as described previously (Olins et al., 1976). Nuclei to be stored at -70°C were washed twice in TSM buffer (10 mM Tris-HCl pH 7.5, 50 mM NaCl, 3 mM MgCl₂) and frozen as a loosely-packed pellet in a minimum volume of TSM diluted 1:1 with glycerol.

2.2 Histone extraction

Standard methods for extracting histones were used. All steps were performed at 4°C. Total histone was extracted from freshly prepared or frozen nuclei with 10 volumes (relative to the nuclear pellet) of 0.4 N H₂SO₄ and core histones selectively precipitated by the addition of perchloric acid (PCA, 5% (w/v) final). The crude H1/H5 mixture was recovered from the 5% PCA supernatant by trichloroacetic acid (TCA) precipitation (20% (w/v) final). Protein precipitates were washed twice with acetone containing 0.1% (v/v) HCl, twice with acetone, dried under vacuum and stored dry at -70°C. All data shown here for H1 and H5 from pooled blood was obtained using histone prepared in this fashion. Alternate methods of preparing crude H1/H5, including direct extraction of H1 and H5 from nuclei with 5% perchloric acid followed recovery by TCA precipitation as above, or extraction with 0.6 M NaCl containing 0.5 mM PMSF, gave equivalent results. Solutions of acid-extracted histones, dissolved in dH₂O, were stored at -70°C. Salt-extracted H1/H5 were dialysed against 150 mM NaCl, 0.5 mM PMSF, 10 mM sodium phosphate pH 7.0 and stored at -70°C.

2.3 Preparation of H1_total and H5_total from pooled blood

The standard sample of H1_total for HPLC method development was prepared by gel filtration of the acid-extracted crude H1/H5 mixture from pooled erythrocyte nuclei. Two columns (5.0 cm i.d. x 100 cm) packed with Bio-Gel P100 (100 - 200 mesh; Bio-Rad) were eluted in series
at 50 ml/h with 10 mM HCl, 3 mM NaN₃ at 22°C after application of crude H1/H5 in a minimum volume of 10 mM HCl to the bed of the lead column. Elution was monitored at 218 nm. Fractions were analysed by PAGE and those containing H1 (or H5 in the case of H5̄̅ in the case of H5 total) with little or no significant contamination by other proteins were pooled and the protein recovered by TCA precipitation. In the case of salt-extracted crude H1/H5 from pooled blood, H1 total was prepared by chromatography on Sephadex CMC-25 (Pharmacia) as described previously (Harborne and Allan, 1986).

2.4 Polyacrylamide gel electrophoresis

Acrylamide, bis-acrylamide and sodium dodecyl sulfate (SDS) were from Bio-Rad. Electrophoresis in 12% and 15% (w/v) polyacrylamide gels containing SDS (SDS-PAGE) was performed according to Laemmli (1970). Electrophoresis in 50 cm long 15% (w/v) polyacrylamide gels containing 2.5 M urea and 0.9 N acetic acid (AU-PAGE) was performed as described by Lennox and Cohen (1989). Gels were stained with Coomassie Blue R-250 or silver nitrate (Morrissey, 1981).

2.5 HPLC

Trifluoroacetic acid (TFA) was from Pierce (Sequanal grade). HPLC-grade sodium perchlorate and acetonitrile were from Fisher Scientific and VWR Scientific, respectively. Water (dH₂O) was prepared with a Milli-Q deionization system (Millipore). All other chemicals employed were ACS grade or better.

Separations were performed at room temperature on a Waters binary gradient chromatograph or a Beckman System Gold quaternary gradient chromatographs with computerized data collection and instrument control. Protein elution was monitored at 214 nm unless noted otherwise. Fraction collectors (Gilson) programmed to collect fractions on a time basis were programmed to function in the appropriate portions of chromatograms based on the results of trial runs. Throughout this work, the concentrations of organic modifiers (e.g. CH₃CN) and ion-
pairing reagents (e.g. TFA) employed are expressed as v/v.

Analytical format columns representative of the various types of RP-HPLC supports available commercially were tested for their ability to resolve $H_{1\text{total}}$. The salient properties distinguishing each column, based on information supplied by the manufacturers are summarized in Table 3.1. Unless otherwise noted, a standard injection of 100 µg of $H_{1\text{total}}$ in 100 µl of buffer A (0.1% TFA in dH$_2$O or 0.1% TFA in 5% CH$_3$CN) was employed and columns were eluted with a linear gradient from 0 to 100% buffer B (0.093% TFA in 60% CH$_3$CN) over 2 hours at 0.8 ml/minute. Proteins were recovered from RP-HPLC eluents by solvent evaporation using a Speed-Vac apparatus (Savant Inc., Farmingdale, NY, USA).

The properties of the CX-HPLC columns tested for the resolution of pooled erythrocyte $H_{1\text{total}}$ are summarized in Table 3.3. Columns were initially compared using injections of 100 µg $H_{1\text{total}}$ and 2 hour linear gradients with equivalent slopes generated by mixing buffer A (20 mM sodium phosphate pH 6.5) with buffer B (buffer A containing 2 M NaCl). Because different supports bound H1 with different strengths, the %B employed at the initial and final points of these gradients were adjusted on a case-by-case basis so that the earliest H1 component eluted at approximately 1 hour after injection. In addition to trials at pH 6.5, the resolution of $H_{1\text{total}}$ on strong cation-exchange supports was also compared at pH 3.0 in 20 mM sodium phosphate buffer. Compared to all columns tested, the PolyCAT A support resolved the greatest number of peaks in these tests and was selected for further method development as described in section 3.2.6. The optimized method, referred to as cation-exchange-hydrophilic-interaction chromatography (CX-HILIC) employed PolyCAT A supports with 30 or 100 nm diameter pores (depending on the sample) eluted with a linear gradient of sodium perchlorate (see figures) in 10 mM propionic acid containing 70% acetonitrile, adjusted to pH 6.5 with sodium hydroxide. CX-HPLC buffers containing more than 50% acetonitrile were prepared by dissolving all components except water in the desired volume of acetonitrile using a magnetic stirrer. Water was then added to bring the volume to 95% of the intended final volume and the solution allowed to come to room temperature.
(22°C). The pH was adjusted using a glass combination electrode before dH₂O was added to give the final intended volume. This method of preparation was adopted to avoid errors due to the effects of organic solvents on buffer ionization constants. Buffers were vacuum-filtered using a 0.45 μm pore diameter Durapore membrane (Millipore, Bedford, MA, USA) prior to use. Proteins were recovered from CX-HILIC eluents by TCA precipitation after evaporation of acetonitrile under vacuum in a Speed-Vac apparatus.

2.6 Circular dichroism spectroscopy

CD spectra were recorded on a J720 spectropolarimeter (JASCO) using a 0.1 mm pathlength quartz cell. Identical samples (80 μg) of pooled erythrocyte H₁₅O_dkd dissolved in microcentrifuge tubes were dissolved in 10 μl of water and then brought to a final volume of 200 μl with the appropriate solvents several hours prior to measurement. Ten scans were averaged for each sample.

2.7 Mass spectrometry

Electrospray mass spectrometry (ESMS) was performed at the University of Michigan Protein and Carbohydrate Structure Facility in collaboration with Drs. P. C. Andrews (Director) and R. Loo. Mass spectra were obtained on a Vestec model 201 single quadrupole mass spectrometer fitted with a Vestec electrospray interface. Lyophilized samples (10 μg, approx. 500 pmol) were dissolved in 1-4% acetic acid, 50% CH₃CN prior to analysis. Samples were introduced via a flow-injection system and analysed by standard procedures.

Ion spray mass spectrometry (ISMS) was performed at the University of Toronto in collaboration with Drs. H. Pang and M. Cheung. Mass spectra were obtained on a Sciex API III triple quadrupole mass spectrometer fitted with a Sciex ion spray interface. Lyophilized samples (10 μg, approx. 500 pmol) were dissolved in 1% acetic acid, 2 mM ammonium acetate, 50%
CH$_3$CN prior to analysis. Samples were introduced via a flow-injection system and analysed by standard procedures.

2.8 NBS cleavage procedures

The method of Sherod et al., (1974) was adapted for use with small quantities of protein and direct recovery of cleaved products by RP-HPLC. Aliquots corresponding to 10-20 ug of H1 fractions recovered from CX-HILIC were cleaved during initial trials to optimize the method. Cleavages employing 20-200 µg of CX-HILIC H1 fractions were performed for analyses by mass spectrometry. Samples containing approx. 500 pmol of the recovered N- and C-termini (3.4 and 7.6 µg, respectively) were submitted for mass spectrometry. N-bromosuccinimide (NBS) was ACS grade from Fisher Scientific. NBS powder and working solutions were shielded from light using aluminum foil to prevent photodegradation.

Aliquots of CX-HILIC H1 fractions dissolved in water were vacuum-dried in 1.5 ml polypropylene microcentrifuge tubes and redissolved in 5% acetic acid to give a final concentration of H1 of 1 mg/ml. An aliquot of 1/10 this volume of 4.5 mM NBS in 5% acetic acid (freshly prepared that day) was added and the reaction mixed briefly and incubated for 5 minutes at room temperature (22-25°C). This amount of NBS gives an approximate 10-fold molar excess over the single tyrosine residue reported to exist in each of the six chicken H1 non-allelic variants (Coles et al., 1987). The reaction was then terminated by adding an aliquot of 45.45 mM tyrosine in 0.1 M NaOH corresponding to 1/10 of the volume of the original H1 solution. The reaction tube was vortexed while adding the tyrosine solution to avoid tyrosine precipitation. The mixture was then allowed to stand in the dark for at least 5 minutes prior to clarification in a microcentrifuge (13,000 x g, 5 min.) and injection into the HPLC. Integration of RP-HPLC chromatograms indicates that approximately 80% of the sample was cleaved under these conditions. The NBS solution was prepared by dissolving 8.0 mg NBS in 10 ml of 5% acetic acid and kept in a sealed tube wrapped in foil. The tyrosine solution was prepared by dissolving 82.36 mg of tyrosine in 1 ml of 1.0 M NaOH and then adjusting the final volume to 10 ml with dH$_2$O.
Peptides of interest were recovered by injecting clarified reactions directly onto a Chromegabond MC-18 column (4.6 x 250 mm, 5 μ silica with 30 nm pores; E.S. Industries). The column was equilibrated with 95% buffer A/5% buffer B where A = 0.1% TFA in water and B = 0.094% in 60% CH₃CN prior to sample injection. Peptides were eluted using a gradient from 5% B to 100% B over 1 hour and the column washed with 100 %B for 10 minutes prior to equilibration at 5% B for subsequent separations. Chromatography was performed at room temperature on the Beckman System Gold chromatograph fitted with a photodiode array detector. A flow rate of 0.8 ml/min and absorbance at 214 was monitored to detect all peptides. N-terminal peptides were identified by the characteristic absorbancy at 260 nm of the dibromodienone spirolactone adduct of tyrosine formed upon polypeptide cleavage by NBS (Fontana and Gross, 1986).

2.9 α-chymotrypsin digestion procedures and microsequencing

Methods described previously for preferential cleavage at phenylalanine in histone H1 (Bradbury et al., 1975; Clark et al., 1988) were modified to facilitate work with smaller amounts of H1 and recovery of peptides by RP-HPLC. The time course of digestion in trials employing 200 μg of H1total were monitored on 16% SDS gels to identify conditions that maximized cleavage at the unique phenylalanine in the chicken H1 variants (residue 110 of H1α, Coles et al., 1987) with minimal cleavage at residues C-terminal to this to generate a full length C-terminal fragment (i.e. residues 111-224 of H1α). This approach greatly simplified identification of the appropriate peptide to be recovered for sequencing in RP-HPLC chromatograms. Digestions of H1 CX-HILIC fractions for sequence analysis employed 40 μg of purified H1 (approx. 2 nmols).

α-chymotrypsin (Boehringer Mannheim, sequencing grade) was dissolved in 1 mM HCl to make a 0.5 mg/ml “concentrated stock” that was kept on ice for up to one week as recommended by the manufacturer. Prior to use, an aliquot was diluted to 0.1 mg/ml in 1 mM HCl. H1,
dissolved at 1 mg/ml in 10 mM Tris-HCl pH 8.0 was prewarmed to 37°C and then digested with α-chymotrypsin using an enzyme:H1 ratio of 1:500 by weight at 37°C for 12 minutes. The reaction was then terminated by the addition of TFA to 0.2% (final) and cooling on ice.

Peptides of interest were recovered by RP-HPLC as described in section 2.9 except that a Chromegabond MC-18 column packed with 100 nm diam. pore silica was used and the gradient was from 5-70% B over one hour. Confirmation that the appropriate C-terminal fragment had been recovered was obtained by analysis of fractions in 16% SDS gels. C-terminal fragments were also analysed by ion spray mass spectrometry prior to submission for sequencing.

Aliquots representing approx. 100 pmol of the C-terminal fragments of CX-HILIC peaks 10 and 11 were sequenced using standard methods on an Applied Biosystems model 470 sequenator at the University of Michigan Protein and Carbohydrate Structure Facility in collaboration with Dr. P. C. Andrews (Director).

2.10 Preparation of H1 from non-erythroid chicken tissues

In order to prepare H1 from various tissues with minimal contamination from erythrocyte H1+H5, chickens were exhaustively perfused prior to organ removal. Two live adult chickens, each approx. 2.3 kg, were obtained from a local supplier and anesthetized initially with ketamine (400 mg/kg i.m.). Once light anesthesia was apparent, heparin (600 U/kg) was injected into the wing vein and the animal brought to surgical anesthesia by halothane inhalation. The thoracic cavity was then opened, the vena cava or right atria bisected, and the animal perfused via the left ventricle using a 60 ml syringe and 21G catheter. Approx. 200 ml of perfusate (20 mM Tris-HCl pH 7.4, 154 mM NaCl) containing 2U heparin/ml was required to produce maximal blanching of the liver and 400 ml total was administered to each animal as rapidly as possible while avoiding severe distension of the heart. The liver, spleen, heart and kidneys were then removed and kept on ice until extracted.
Tissues, free of fat and excess connective tissue, were weighed and then coarsely minced with scalpels in plastic weighboats over ice in a 4°C room. They were then extracted in 5 volumes (relative to weight) of cold 5% perchloric acid using two 30 second bursts of a Polytron homogenizer (Brinkman) fitted with a PT10ST or PT20ST probe depending on the extract volume, at maximum speed with 5 - 10 minutes cooling in an ice-water bath between bursts. Samples were then clarified by centrifugation at 10,000 x g for 20 min. at 4°C. The supernatants were transferred to new tubes and H1 recovered by TCA precipitation. TCA precipitated H1 was washed twice with acetone/0.1% HCl, twice with acetone, air dried and dissolved in dH2O prior to storage at -70°C. Extremely low levels of H5 in these samples, as detected on SDS and AU gels, demonstrated that the perfusion procedure had been effective in minimizing contamination of organ H1 by blood H1+H5.

2.11 Preparation of H1 and H5 from blood of single chickens

Essentially undiluted blood (30-50 ml) collected separately from each chicken early during the perfusions described in section 2.11 was kept on ice until nuclei were prepared separately from each sample by detergent lysis as described previously (Olins et al., 1976). Histones were extracted from isolated nuclei as in section 2.2. Crude H1+H5 (10 mg PCA soluble protein) were separated by gel filtration chromatography on Bio-Gel P100 as described in section 2.3 except that a single 1.5 x 170 cm column, eluted with 10 mM HCl, 0.02% NaN3, at a flow rate of 5 ml/hour, was used. Fractions (2.5 ml) were collected and protein elution determined by the optical density of fractions at 218 nm. Purified H1 and H5 were recovered by TCA precipitation of fractions pooled after characterization on 12% SDS gels. Analyses of the final H1 and H5 preparations on SDS and AU gels revealed only minor contamination by non-histone proteins with no detectable cross-contamination between the H1 and H5.

2.12 Assays for post-translational modifications of H1_total

To test the possibility that post-translational modifications including glycosylation, phosphorylation and poly(ADP-ribosylation) were responsible, in whole or in part, for the chromatographic heterogeneity of pooled blood H1_total in CX-HILIC, aliquots of H1_total were
treated using procedures designed to remove these modifications prior to CX-HILIC chromatography and the profiles of treated and control samples compared to reveal any potential changes in the abundance or retention of the 14+ peaks resolved for H1̅ total.

Phosphorylation was assayed for by incubating 100 μg of H1̅ total dissolved in 100 μl of 50 mM Tris-HCl pH 8.0, 1 mM PMSF, 1 μg/ml α-2-macroglobulin (Boehringer Mannheim) with 1.0 ul of bacterial alkaline phosphatase (Worthington) at 30°C for 2 hours. Purified, highly phosphorylated Tetrahymena macronuclear H1̅ (provided by C.D. Allis) was used as a positive control for the digestion. Significant dephosphorylation of Tetrahymena H1̅ by the treatment was confirmed by the appearance of the dephosphorylated form possessing a characteristic mobility on SDS gels (Lu et al., 1994). Neither the SDS gel electropherogram or the CX-HILIC profile of pooled chicken blood H1̅ total was altered by incubation with alkaline phosphatase compared to the control sample incubated in parallel without alkaline phosphatase.

The possible involvement of glycosylation was assessed by incubating 100 μg of H1̅ total dissolved in 200 μl of 250 mM sodium phosphate pH 7.5, 10 mM EDTA, 10 mM 2-mercaptoethanol, 1 mM PMSF and 1 μg/ml α-2-macroglobulin (Boehringer Mannheim) with 5.0 units glycopeptidase F (Boehringer Mannheim) for 16 hours at 37°C as suggested by the manufacturer. The SDS gel mobility of fetuin (20 μg, Sigma) incubated in parallel was used as a positive control and incubation with glycopeptidase F resulted in the formation of faster migrating species as expected upon loss of glycosyl moieties. The mobility of H1̅ total on SDS gels and retention in CX-HILIC were unaltered by this treatment, suggesting that N-linked glycosylation does not contribute to the heterogeneity of H1̅ total.

Poly(ADP-ribosylation) was assayed for by incubating 100 μg H1̅ total in 0.1 M NaOH at 56°C for 20 min. (Braeuer, et al., 1981). No differences were noted in the SDS gel
electropherogram or the CX-HILIC profiles of samples treated with NaOH compared to controls incubated in parallel in dH₂O. Prolonged treatment with NaOH appeared to result in non-specific degradation of H1 proteins as determined by smearing on SDS gels and generalized decreased height and increased width of H1 peaks in CX-HILIC. Attempts to assay for poly(ADP-ribosylation) using snake venom phosphodiesterase (Sigma) (Brauer et al., 1981; Aboul-Ela et al., 1988) were unsuccessful due to proteolytic activities contaminating the phosphodiesterase preparation that were not completely inhibited by 1 mM PMSF plus 1 μg/ml α-2-macroglobulin.
Chapter 3. Resolution of pooled erythrocyte $H_{I_{total}}$ by HPLC
3.1 Introduction

In this chapter, I describe the results of a survey of the resolution of chicken H1 variants attainable by various HPLC techniques. It is commonly acknowledged that analytes can interact with chromatographic supports by more than one mechanism (e.g. hydrophobic and electrostatic interactions may occur between a protein and an ion-exchange support) leading to so-called “mixed-mode effects” that contribute to the overall resolution of a chromatographic system (e.g. Hancock and Sparrow, 1981; Mant et al., 1987; Zhu et al., 1992). Mixed-mode effects can be particularly pronounced with macromolecules such as proteins which can interact with a chromatographic support through charged, hydrophobic and hydrophilic amino acid side chains simultaneously depending on the nature of the support, the amino acid sequence of the protein, and its secondary and tertiary structure under the conditions employed (e.g. Regnier, 1987). Thus, it was anticipated that a variety of eluent compositions would have to be tried with each support tested in order to adequately assess it’s applicability for resolving H1 variants.

To minimize the time required to test a wide variety of chromatographic methods, a large quantity of electrophoretically pure H1\textsubscript{total} (i.e., containing the six non-allelic variants of H1 resolved by AU-PAGE with minimal contamination by other proteins) was prepared for use as a standard sample in chromatographic trials. This eliminated the need for ancillary analyses (e.g. PAGE) to determine whether components separated in chromatographic trials represented H1 or other proteins present in crude extracts. Nuclei were isolated from a liter of pooled chicken blood and crude lysine-rich histone (H1+H5) was prepared by acid extraction. H1\textsubscript{total}, containing the AU gel subtypes H1\textsubscript{a}, H1\textsubscript{a’}, H1\textsubscript{b}, H1\textsubscript{c}, H1\textsubscript{c’} and H1\textsubscript{d}, but few other proteins as judged by SDS and AU-PAGE analyses, was prepared from crude H1+H5 by gel filtration chromatography. Following recovery and quantitation, H1\textsubscript{total} was employed as the standard sample in the various chromatographic trials described below. The resolution of H5\textsubscript{total}, recovered from this procedure was also examined in some cases.
3.2 Results and Discussion

3.2.1 Preparation and characterization of pooled erythrocyte $H_1_{total}$ and $H_5_{total}$

The separation of $H_1_{total}$ and $H_5_{total}$ from acid-extracted crude $H_1/H_5$ of pooled erythrocyte nuclei by preparative gel filtration chromatography on Bio-Gel P100 is shown in Fig. 3.1A. SDS-PAGE analyses of fractions from this column revealed that the first peak contained $H_1$ and that the second peak contained $H_5$ (Fig. 3.1B). Aliquots of fractions estimated to contain at least 5 ug of protein according to absorbance at 218 nm were analysed on SDS gels. This approach gave sensitive detection of contaminants at edges of peaks but resulted in overloading of $H_1$ and $H_5$ proteins in the central fractions of the peaks due to nonlinearity of spectrophotometric quantitation at high absorbance values. The trailing edge of the $H_1$ peak was contaminated to a minor extent by other polypeptides while the leading edge of the $H_5$ peak was contaminated by a protein with an approx. $M_r$ of 60,000. Small amounts of this protein eluted across the entire $H_5$ peak. Minor contamination of the later half of the $H_5$ peak by smaller proteins was also detected. Note that some separation of the $H_1$ variants resolved by SDS-PAGE was apparent within the $H_1$ peak with earlier fractions (58-70) enriched in the upper band relative to later fractions (100-110) that were enriched in the faster migrating doublet. To maintain the natural stoichiometry of $H_1$ subtypes present in the sample, fractions 56-108 were included in the $H_1_{total}$ pool despite the minor contamination apparent in fractions 106-108. Fractions 132-190 were pooled separately to recover the corresponding $H_5_{total}$ peak. After recovery by TCA precipitation, the $H_1_{total}$ and $H_5_{total}$ fractions were dissolved in $dH_2O$ and quantitated by absorbance at 218 nm. Approximately 268 mg and 766 mg of $H_1_{total}$ and $H_5_{total}$, respectively, were recovered. Working stocks of the proteins dissolved in $dH_2O$ at 10 mg/ml were kept at -70°C while aliquots of the remaining protein were lyophilized and stored as dry powders in tightly-sealed vials at -20°C.

3.2.2 Resolution by RP-HPLC

At the time this work was initiated, several reports had been published describing RP-HPLC separations of the five classes of histones obtained from a variety of organisms. As discussed in Chapter 7, these methods generally employed conventional mobile phases containing TFA and acetonitrile but used a variety of RP supports. In most instances, samples of $H_1$ from
Figure 3.1. Resolution of $H_{1\text{total}}$ and $H_{5\text{total}}$ from pooled chicken erythrocyte $H_1+H_5$.

(A) Approx. 1.2 g of PCA soluble protein extracted from pooled erythrocyte nuclei was dissolved in 50 ml of 10 mM HCL, 0.02% NaN$_3$, 1 mM PMSF and separated on two 5.0 x 100 cm columns of Bio-Gel P100 in series as described in Chapter 2.3. Eluent absorbance at 218 nm is plotted versus fraction number.

(B) Aliquots of fractions from (A) estimated to contain at least 5 ug of protein according to absorbance at 218 nm were analysed on 12% SDS gels. Proteins were detected by Coomassie blue staining. Based on these gels, fractions 56-108 were pooled for $H_{1\text{total}}$ and fractions 132-190 were pooled for $H_{5\text{total}}$. The yields of $H_{1\text{total}}$ and $H_{5\text{total}}$ after recovery were approx. 268 mg and 766 mg, respectively.
metazoans known to express variants comprising several bands in PAGE analyses, including chickens, eluted as a single peak (e.g. Gurley et al., 1984, Helliger et al., 1988). However, the separation of four peaks for mouse, rat and human H1 by RP-HPLC (e.g. Lindner et al., 1988; Lindner et al., 1990; Ohe et al., 1986) suggested the possibility that RP-HPLC could potentially separate chicken H1 variants if the right combination of chromatographic support and elution conditions were employed. Given the potential of this methodology for high sensitivity and ease of protein recovery from volatile mobile phases, considerable effort was invested in developing an RP-HPLC method to resolve the variants in pooled blood H1 \textit{total}. The effects that differences in support characteristics and mobile phase composition had on the resolution of H1 \textit{total} were examined systematically and are summarized below.

Table 3.1 describes the chief characteristics of the RP supports tested. Chromatography of H1 \textit{total} on the \mu Bondapack C18 column was unsuccessful (i.e. elution of H1 was not detected using a variety of mobile phases), presumably due to strong electrostatic interaction of the very basic H1 molecules with ionized silanols present on the surface of this support. Subsequent work confirmed that the recovery of H1 was highest from RP supports that were “endcapped”, i.e. treated with trimethylsilane or similar reagents after application of the primary bonded phase in order to minimize the content of ionizable silanols on the support surface. Consequently, only endcapped materials were selected for further method development.

The alkyl chain length of the bonded phase of RP supports appeared to have little effect on the resolution of chicken H1 subtypes. Similar resolution was achieved when H1 \textit{total} was chromatographed on the Hi-Pore RP-304 and Hi-Pore RP-318 columns using several different mobile phase systems. These columns are packed with the same 5 \mu diameter silica particles with 30 nm diameter pores that have been derivatized with butylsilane (C4) or octadecylsilane (C18) reagents, respectively, prior to endcapping (Table 3.1). In comparisons employing a 2 hour linear gradient from 20 to 60% CH3CN in 0.1% TFA, H1 \textit{total} eluted as a single asymmetric peak with multiple shoulders on the trailing edge from both supports (data not shown). Surprisingly, H1

88
was retained to a similar degree by both columns, eluting at approximately 32% and 34% CH₃CN from the C4 and C18 supports, respectively, under these conditions. Similar results were obtained with C8 columns supplied by two other manufacturers that also utilized 30 nm pore silicas (data not shown, Table 3.1).

Unexpectedly, the pore diameter of RP supports was found to affect the resolution of variants in H1_total significantly. When this work was initiated, the vast majority of RP-HPLC polypeptide fractionations reported in the chromatographic literature utilized supports based on 30 nm pore silicas, a trend which continues today. Nonetheless, a manufacturer (E. S. Industries, Table 3.1) able to provide columns packed with silicas with 30, 100 or 400 nm average diameter pores, that had been derivatized with C18 and endcapped equivalently, was located. As shown in Fig. 3.2, when H5_total and H1_total were co-injected, H5 proteins were better separated from H1 proteins on the 100 nm diameter pore support than they were the 30 nm diameter pore support using the same 2 hour linear gradient from 20 to 60% CH₃CN in 0.1% TFA used above to test the Bio-Rad columns. Moreover, subcomponents of both the H5 peak and the H1 peak were markedly better separated on the 100 nm pore support, appearing as distinct peaks in contrast to the poorly resolved shoulders obtained with the 30 nm pore support under identical conditions. The results shown in Fig. 3.2A for the 30 nm pore Chromegabond C18 support appear to be representative of 30 nm pore RP supports in general since they are similar to those obtained with the Bio-Rad Hi-Pore 304 and 318 (C4 and C18, respectively) columns discussed above and are also similar to those obtained with the Aquapore RP 300 and Vydac 218BTP54 C8 columns (Table 3.1, data not shown). Thus, although it is well known among chromatographers that differences in manufacturing procedures can result in significant differences in the selectivity of RP supports for certain analytes, only minor differences in the resolution of H1_total were observed for the 30 nm pore RP supports tested under the conditions employed.

Samples (50 µg) of H1_total were chromatographed on all the columns listed in Table 3.1 using a 2 hour linear gradient from 20 to 60% CH₃CN in 0.1% TFA. Thus, the entry "% CH₃CN
for $H_{1\text{total}}$ elution”, which lists the approximate value (rounded to the closest integer) of the % 
CH$_3$CN in the actual gradient coincident with the center of the main peak of $H_{1\text{total}}$ can be used as 
an approximate measure of the overall retention of H1 proteins by the RP columns tested. 
Surprisingly, this parameter varied little with the alkyl chain length of the bonded phase or the 
average pore diameter of the supports. Given that pore diameters affect the accessibility of the 
interior surfaces of pores to analytes and are inversely related to the chromatographic surface area 
of the support, these findings suggest that the dominant factor(s) determining the retention of 
$H_{1\text{total}}$ in RP-HPLC under these conditions are inherent to the H1 proteins rather than the supports. 
Because the composition of the mobile phase could influence the retention and resolution of H1 
proteins by multiple mechanisms, a wide variety of mobile phase compositions were tested as 
described below.

The ion-pair reagent TFA is typically used at a concentration of 0.1%, but advantages have 
been claimed for the use of higher concentrations of TFA in RP-HPLC of histones (e.g. Gurley et 
al., 1990), so the effect of concentration of TFA on the retention of $H_{1\text{total}}$ on several RP columns 
was examined. On both 30 nm and 100 nm Chromegabond MC18 supports, the retention of 
$H_{1\text{total}}$ was increased in 0.2% and 0.3% TFA relative to that in 0.1% TFA, but no differences in 
resolution were observed and baseline noise (at 214 nm) was increased markedly compared to that 
in 0.1% TFA (data not shown). Similarly, no advantages were associated with the use of a lower 
concentration (0.05%) TFA (data not shown).

After TFA, HFBA is possibly the next most-frequently used ion-pair reagent for 
polypeptide fractionations. The retention of $H_{1\text{total}}$ on the 30 nm and 100 nm diameter pore 
Chromegabond MC18 columns was greater in 0.1% HFBA compared to 0.1% TFA, but the 
separation of $H_{1\text{total}}$ from $H_{5\text{total}}$, and the resolution of the subcomponents of the $H_{5}$ and $H_{1}$ peaks 
noted in Fig. 3.2B was decreased in HFBA compared to that in TFA (data not shown). $H_{1\text{total}}$ and 
$H_{5\text{total}}$ coeluted from the 100 nm pore C18 column in 0.1% HFBA (data not shown).
<table>
<thead>
<tr>
<th>Name</th>
<th>Format</th>
<th>Particle type</th>
<th>Bonded phase</th>
<th>% CH$<em>3$CN for H1$</em>{total}$ elution</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>µBondapak C18</td>
<td>4.6 x 250</td>
<td>ir Si, 10 µ 8 nm</td>
<td>C18, pe</td>
<td>n.d.</td>
<td>Waters Associates</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bedford, MA, USA</td>
</tr>
<tr>
<td>Aquapore RP-300</td>
<td>4.6 x 250 (cartridge)</td>
<td>sp Si, 7 µ 30 nm</td>
<td>C8, e</td>
<td>32</td>
<td>P-E/Applied Biosystems</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Foster City, CA, USA</td>
</tr>
<tr>
<td>Hi-Pore RP304</td>
<td>4.6 x 250</td>
<td>sp Si, 5 µ 30 nm</td>
<td>C4, e</td>
<td>32</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hercules, CA, USA</td>
</tr>
<tr>
<td>Hi-Pore RP318</td>
<td>4.6 x 250</td>
<td>sp Si, 5 µ 30 nm</td>
<td>C18, e</td>
<td>34</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hercules, CA, USA</td>
</tr>
<tr>
<td>Vydac 218BTP54</td>
<td>4.6 x 250</td>
<td>sp Si, 5 µ 30 nm</td>
<td>C8, e</td>
<td>34</td>
<td>Separations Group</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hesperia CA, USA</td>
</tr>
<tr>
<td>Chromegabond MC-18</td>
<td>4.6 x 250</td>
<td>sp Si, 5 µ 30 nm</td>
<td>C18, e</td>
<td>35</td>
<td>E.S. Industries,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Berlin, NJ, USA</td>
</tr>
<tr>
<td>Synchropak RPP4000</td>
<td>4.6 x 250</td>
<td>sp Si, 10 µ 400 nm</td>
<td>C18, e</td>
<td>33</td>
<td>Synchrom Inc.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lafayette, IN, USA</td>
</tr>
<tr>
<td>Mono-RPC</td>
<td>5.0 x 50 (FPLC)</td>
<td>sp poly, 10 µ 25 nm</td>
<td>C8, na</td>
<td>34</td>
<td>Amersham Pharmacia</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Piscataway, NJ, USA</td>
</tr>
</tbody>
</table>
inner diameter x length in mm
2 type of particle (ir = irregular, sp = spherical, Si = silica, poly = polymeric); average particle diameter (μ); average pore diameter (nm)
3 pe = partially endcapped, e = endcapped, na = not applicable
4 approximate % CH₃CN in which the first H1 component is eluted
Acidic ion-pair reagents like TFA and HFBA are thought to affect RP-HPLC separations of polypeptides in multiple ways. In addition to “pairing” with the basic side chains of lysine and arginine to convert these polar moieties into less polar ones and enhance retention through hydrophobic interactions, the low pH (approx. 2.1) of mobile phases containing these reagents enhance polypeptide retention via suppression of the ionization of the carboxylic acid side chains of aspartic and glutamic acids. The low pH of these mobile phases also suppresses silanol ionization, a factor commonly thought to be important for high recoveries of basic polypeptides from silica-based supports. The contribution of these factors to the retention and resolution of H1_total was examined by substituting other reagents for TFA. H1_total was less retained when chromatographed on various RP supports in mobile phases containing 10 mM HCl (pH 2.1) or 10 mM formic acid (pH 2.7) compared to 0.1% TFA, suggesting that ion-pairing by the fluoro-substituted acids contributed to H1 retention (data not shown). Subcomponents of the H5_total and H1_total peaks were less resolved on the 100nm diam. pore Chromegabond MC-18 column under these conditions, suggesting that differences in the degree of ion-pairing with the fluoro-substituted acids may exist between components comprising these peaks (data not shown).

Unexpectedly, it was found that low pH was not required for good recovery of H1_total from any of the well-endcapped RP supports tested provided that appropriate concentrations of salts were substituted for the acids. A variety of monovalent salts including KCl, NaCl, NaF, NaSO4, were effective in the range of 100-200 mM, but 100 mM NaClO4 was the most effective and this reagent has the advantage of possessing greater solubility in organic solvents such as acetonitrile (data not shown). However, the resolution of H5_total from H1_total and the resolution of subcomponents of H1_total was less than that achieved in the conventional TFA/CH3CN mobile phase.

Substitution of other organic modifiers such as methanol, ethanol and propanol in place of CH3CN resulted in minor differences in the resolution of H1_total on several RP supports but generally was without benefit. Ultimately, it was determined that for the RP supports and conditions tested, the best resolution of H1_total was obtained with the 100 nm diam. pore
Chromegabond MC18 support using a conventional TFA/CH$_3$CN mobile phase. As shown in Fig. 3.2C, more than six peaks were evident when H$_{1\text{total}}$ was chromatographed on this support using a shallow gradient of CH$_3$CN in 0.1% TFA. However, SDS and AU-PAGE analyses revealed significant cross-contamination between the H1 variants eluting in these peaks so alternate modes of chromatography were examined to find a method capable of fractionating H$_{1\text{total}}$ with greater resolution. Nonetheless, the experiments described above identified a number of chromatographic systems suitable for separating H$_{5\text{total}}$ from H$_{1\text{total}}$ and the insight gained into factors governing the separation of chicken H1 proteins by RP-HPLC may be useful in developing methods to separate H1 variants of other species.

### 3.2.3 Resolution by SE-HPLC

At the time this work was initiated, it was commonly acknowledged among chromatographers that the supports available for size exclusion HPLC (SE-HPLC) were of limited value in polypeptide fractionations and given the similar sizes predicted for the chicken H1 variants, it seemed unlikely these proteins could be resolved by a pure size exclusion mechanism. However, given the possibility that mixed-mode effects similar to those thought to contribute to the resolution of H1 and H5 (Fig. 3.1) and mammalian H1 and H1\(^v\) (Pehrson and Cole, 1980) on Bio-Gel P100 could be exploited during chromatography to achieve a higher degree of resolution, SE-HPLC of H$_{1\text{total}}$ was explored. Only one type of support, the SW series manufactured by the Toyo Soda Company (TSK) of Japan, was tested. Columns marketed under the name Protein-Pak 300SW by Waters Associates (Table 3.2) were tested using a variety of eluents. Early tests revealed that optimal chromatography (good peak shapes, detection sensitivity) of H$_{1\text{total}}$ at pH 2.5 required the presence of moderate salt concentrations and that higher ionic strengths were required to maintain peak shapes as eluent pH was raised, suggesting that mixed-mode effects due to the presence of accessible silanols were apparent. This was also supported by the finding that the addition of ions such as triethylamine, thought to interact more strongly than H\(^+\) or glycine with silanols, were more effective in suppressing the loss of H1 due to adsorption to the support (data not shown). Ultimately, it was found that a mobile phase containing 20 mM triethylammonium phosphate (TEAP) pH 2.5, 250 mM NaClO$_4$ provided optimal recovery and resolution of H$_{1\text{total}}$. 94
Figure 3.2. Resolution of $H_{1\text{total}}$ and $H_{5\text{total}}$ by RP-HPLC on 30 nm and 100 nm diam. pore Chromegabond C18.

(A) $H_{1\text{total}}$ and $H_{5\text{total}}$ (50 ug each) were co-injected onto an analytical column (4.6 x 250 mm) packed with Chromegabond MC-18 (base = 5 u spherical silica with 30 nm diam. pores). Proteins were eluted with a 2 hour linear gradient from 20% to 60% CH$_3$CN in 0.1% TFA using a flow rate of 0.8 ml/min.

(B) All conditions as in (A) except that the chromatographic support was based on 5 u spherical silica with 100 nm diam. pores.

(C) All conditions as in (B) except that a two hour linear gradient from 28% to 36% CH$_3$CN was employed.
Table 3.2. SEC and HIC HPLC columns tested for the resolution of chicken H1 variants

<table>
<thead>
<tr>
<th>Name</th>
<th>Format</th>
<th>Particle type</th>
<th>Bonded phase</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SEC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein-Pak 300SW</td>
<td>7.5 x 300</td>
<td>sp Si, 10 μ</td>
<td>PEG</td>
<td>Waters Associates</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30 nm</td>
<td></td>
<td>Bedford, MA, USA</td>
</tr>
<tr>
<td><strong>HIC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PolyPROPYL A</td>
<td>4.6 x 200</td>
<td>sp Si, 5 μ</td>
<td>C3</td>
<td>PolyLC Inc.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100 nm</td>
<td></td>
<td>Columbia, MD, USA</td>
</tr>
<tr>
<td>Phenyl-Superose</td>
<td>5.0 x 50 (FPLC)</td>
<td>sp poly, 10 μ</td>
<td>C6</td>
<td>Amersham Pharmacia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25 nm (aromatic ring)</td>
<td></td>
<td>Piscataway, NJ, USA</td>
</tr>
</tbody>
</table>

1 inner diameter x length in mm  
2 type of particle (sp = spherical, Si = silica, poly = polymeric); average particle diameter (μ); average pore diameter (nm)  
3 PEG = polyethylene glycol
with this support. One column (7.5 mm i.d. x 300 mm) alone was sufficient to separate H1\textsubscript{total} from H5\textsubscript{total} but provided no resolution of H1 subtypes. When 3 columns were used in series, two peaks were partially resolved for H1\textsubscript{total} while H5\textsubscript{total} eluted as a single peak (Fig. 3.3). In summary, this method has some utility for separating H1 from H5 but is inferior to RP-HPLC in terms of speed, resolution and loading capacity.

### 3.2.4 Resolution by HI-HPLC

Hydrophobic interaction chromatography (HIC) is a technique in which hydrophobic interactions between proteins and alkyl moieties on the surface of an otherwise hydrophilic stationary phase are promoted in solutions of high ionic strength. Because these fractionations are performed using mobile phases containing high concentrations (e.g. 1 - 2 M) of salts (typically cosmotropic salts like (NH\textsubscript{4})\textsubscript{2}SO\textsubscript{4}), they are frequently found to be less denaturing than other modes of chromatography and it seems likely in many instances that residues interacting with the column under these conditions are ones that are localized to the surface of proteins in their normal cellular context. Mixtures of proteins are typically fractionated using an inverse concentration gradient of the salt initially employed to promote protein binding although other strategies such as increasing concentration gradients of organic solvents or chaotropic salts have also been employed (Scopes, 1994).

In the section on CX-HILIC below (section 3.2.6), an optimized method is described which separates more than 14 peaks for H1\textsubscript{total}. The resolution of H1\textsubscript{total} by HI-HPLC described here was investigated primarily as an adjunct method to see if this technique revealed a similar degree of heterogeneity in H1\textsubscript{total}. Consequently, only two HIC supports and a limited number of mobile phase conditions were tested. Phenyl-Superose was selected as a prototype for supports containing a cyclic alkyl chain bonded phase while PolyPROPYL A was selected as a prototype for supports containing a linear alkyl chain bonded phase (Table 3.2). Initial trials using decreasing gradients of (NH\textsubscript{4})\textsubscript{2}SO\textsubscript{4} in 50 mM phosphate buffer at pH 7.0 revealed that Phenyl-Superose resolved H1\textsubscript{total} from H5\textsubscript{total} (H1 elutes first) but was not effective for resolving H1 subtypes (data not shown). Similar results were obtained using other salts (Na\textsubscript{2}SO\textsubscript{4}, KH\textsubscript{2}PO\textsubscript{4}, NaCl) in place of
Higher initial concentrations of the various salts tested were required to achieve sufficient binding of H1\textsubscript{total} to PolyPROPYL A compared to Phenyl-Superose. Although the strength of binding of H1\textsubscript{total} (and H5\textsubscript{total}) to Phenyl-Superose in 2.0 M KH\textsubscript{2}PO\textsubscript{4} or 5.0 M NaCl (near saturation for each salt at 22°C) was adequate for chromatography at pH 7.0, these proteins were too weakly retained on PolyPROPYL A for efficient chromatography using these salts. However, the binding of H1\textsubscript{total} to PolyPROPYL A in 1.5 M Na\textsubscript{2}SO\textsubscript{4} or 2.2 M (NH\textsubscript{4})\textsubscript{2}SO\textsubscript{4} was sufficient for high resolution fractionations to be performed.

Figure 3.4 shows a separation of H1\textsubscript{total} on PolyPROPYL A at pH 7.0 using an inverse gradient of (NH\textsubscript{4})\textsubscript{2}SO\textsubscript{4}. Approximately 11 components were resolved to varying degrees. Because greater resolution of H1\textsubscript{total} had already been attained using CX-HILIC (section 2.3.2f), this method was not optimized further and the fractions resolved by this technique are not characterized here. Nonetheless, this work demonstrated that HI-HPLC is capable of resolving heterogeneity in H1\textsubscript{total} to an extent similar to that of CX-HILIC. Despite disadvantages associated with the high ionic strengths used in this method (difficulties in protein recovery, high backpressures during chromatography), the HIC technique is likely to be useful for resolving heterogeneity of H1, and possibly other histones, in a variety of species. The technique may be particularly appropriate for studies attempting to avoid denaturation of H1 and appears to be well suited for the preparation of H1 fractions from the salt extracts of nuclei frequently used for the isolation of transcriptional regulators (e.g. Dignam et al., 1983; Croston et al., 1991).

### 3.2.5 Resolution by CX-HPLC

Table 3.3 lists key properties of the various CX-HPLC columns tested. Initial comparisons of the resolution of H1\textsubscript{total} on this array of columns using linear gradients of NaCl in 20 mM phosphate buffer pH 6.5 (see Methods, section 2.5), revealed that the highest resolution under these conditions was obtained with PolyCAT A supports having certain pore diameters. Mixed-mode effects were apparent for H1\textsubscript{total} on all the supports tested and in the case of Poly CAT
Figure 3.3. Resolution of H₁total and H₅total by SE-HPLC.

H₁total and H₅total (approx. 100 μg of each) were co-injected and separated on three Waters Protein Pak 300SW columns (each 7.5 x 300 mm) eluted in series at 0.1 ml/min with 250 mM NaClO₄ in 20 mM phosphoric acid-triethylamine pH 2.5.
Figure 3.4. Resolution of H1\textsubscript{total} by HI-HPLC on PolyPROPYL A.

H1\textsubscript{total} (approx. 50 ug) was injected onto an analytical column (4.6 x 200 mm) packed with PolyPROPYL A (base = 5 u spherical silica with 100 nm diam. pores) equilibrated with 2.2 M (NH\textsubscript{4})\textsubscript{2}SO\textsubscript{4} in 50 mM ethylene diamine pH 7.0. H1 variants were separated with a two hour linear gradient from 2.2 M to 1.54 M (NH\textsubscript{4})\textsubscript{2}SO\textsubscript{4} in 50 mM ethylene diamine pH 7.0 using a flow rate of 0.8 ml/min.
A supports, these were exploited to develop an optimized method described in section 3.2.6 below. Here, the best results obtained with other supports are summarized to facilitate comparison to the results obtained with PolyCAT A.

Initial testing of the Mono S strong cation-exchange support revealed that the profile reported previously (Shannon et al., 1987; see Fig. 1.10C) could be reproduced for the H1_{total} sample using a similar gradient of GuHCl in 20 mM sodium phosphate buffer pH 6.5 (data not shown). Moreover, similar profiles were obtained using NaCl or sodium phosphate as gradient materials in place of GuHCl at this pH with the advantage detection sensitivity was enhanced due to lower baseline absorbancy (note that the chromatogram shown in Fig. 1.10C represents 1 mg of sample equivalent to H1_{total}). PAGE analyses revealed that each peak contained significant amounts of the H1 variants eluting in adjacent peaks regardless of the gradient material used (data not shown). Extensive variation of mobile phase composition within the useful pH range and chemical stability of this polymeric material only decreased the resolution of H1_{total}. In mobile phases adjusted to pH 2.2 - 4.0, H1_{total} was strongly retained, eluting as a single broad peak regardless of whether NaCl, NaClO_4 or GuHCl was employed as the gradient material, and inclusion of up to 40% CH_3CN in the mobile phase did not alter the profile (data not shown). H1_{total} was poorly retained, eluting as a single broad peak when phosphate buffers adjusted to pH 10-11.5 were used with a NaCl gradient, suggesting that the initial binding of the protein to the support under these conditions was too weak to form a narrow starting zone.

Given some similarities in composition, it was anticipated that another polymeric strong cation exchanger, Waters TSK SP-5PW, would give results similar to those obtained with the Mono S support. However, significant differences were observed. Extensive experimentation with mobile phase conditions determined that the best resolution of H1_{total} on this support was obtained at low pH. H1_{total} eluted as a single peak using linear gradients of NaCl in 20 mM sodium phosphate pH 6.5 (data not shown), but approximately seven components were partially resolved using the same mobile phase adjusted to pH 2.5 (Fig. 3.5A). PAGE analyses revealed extensive cross-contamination of H1 variants under these conditions (data not shown). Inclusion of up to
Figure 3.5. Resolution of H1<sub>total</sub> by CX-HPLC on TSK SP-5PW and CM-3SW.

(A) H1<sub>total</sub> (approx. 50 ug) was injected onto an analytical (7.5 x 75 mm) Waters TSK SP-5PW column equilibrated in 0.55 M NaCl in 10 mM H<sub>3</sub>PO<sub>4</sub>-NaOH pH 2.5. H1 variants were separated with a two hour linear gradient from 0.55 M to 1.0 M NaCl in 10 mM H<sub>3</sub>PO<sub>4</sub>-NaOH pH 2.5 using a flow rate of 1.0 ml/min.

(B) H1<sub>total</sub> and H5<sub>total</sub> (approx. 100 and 50 ug, respectively) were co-injected onto an analytical (7.5 x 75 mm) Beckman Spherogel TSK CM-3SW column equilibrated with 0.1 M NaCl in 20% CH<sub>3</sub>CN, 10 mM H<sub>3</sub>PO<sub>4</sub>-NaOH pH 6.5. Proteins were separated using a one hour linear gradient from 0.1 M to 1.0 M NaCl in 20% CH<sub>3</sub>CN, 10 mM H<sub>3</sub>PO<sub>4</sub>-NaOH pH 6.5 using a flow rate of 1.0 ml/min.
40% CH₃CN in mobile phases and/or the use of gradient materials other than NaCl did not significantly alter the elution profile at pH 2.5. Although cross-contamination was too great for purification of H1 variants under these conditions, the efficacy of this support with low pH mobile phases suggests that the material may be useful for recovering H1, and possibly core histones, directly from acid extracts of nuclei or cells.

After PolyCAT A (and the related PolySULFOETHYL A), the Beckman Spherogel TSK CM-3SW was found to be the best-performing silica-based support tested. Optimal resolution of H1_total on this column is shown in Fig. 3.5B. Evidence of hydrophobic interaction of H1 with this support was suggested by the finding that inclusion of 20% CH₃CN gave sharper peaks and decreased retention slightly compared to mobile phases containing 0% CH₃CN. Extensive variation of mobile phase conditions suggested that NaCl was the best choice of gradient material with this support. Three peaks were partially resolved for H1_total using a relatively steep gradient of NaCl in 20 mM sodium phosphate pH 6.5 containing 20% CH₃CN. PAGE analyses revealed that each of these peaks contained multiple H1 variants (data not shown). Attempts to increase the resolution by using shallower gradients of NaCl were unsuccessful due to band spreading.

3.2.6 Resolution by CX-HILIC

During the initial comparisons of the resolution of H1_total on CX-HPLC columns using linear gradients of NaCl in 20 mM sodium phosphate pH 6.5, the greatest resolution was obtained with the weak cation exchanger PolyCAT A based on 5 μm diam. silica with 30 nm diam. pores. Fig. 3.6A shows a typical result for this support that employing NaClO₄ as the gradient material. Similar resolution was also obtained on the related strong cation exchanger PolySULFOETHYL A (data not shown). We found that eluent composition significantly affected the retention and resolution of H1_total on PolyCAT A. As shown in Figs. 3.6A-C, when the same NaClO₄ gradient was employed with buffers containing 0%, 40% or 70% acetonitrile, the retention of H1_total was lower in buffers containing 40% acetonitrile. The inset of Fig. 3.6B shows a separation in 40% acetonitrile using a NaClO₄ gradient with the same slope as those in Figs. 3.6A-C, but after
Table 3.3. CX-HPLC columns tested for the resolution of chicken H1 variants

<table>
<thead>
<tr>
<th>Name</th>
<th>Format</th>
<th>Particle type</th>
<th>Ion-exchanger</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mono-S</td>
<td>5.0 x 50 (FPLC)</td>
<td>sp poly, 10 μ</td>
<td>SO₃⁻</td>
<td>Amersham Pharmacia Piscataway, NJ, USA</td>
</tr>
<tr>
<td>Protein-Pak SP-5PW</td>
<td>7.0 x 50</td>
<td>sp poly, 10 μ</td>
<td>SO₃⁻</td>
<td>Waters Associates Bedford, MA, USA</td>
</tr>
<tr>
<td>Tessek Hema-Bio 1000CM</td>
<td>5.0 x 50</td>
<td>sp poly, 10 μ</td>
<td>SO₃⁻</td>
<td>Fisher Scientific Fairlawn, NJ, USA</td>
</tr>
<tr>
<td>Synchropak CM-300</td>
<td>4.6 x 250</td>
<td>sp Si, 5 μ</td>
<td>SO₃⁻</td>
<td>Synchrom Inc. Lafayette, IN, USA</td>
</tr>
<tr>
<td>Spherogel TSK CM3SW</td>
<td>4.6 x 250</td>
<td>sp Si, 10 μ</td>
<td>SO₃⁻</td>
<td>Beckman Instruments Fullerton, CA, USA</td>
</tr>
<tr>
<td>Spherogel TSK IEX-535CM</td>
<td>4.6 x 250</td>
<td>sp Si, 10 μ</td>
<td>SO₃⁻</td>
<td>Beckman Instruments Fullerton, CA, USA</td>
</tr>
<tr>
<td>Aquapore CX-300</td>
<td>4.6 x 250</td>
<td>sp Si, 7 μ</td>
<td>COO⁻</td>
<td>P-E/Applied Biosystems Foster City, CA, USA</td>
</tr>
<tr>
<td>PolySULFOETHYL A</td>
<td>4.6 x 200</td>
<td>sp Si, 5 μ</td>
<td>COO⁻</td>
<td>PolyLC Inc. Columbia, MD, USA</td>
</tr>
<tr>
<td>PolyCAT A</td>
<td>4.6 x 200</td>
<td>sp Si, 5 μ</td>
<td>COO⁻</td>
<td>PolyLC Inc. Columbia, MD, USA</td>
</tr>
</tbody>
</table>
inner diameter x length in mm

composition (sp = spherical, Si = silica, poly = polymeric); average particle diameter (μ); average pore diameter (nm)

SO_3^- = sulfonic acid based, COO^- = carboxylic acid based

columns packed with material corresponding to each pore diameter were tested
reduction of the initial and final salt concentrations to make the overall retention of H1_total similar to that in Figs. 3.6A-C. Despite this adjustment, the resolution of H1 variants in 40% acetonitrile was similar or slightly inferior to that achieved in completely aqueous buffer. In contrast, the resolution of H1 variants in buffer containing 70% acetonitrile (Fig. 3.6C) was markedly greater than that achieved in buffers containing less acetonitrile. In the absence of acetonitrile, various gradient materials, including potassium chloride, sodium chloride and sodium sulfate gave results similar to those shown for sodium perchlorate in Fig. 3.6A (data not shown). However, sodium perchlorate was the only salt available in sufficiently pure form found to be compatible with chromatography in buffers composed primarily of acetonitrile. Detailed optimization experiments established that the retention and resolution of H1 variants were decreased when buffers containing 70% acetonitrile were adjusted to pH values other than 6.5 (data not shown). Additional data relevant to the mechanisms governing this separation are presented below, but the effects of CH3CN on the resolution and retention of H1_total clearly indicate that more than just electrostatic interactions are involved. Because recent examples of separations performed on hydrophilic supports with largely non-aqueous mobile phases have been attributed to hydrophilic interactions (see Chapter 7), this mixed-mode form of cation-exchange chromatography will be referred to by the acronym CX-HILIC.

3.2.6a Effect of CH3CN on H1 retention on PolyCAT A

To better understand the effect of acetonitrile on the retention of H1_total by PolyCAT A, a series of separations was performed using identical NaClO4 gradients in which buffer acetonitrile content was increased by increments of 5 or 10%. The retention time of the earliest eluting H1 peak was plotted as a function of buffer acetonitrile content (Fig. 3.7). H1 retention decreased gradually as buffer acetonitrile content was increased from 0% to 40%, reached a minimum at approximately 40% acetonitrile, increased gradually as acetonitrile content was raised from 40 to 60%, and then increased steeply between 60% and 70% acetonitrile. Even though H1_total remained soluble at a concentration of 1 mg/ml in buffers containing 90% acetonitrile and 0.4 M NaClO4, elution of H1 was not observed when acetonitrile concentrations greater than 72% were employed
Figure 3.6. Resolution of $H_{\text{total}}$ on PolyCAT A in 0, 40 and 70% CH$_3$CN.

Identical 100 $\mu$g samples of pooled chicken erythrocyte H1 were chromatographed by CX-HPLC (A and B) and CX-HILIC (C). (A-C) An analytical PolyCAT A column (4.6 x 200 mm; 5 μm spherical silica, 100 nm diam. pores) was eluted with a 4 hour linear gradient from 380 mM to 590 mM NaClO$_4$. Buffers were 10 mM sodium phosphate pH 6.5 (A), 10 mM sodium phosphate pH 6.5 containing 40% (v/v) CH$_3$CN (B), and 10 mM propionic acid pH 6.5 containing 70% (v/v) CH$_3$CN (C). The profile obtained with a 4 hour linear gradient from 260 mM to 440 mM NaClO$_4$ in 10 mM sodium phosphate pH 6.5 containing 40% (v/v) CH$_3$CN is shown in the inset of (B). The flow rate was 0.8 ml/min in all panels.
for chromatography, apparently due to precipitation of H1 within the HPLC system following injection (data not shown).

3.2.6b H1 conformation in buffers containing CH₃CN and NaClO₄

To ascertain whether alterations in H1 conformation contributed to the resolution of variants achieved in buffers containing 70% acetonitrile, we studied the effects of acetonitrile and sodium perchlorate on the conformation of H1_total using circular dichroism (CD) spectroscopy (Fig. 3.8). NaClO₄ (0.6 M) and 70% CH₃CN each increased the α-helix content of H1_total to a similar extent (enhanced negative ellipticity at 222 nm) relative to its conformation in 10 mM sodium phosphate (pH 6.5) alone. When acetonitrile and sodium perchlorate were used in combination, i.e. as in CX-HILIC, the effects were additive. Thomas and colleagues (Clark et al., 1988; Hill et al., 1989) have shown previously that 1 M NaClO₄ and 50-65% (v/v) 2.2.2-trifluoroethanol (TFE) stabilize helices in chicken H1.

3.2.6c Effect of PolyCAT A support pore diameter

The pore size of the PolyCAT A packing had a significant effect on H1 retention and variant resolution. Using gradients that were 2 hours long and had identical slopes, the resolution of H1_total on PolyCAT A packings prepared using silicas with average pore diameters of 20, 30, 100 and 400 nm (according to manufacturers specifications) under CX-HPLC (Fig. 3.9A-D, 0% acetonitrile) and CX-HILIC conditions (Fig. 3.9E-H, 70% acetonitrile) was compared. The overall retention of H1_total decreased with increasing support pore diameter under both CX-HPLC and CX-HILIC conditions. In aqueous buffer, H1 variants were best resolved on the 30 nm diameter pore support (Fig. 3.9B), but this was inferior to the resolution achieved with any of the four different pore diameter supports using CX-HILIC conditions (Fig. 3.9E-H). Under the conditions employed in this comparison, the greatest number of variants was resolved by CX-HILIC using the 100 nm diameter pore packing (Fig. 3.9G). However, careful comparison of the numbers, shapes and spacing of the peaks resolved on the various supports suggested the possibility that different sets of variants were resolved on the 30 nm and 100 nm diameter pore
Figure 3.7. Effect of CH$_3$CN on H1 retention by PolyCAT A.

Identical 100 µg samples of H$_1$$_{total}$ were chromatographed on the same analytical PolyCAT A column (4.6 x 200 mm; 100 nm diam. pores) with a 4 hour gradient from 380 mM to 590 mM NaClO$_4$ in mobile phases containing 0, 10, 20, 30, 40, 45, 50, 55, 60, 65 and 70% (v/v) CH$_3$CN. Buffers were 10 mM sodium phosphate pH 6.5 for mobile phases containing 0-50% (v/v) CH$_3$CN and 10 mM propionic acid pH 6.5 for mobile phases containing 55-70% (v/v) CH$_3$CN. All other conditions were as in Fig. 3.6. The retention time of the first H$_1$ component to elute is plotted as a function of the mobile phase CH$_3$CN concentration.
Figure 3.8. Conformation of H1 in buffers containing CH3CN and NaClO4.

CD spectra are shown of identical samples of H1_total dissolved in 10 mM sodium phosphate pH 6.5, 10 mM sodium phosphate pH 6.5 containing 0.6 M NaClO4, 10 mM sodium phosphate pH 6.5 containing 70% (v/v) CH3CN, and 10 mM sodium phosphate pH 6.5 containing 0.6 M NaClO4 and 70% (v/v) CH3CN.
Figure 3.9.  Effect of PolyCAT A support pore diameter on resolution of H1_{total}.

Identical 100 ug samples of pooled chicken erythrocyte H1 were chromatographed on analytical columns (4.6 x 200 mm) packed with PolyCAT A supports with 20 (A and E), 30 (B and F), 100 (C and G) and 400 (D and H) nm diameter pores. Two hour linear gradients were employed from 400 mM to 600 mM NaClO4 in 10 mM sodium phosphate pH 6.5 (A-D) or from 500 mM to 700 mM NaClO4 in 10 mM propionic acid pH 6.5 containing 70% (v/v) CH3CN (E-H). All other conditions were as in Fig. 3.6.
supports (Fig. 3.9F and G, see below).

3.2.6d Optimized resolution of H1\textsubscript{total} by CX-HILIC

Optimized CX-HILIC separations of pooled blood H1\textsubscript{total} on PolyCAT A packings with 30 or 100 nm diameter pores are shown in Fig. 3.10A and B, respectively. Doubling the gradient time relative to that used in Fig. 3.9 (i.e., increase the concentration of NaClO\textsubscript{4} by 200 mM linearly over 4 hours) provided the greatest resolution of H1 variants on both supports. Under these conditions, the resolution of variants appeared to be similar on both columns: 11 and 12 major peaks were resolved on the 30 and 100 nm diameter pore supports, respectively. However, electrophoretic analyses, and analyses of peaks resolved on one support by rechromatography on the other, revealed differences in the resolution of particular H1 variants on the two supports. These differences are annotated in the peak numbering used in Fig. 3.10. H1 variants 1 and 2, variants 7 and 8, and variants 12 and 13, comprising the first, sixth and tenth major peaks, respectively, to elute from the 30 nm diameter pore support, were resolved on the 100 nm diameter pore support. Conversely, variants 4 and 5, and variants 8 and 9, which comprised the fourth and seventh major peaks, respectively, to elute from the 100 nm diameter pore support were resolved on the 30 nm diameter pore support. Additionally, several minor H1 variants (indicated by the horizontal bars) appeared to be better resolved on the 30 nm pore support.
Figure 3.10. Optimized CX-HILIC resolution of H1 variants on 30 and 100 nm diameter pore PolyCAT A.

Identical 100 μg samples of pooled chicken erythrocyte H1 were chromatographed on analytical columns (4.6 x 200 mm) packed with 5 μspherical PolyCAT A supports with 30 (A) and 100 (B) nm diameter pores. H1 variants were separated using four hour linear gradients from 460 to 660 mM NaClO₄ (A), and 420 to 620 mM NaClO₄ (B), in 10 mM propionic acid pH 6.5 containing 70% (v/v) CH₃CN at a flow rate of 0.8 ml/min. Peaks are numbered according to order of elution, the same numbering is used for the PAGE analyses shown in Fig. 4.1. The differences in resolution between the two supports indicated by the peak numbering were confirmed by rechromatography and PAGE analyses. Horizontal bars indicate minor components confirmed to be H1 variants by PAGE that are not characterized here.
Chapter 4. Characterization of CX-HILIC fractions of $H_{1\text{total}}$

from pooled erythrocytes
4.1 Introduction

Because six bands were resolved for $H_1_{\text{total}}$ on long AU gels (see Fig. 4.1B) and six genes were estimated to comprise the entire set of $H_1$ genes in chickens (Coles et al., 1987), the resolution of more than 14 components for $H_1_{\text{total}}$ by CX-HILIC was unexpected. In this chapter, I describe characterization of the major CX-HILIC peaks of $H_1_{\text{total}}$ by PAGE and electrospray mass spectrometry. Long AU gels were used to identify the proteins separated by CX-HILIC according to the nomenclature described previously for chicken $H_1$ variants (Shannon et al., 1987) which permits correlation with the reported gene sequences (Coles et al., 1987). All of the CX-HILIC peaks contained a predominant form corresponding to one of the AU gel variants. However, in almost all cases, mass spectrometry revealed that the proteins were significantly smaller than predicted by the corresponding gene sequences and it is shown here that these discrepancies are not the result of simple misidentification of variants or artifacts of the methods of protein preparation and analysis.

Most CX-HILIC peaks appeared to be homogenous in PAGE after secondary purification to eliminate cross-contamination between proteins that eluted adjacent to one another in the primary chromatography. However mass spectrometry revealed heterogeneity in all of the CX-HILIC peaks. In some cases, this was potentially attributable to residual cross-contamination between peaks but other cases appeared to be more likely due to co-elution of distinct forms of the same $H_1$ variant within an CX-HILIC peak that possessed identical mobilities in PAGE analyses.

Although somewhat novel at the time this work was initiated, protein characterization by mass spectrometry has become commonplace. The principles behind this technique and the high precision of the molecular masses determined routinely for proteins using electrospray instruments have been reviewed elsewhere (e.g. Smith et al., 1990; Smith et al., 1991; Chait and Kent, 1992).

4.2 Results and Discussion

4.2.1 PAGE identification of $H_1_{\text{total}}$ CX-HILIC peaks

Conventional SDS-PAGE resolves only three bands for chicken erythrocyte $H_1$ (e.g. Fig
4.1B). However, electrophoresis in long AU gels resolves 6 bands, a, a', b, c, c' and d, that have been shown to correspond to the six H1 genes characterized in chicken erythrocytes (Shannon and Wells, 1987). To prepare sufficient amounts of the H1 variants resolved in the major CX-HILIC peaks to permit both PAGE identification and other analyses, 50 mg of H1 \text{total} was separated in three runs (10, 20 and 20 mg) on a preparative column (21.1 x 250 mm) packed with 100 nm diam. pore PolyCAT A as shown in Fig. 4.1A. H1 proteins in pooled fractions representing each peak were recovered by TCA precipitation and rechromatographed on semi-preparative columns (9.4 x 200 mm) packed with 30 nm or 100 nm diam. pore PolyCAT A as required (data not shown). Rechromatography of the peaks labeled 4+5 and 8+9 in Fig. 4.1A was performed on the 30 nm diam. pore semi-prep PolyCAT A column in order to separate the forms known to co-elute from 100 nm pore PolyCAT A in these two peaks (Fig. 3.10). SDS and AU-PAGE analyses of pooled fractions from the primary chromatography revealed moderate cross-contamination between adjacent peaks (data not shown). After rechromatography, most of the CX-HILIC peaks contained a single H1 band in SDS and AU-PAGE that corresponded to one of the six bands comprising total H1 (lanes S) in AU-PAGE (Fig. 4.1B). However, in addition to a predominant band, peaks 6, 8, 11, 12 and 13 were each found to contain a less abundant band that appeared to be attributable to residual cross-contamination. The less abundant component (H1a) in peaks 6, 8, and 13 was resolved from the predominant forms (H1c, H1c and H1c', respectively) in these peaks on both SDS and AU gels. Cross-contamination of H1a' in peak 11 by H1a (originating from peaks 10 and/or 11) and cross-contamination of H1a in peak 12 by H1a' from peak 11 was apparent only on AU gels due to comigration of the H1a and H1a' proteins in SDS gels (Fig. 4.1B). The apparent identity of the predominant band in each PolyCAT A peak according to AU-PAGE mobility, together with characteristics of each H1 variant predicted from the corresponding gene sequence, is listed in Table 4.1. Remarkably, multiple forms of H1a, H1b and H1c were resolved by CX-HILIC. Peaks 1 through 4 all migrated equivalently to H1b, peaks 7, 9, 10 and 12 migrated equivalently to H1a and peaks 5, 6 and 8 migrated equivalently to H1c (Fig. 4.1B and Table 4.1). Of the three forms of H1c (lanes 5, 6 and 8 in Fig. 4.1B), two coeluted with forms of H1b and H1a (i.e., variants 4 and 9, respectively), in the peaks labeled 4+5 and 8+9 from 100 nm diameter pore PolyCAT A (Fig. 4.1A). AU-PAGE analyses of the late-eluting minor components (bar in
Figure 4.1 Preparative scale purification of H1<sub>total</sub> CX-HILIC peaks.

(A) Pooled blood H1<sub>total</sub> (approx. 10 mg) was separated on a preparative column (21.1 x 250 mm) of Poly CAT A (100 nm diam. pores) using a four hour linear gradient from 380 mM to 590 mM NaClO₄ in 70% CH₃CN in 10 mM propionic acid-NaOH pH 6.5 at a flow rate of 12.5 ml/min. As described in the text, fractions from other runs were pooled and rechromatographed on semi-preparative columns (9.4 x 200 mm) packed with 30 nm or 100 nm diam. pore PolyCAT A as appropriate to remove cross-contamination.

(B) After rechromatography, aliquots representing approximately 500 ng of peaks recovered from (A) were analysed in a 12% SDS gel. Proteins detected by silver staining. Lane S = H1<sub>total</sub> (starting material). Lane numbering corresponds to the peak numbering used in Fig. 3.10. The relative positions of the six H1 variants based on AU gel mobility (determined in (C)) are shown on the right side.

(C) Aliquots of the recovered fractions (approx. 500 ng) were analysed on a 50 cm long AU gel. Lane numbering as in (B). Proteins detected by silver staining. The six H1 variants based on AU gel mobility in the nomenclature of Shannon and Wells (1987) are shown on the right side.
Table 4.1. Predicted properties of chicken H1 variants

<table>
<thead>
<tr>
<th>CX-HILIC peak</th>
<th>AU-PAGE variant</th>
<th>predicted net charge (pH 6.5)</th>
<th>predicted length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>b</td>
<td>+59</td>
<td>223</td>
</tr>
<tr>
<td>2</td>
<td>b</td>
<td>+59</td>
<td>223</td>
</tr>
<tr>
<td>3</td>
<td>b</td>
<td>+59</td>
<td>223</td>
</tr>
<tr>
<td>4</td>
<td>b</td>
<td>+59</td>
<td>223</td>
</tr>
<tr>
<td>5</td>
<td>c</td>
<td>+59</td>
<td>219</td>
</tr>
<tr>
<td>6</td>
<td>c</td>
<td>+59</td>
<td>219</td>
</tr>
<tr>
<td>7</td>
<td>a</td>
<td>+59</td>
<td>224</td>
</tr>
<tr>
<td>8</td>
<td>c</td>
<td>+59</td>
<td>219</td>
</tr>
<tr>
<td>9</td>
<td>a</td>
<td>+59</td>
<td>224</td>
</tr>
<tr>
<td>10</td>
<td>a</td>
<td>+59</td>
<td>224</td>
</tr>
<tr>
<td>11</td>
<td>a'</td>
<td>+58</td>
<td>218</td>
</tr>
<tr>
<td>12</td>
<td>a</td>
<td>+59</td>
<td>224</td>
</tr>
<tr>
<td>13</td>
<td>c'</td>
<td>+59</td>
<td>218</td>
</tr>
<tr>
<td>14</td>
<td>d</td>
<td>+60</td>
<td>217</td>
</tr>
</tbody>
</table>

1 peaks are numbered according to elution order in Fig. 4.1.

2 nomenclature based on mobility in AU-PAGE as described previously (Shannon and Wells, 1987; also see Fig. 4.1)

3 derived from the gene sequences (Coles et al., 1987; see Fig. 4.2) assuming free amino and carboxyl termini and assigning +1 for all lysine and arginine residues and -1 for all aspartic acid and glutamic acid residues.
**Figure 4.2.** Amino acid sequences of chicken H1 variants predicted from the corresponding genes.

The protein sequences predicted from the genes described previously (Coles et al., 1987) are shown aligned for maximum homology. The nomenclature used to identify the genes is shown in brackets. The corresponding H1 variants based on AU gel mobility identified by Shannon and Wells (1987) are also shown. Stars indicate residues identical to the H1a sequence, gaps introduced for alignment are indicated by hyphens. The predicted protein lengths are indicated at the end of each sequence. The approximate boundaries of the conserved globular domain are indicated by the line above the sequences. The solid and empty arrowheads indicate the unique tyrosine and phenylalanine residues, respectively that are present in all six variants.
Fig. 4.1A) suggested that additional forms of H1a' and H1d were also present in H1_{total} (data not shown).

4.2.2 Electrospray MS characterization of H1_{total} CX-HILIC peaks

Aliquots of each of the recovered H1_{total} CX-HILIC peaks were analysed by electrospray mass spectrometry (ESMS). It was expected that the relative molecular mass (M_r) determined by this technique could be used to confirm the H1 subtype identifications assigned on the basis of AU gel mobilities (Fig. 4.1 and Table 4.1). Moreover, any post-translational modifications contributing to the resolution of the multiple forms of variants H1a, H1b and H1c by CX-HILIC would be revealed by M_r values that deviated from those predicted by the respective gene sequences.

Mass spectra were obtained on a single quadrupole mass spectrometer fitted with an electrospray interface at the University of Michigan Protein and Carbohydrate Structure Facility in collaboration with Drs. Phil C. Andrews (Director) and Rachel Loo. The results of these analyses are summarized in Table 4.2 where the M_r values of all the polypeptides detected in H1_{total} CX-HILIC peaks 1-14 are aligned with the values for protein length and M_r predicted by the gene sequence of the corresponding H1 variant based on AU gel mobility (Shannon et al., 1987; Coles et al., 1987). Predicted M_r values corresponding to both the α-N-acetylated and non-α-N-acetylated forms of the H1 variants are listed in Table 4.2 with the non-α-N-acetylated forms shown in brackets. Because evidence from protein sequence analyses indicates that H1 variants in chickens and other metazoans are usually α-N-acetylated (Von Holt et al., 1989; van Holde, 1989), the M_r predicted for the α-N-acetylated variants are used for all comparisons listed in Table 4.2. The M_r values determined for each CX-HILIC peak are listed in order of their apparent abundance in the mass spectra. The M_r of relatively minor components, estimated to represent less than 10% of the sample, are shown in square brackets. Two striking findings were revealed in these analyses: (1) all of the CX-HILIC peaks, including those that appeared to be homogenous on
both SDS and AU gels, were heterogeneous, and (2) except for peak 11 which contained proteins whose \( M_r \) was larger than predicted for H1a', the \( M_r \) of the major forms detected by ESMS in all other CX-HILIC peaks were significantly smaller than the values predicted for the corresponding H1 variants.

The heterogeneity associated with the H1\text{total} CX-HILIC peaks and the discrepancies between the measured and predicted \( M_r \) values were unexpected and experiments were performed to ascertain whether these observations were attributable to artifacts of the methods of protein preparation or ESMS analysis. As described below, mass spectrometry of a subset of the peaks listed in Table 4.2 at a second laboratory gave equivalent results. As described in Chapter 5, the \( M_r \) observed for CX-HILIC peaks of H5\text{total} processed in parallel were identical to the predicted value. Together, these data strongly suggest that the discrepancies in the measured and predicted \( M_r \) for the H1\text{total} CX-HILIC peaks were not attributable to artifacts of the methods used. Before describing further experimental work, several features of the data in Table 4.2 merit consideration here. Careful examination of these data reveals that the discrepancies between the predicted and measured \( M_r \) values are not readily explained by simple mismatching of CX-HILIC H1 peaks with inappropriate gene sequences since in most cases the \( M_r \) values observed for a given CX-HILIC peak are not better matched with values predicted for H1 variants other than the one selected based on AU gel mobility. For example, all the \( M_r \) values observed for CX-HILIC peaks 1-4, which contained proteins with mobility indistinguishable from that of H1b on AU gels, are smaller than the \( M_r \) value predicted for any H1 subtype regardless of whether \( \alpha \)-N-acetylation is assumed or not (Table 4.2).

The magnitude of the differences between the observed and predicted \( M_r \) values vary considerably, ranging from +685 to -1745 when the predicted value is subtracted from the measured \( M_r \) (Table 4.2). In cases where the observed \( M_r \) values were less than expected, these differences could possibly be due to proteolysis of the H1 proteins \textit{in vivo} and/or during purification. However, the finding that the \( M_r \) observed for H5 proteins processed in parallel
Table 4.2. Electrospray MS of pooled chicken erythrocyte H1 CX-HILIC fractions.

<table>
<thead>
<tr>
<th>HPLC peak#</th>
<th>AU gel subtype</th>
<th>predicted length</th>
<th>predicted M&lt;sub&gt;r&lt;/sub&gt;</th>
<th>measured M&lt;sub&gt;r&lt;/sub&gt;</th>
<th>difference&lt;sup&gt;5&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>b</td>
<td>223</td>
<td>22437.3 (22395.3)</td>
<td>21318 ± 6 [21004 ± 8]</td>
<td>-1119 [-1433 [20693 ± 8] [-1744 [21740 ± 10] [-697]</td>
</tr>
<tr>
<td>2</td>
<td>b</td>
<td>223</td>
<td>22437.3 (22395.3)</td>
<td>21300 ± 13 [20978 ± 7]</td>
<td>-1137 [-1459]</td>
</tr>
<tr>
<td>3</td>
<td>b</td>
<td>223</td>
<td>22437.3 (22395.3)</td>
<td>21300 ± 15 [20975 ± 7]</td>
<td>-1137 [-1462]</td>
</tr>
<tr>
<td>4</td>
<td>b</td>
<td>223</td>
<td>22437.3 (22395.3)</td>
<td>20956 ± 12 [21263 ± 8]</td>
<td>-1481 [-1174]</td>
</tr>
<tr>
<td>5</td>
<td>c</td>
<td>219</td>
<td>21843.4 (21801.4)</td>
<td>20991 ± 6 [21686 ± 6]</td>
<td>-852 [-157]</td>
</tr>
<tr>
<td>6</td>
<td>c</td>
<td>219</td>
<td>21843.4 (21801.4)</td>
<td>20979 ± 9 [21670 ± 5]</td>
<td>-864 [-173]</td>
</tr>
<tr>
<td>7</td>
<td>a</td>
<td>224</td>
<td>22439.2 (22397.2)</td>
<td>22017 ± 9 [21666 ± 11]</td>
<td>-422 [-773]</td>
</tr>
<tr>
<td>9</td>
<td>a</td>
<td>224</td>
<td>22439.2 (22397.2)</td>
<td>22064 ± 8 [21977 ± 4]</td>
<td>-375 [-462 [22497 ± 4] [+58 [22400 ± 5] [-39]</td>
</tr>
<tr>
<td>10</td>
<td>a</td>
<td>224</td>
<td>22439.2 (22397.2)</td>
<td>21348 ± 7 [22033 ± 6]</td>
<td>-1091 [-406 [22465 ± 5] [+26]</td>
</tr>
<tr>
<td>11</td>
<td>a'</td>
<td>218</td>
<td>21714.3 (21672.3)</td>
<td>22015 ± 17 [22400 ± 10]</td>
<td>+301 [+686]</td>
</tr>
<tr>
<td>12</td>
<td>a</td>
<td>224</td>
<td>22439.2</td>
<td>21321 ± 6</td>
<td>-1118</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(22397.2)</td>
<td>[22009 ± 11]</td>
<td>-430</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[22493 ± 4]</td>
<td>+54</td>
</tr>
<tr>
<td>13</td>
<td>c'</td>
<td>218</td>
<td>21955.5</td>
<td>21519 ± 6</td>
<td>-436</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(21913.5)</td>
<td>21198 ± 15</td>
<td>-757</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[21951 ± 7]</td>
<td>-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[22049 ± 3]</td>
<td>+93</td>
</tr>
<tr>
<td>14</td>
<td>d</td>
<td>217</td>
<td>21850.4</td>
<td>21383 ± 10</td>
<td>-467</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(21808.4)</td>
<td>[21042 ± 4]</td>
<td>-808</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[21819 ± 7]</td>
<td>-31</td>
</tr>
</tbody>
</table>

1 peaks are labeled according to elution order as in Fig. 4.1

2 H1 subtype based on AU gel mobility according to Shannon et al., (1987)

3 length (in residues) and $M_r$ of the $\alpha$-N-acetylated form predicted from the corresponding gene sequences (Coles et al., 1987). The $M_r$ predicted for non-$\alpha$-N-acetylated forms are shown in brackets

4 ± std. dev. Listed in order of apparent abundance in MS. The $M_r$ of relatively minor forms are shown in square brackets.

5 (measured $M_r$) - (predicted $M_r$, $\alpha$-N-acetylated form), integer value
matches the expected $M_i$ (Tables 4.3 and 5.1) argues against this possibility, as does the lack of evidence of proteolysis on the SDS and AU gels run just prior to MS (i.e. "smearing" is not evident in Fig. 4.1B). Moreover, the mass spectra were not consistent with extensive random proteolysis because a limited number of molecular species were observed in each spectra (see Appendix Figs. A1-A3 for representative raw mass spectra). This suggests that if proteolysis was involved, it occurred at specific sites and was the result of a regulated process since nearly the entire population of H1 molecules (except the minor species detected in CX-HILIC peaks 9, 10, 12 and 13) was affected. However, as shown in Appendix Tables A1 and A2, attempts to generate values matching the observed set of $M_i$s from the set of predicted $M_i$s through "simulated proteolysis" (i.e. sequential subtraction of masses corresponding to residues in the N and/or C-terminals) of the sequence predicted for each H1 variant provided, at best, only "loose fits" that lacked the precision typically associated with mass spectral analyses (e.g. Smith et al., 1990; Smith et al., 1991; Chait and Kent, 1992; also see Chapters 5 and 7).

Alternatively, many of the observed $M_i$s in Table 4.2 may differ from the predicted values because of genuine differences, apart from proteolysis, between the actual protein sequences and those predicted from the gene sequences. Such differences could be attributable to inaccuracies in the reported gene sequences, or possibly to genetic variation in chickens such that the gene sequences reported for chickens sampled in Australia (Coles et al., 1987), are not typical of chickens in Canada/North America. An appealing aspect of these arguments over the "proteolysis hypothesis" is the ready explanation for the case of CX-HILIC peak 11 whose $M_i$ exceeded that predicted from the H1a' gene sequence, in addition to the more frequent instance where the observed $M_i$ was smaller than that predicted. Moreover, genetic variation in H1 among the chickens sampled here could also explain the resolution of multiple forms of H1a, H1b and H1c resolved by CX-HILIC of pooled blood $H_{1_{total}}$. These possibilities are examined in Chapter 6.0. Below, experiments to further characterize the nature of the heterogeneity of CX-HILIC peaks of $H_{1_{total}}$ are described.
4.2.3 Ion spray MS characterization of $H_{1_{\text{total}}}$ CX-HILIC peaks

To examine the possibility that procedural artifacts affected the accuracy of the electrospray MS analyses performed at the University of Michigan, a subset of pooled erythrocyte $H_{1_{\text{total}}}$ CX-HILIC peaks were analysed in parallel on a Sciex API III triple quadrupole mass spectrometer at the University of Toronto in collaboration with Drs. H. Pang and M. Cheung. This instrument employs a modified version of the electrospray interface, commonly called ion spray, to introduce samples into the mass spectrometer under conditions that have been suggested to be gentler although the overall process is similar (Bruins et al., 1987). The results of these analyses are summarized in Table 4.3. As was found in Michigan, most of the $M_r$ values obtained by ion spray MS (ISMS) were smaller than expected except for those of CX-HILIC fraction 11. For convenience, the two series of measurements are aligned in Table 4.4. Comparison of the $M_r$ values obtained with the two instruments reveals extensive agreement between these series of measurements. Together with the observation that the $M_r$ observed for H5 protein analysed in parallel on both instruments was identical to the predicted $M_r$ (Tables 4.3 and 5.1), the Toronto data confirms that obtained for $H_{1_{\text{total}}}$ CX-HILIC peaks in Michigan. This strongly suggests that the discrepancies between the predicted and observed $M_r$ values for these $H_1$ variants are not an artifact of MS. However, this comparison does reveal that the MS technique employed influenced the apparent abundance of molecular species within the various $H_1$ peaks analysed. In several instances, molecular species that were abundant components of the electrospray mass spectrum of a fraction appeared to be minor components in ionspray mass spectra (e.g. see the results for CX-HILIC peaks 1, 10 and 11 in Table 4.4). The basis for this discrepancy is not known but probably reflects differences in the mechanism of sample ionization on the two instruments that may also have been influenced by differences in the composition of the solvents employed for the introduction of the samples (see Methods, section 2.7).
Table 4.3. Ion spray MS of selected pooled chicken erythrocyte H1 CX-HILIC fractions.

<table>
<thead>
<tr>
<th>HPLC peak</th>
<th>AU gel subtype</th>
<th>predicted length</th>
<th>predicted $M_r$</th>
<th>measured $M_r$</th>
<th>difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>b</td>
<td>223</td>
<td>22437.3 (22395.3)</td>
<td>21743.5 ± 2.1 [21327.3 ± 4.4]</td>
<td>-694</td>
</tr>
<tr>
<td>4</td>
<td>b</td>
<td>223</td>
<td>22437.3 (22395.3)</td>
<td>21686.4 ± 3.2 [21271.0 ± 4.1]</td>
<td>-751</td>
</tr>
<tr>
<td>5</td>
<td>c</td>
<td>219</td>
<td>21843.4 (21801.4)</td>
<td>21699.6 ± 3.4</td>
<td>-144</td>
</tr>
<tr>
<td>10</td>
<td>a</td>
<td>224</td>
<td>22439.2 (22397.2)</td>
<td>22468.3 ± 1.8 [22036.5 ± 3.2]</td>
<td>+29</td>
</tr>
<tr>
<td>11</td>
<td>a'</td>
<td>218</td>
<td>21714.3 (21672.3)</td>
<td>22397.7 ± 1.8 [22026.4 ± 2.1]</td>
<td>+683</td>
</tr>
<tr>
<td>13</td>
<td>c'</td>
<td>218</td>
<td>21955.5 (21913.5)</td>
<td>21956.7 ± 2.0 [21528.1 ± 1.5]</td>
<td>+1</td>
</tr>
<tr>
<td>14</td>
<td>d</td>
<td>217</td>
<td>21850.4 (21808.4)</td>
<td>21827.6 ± 1.6 [21396.4 ± 2.6]</td>
<td>-23</td>
</tr>
<tr>
<td>H5&lt;sub&gt;total&lt;/sub&gt;</td>
<td></td>
<td>189</td>
<td>20602.7&lt;sup&gt;6&lt;/sup&gt;</td>
<td>20605.8 ± 2.2</td>
<td>+3</td>
</tr>
</tbody>
</table>

<sup>1</sup> peaks are labeled according to elution order as in Fig. 4.1

<sup>2</sup> H1 subtype based on AU gel mobility according to Shannon et al., (1987)

<sup>3</sup> length (in residues) and $M_r$ ($\alpha$-N-acetylated form) predicted from the gene sequences (Coles et al., 1987). The $M_r$ predicted for non-$\alpha$-N-acetylated forms are shown in brackets.

<sup>4</sup> ± std. dev. Listed in order of apparent abundance in MS. The $M_r$ of relatively minor forms are shown in square brackets.

<sup>5</sup> (measured $M_r$) - (predicted $M_r$, $\alpha$-N-acetylated form), integer value.

<sup>6</sup> predicted for non-$\alpha$-N-acetylated H5 from the protein sequence of the allelic variant containing Arg at position 15 (Briand et al., 1980).
Table 4.4. Comparison of electrospray and ion spray MS of selected pooled chicken erythrocyte H1 CX-HILIC fractions.

<table>
<thead>
<tr>
<th>HPLC peak</th>
<th>AU gel subtype</th>
<th>predicted length</th>
<th>predicted M$_r$</th>
<th>ESMS M$_r$</th>
<th>ISMS M$_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>b</td>
<td>223</td>
<td>22438.3 (22395.3)</td>
<td>21318 ± 6</td>
<td>[21327.3 ± 4.4]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21004 ± 8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[20693 ± 8]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[21740 ± 10]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21743.5 ± 2.1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>b</td>
<td>223</td>
<td>22438.3 (22395.3)</td>
<td>20956 ± 12</td>
<td>21686.4 ± 3.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21263 ± 8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[21271.0 ± 4.1]</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>c</td>
<td>219</td>
<td>21844.5 (21801.4)</td>
<td>20991 ± 6</td>
<td>21699.6 ± 3.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[21686 ± 6]</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>a</td>
<td>224</td>
<td>22440.2 (22397.2)</td>
<td>21348 ± 7</td>
<td>[21358.0 ± 3.3]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[22033 ± 6]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[22465 ± 5]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22468.3 ± 1.8</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>a'</td>
<td>218</td>
<td>21715.4 (21672.3)</td>
<td>22015 ± 17</td>
<td>[22026.4 ± 2.1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22400 ± 10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22397.7 ± 1.8</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>c'</td>
<td>218</td>
<td>21956.6 (21913.5)</td>
<td>21519 ± 6</td>
<td>21528.1 ± 1.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21198 ± 15</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[21951 ± 7]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[22049 ± 3]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21956.7 ± 2.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[22441.8 ± 1.7]</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>d</td>
<td>217</td>
<td>21851.5 (21808.4)</td>
<td>21383 ± 10</td>
<td>21396.4 ± 2.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[21042 ± 4]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[21819 ± 7]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21827.6 ± 1.6</td>
<td></td>
</tr>
</tbody>
</table>

---

1 peaks are labeled according to elution order as in Fig. 4.1.

2 H1 subtype based on AU gel mobility according to Shannon et al., (1987).

3 length (in residues) and M$_r$ (α-N-acetylated form) predicted from the corresponding gene sequences (Coles et al., 1987). The M$_r$ predicted for α-N-non-acetylated forms are shown in brackets.

4 ± std. dev. Listed in order of apparent abundance in ESMS. The M$_r$ of relatively minor forms are shown in square brackets.

5 ± std. dev. Listed in order of best match with the preceding column. Alignment with a particular ESMS M$_r$ value indicates a match, most of which are within the standard deviations of these measurements. ISMS values that are not matched are listed on separate lines. The M$_r$ of relatively minor forms are shown in square brackets.
Chapter 5. Characterization of CX-HILIC fractions of $H_5^{\text{total}}$

from pooled erythrocytes
5.1 Introduction

The heterogeneity of $H_{1_{\text{total}}}$ CX-HILIC fractions and the discrepancies between the $M_s$s determined for these fractions and the values predicted from the six published $H_1$ gene sequences that became apparent as MS analyses of $H_1$ proteins were pursued (Chapter 4), raised the issue of whether heterogeneity in size was a characteristic of $H_5$. $H_5$ is encoded by a single copy gene in chickens (Ruiz-Carrillo et al., 1983) and the amino acid sequence predicted by the gene matches the protein sequence determined by Edman degradation (Briand et al., 1980). Thus, it was expected that aside from the Arg>Gln allelic variation at residue 15 of chicken $H_5$ described previously (Greenaway and Murray, 1971), the “expected $M_r$,” derived from these sequences could be used as the ultimate arbiter in MS analyses of $H_{5_{\text{total}}}$. If $H_{5_{\text{total}}}$ was found to be heterogeneous in size (apart from post-translational modifications), then this would be evidence of genetic variation or of processing at the protein or mRNA levels that contribute to heterogeneity in linker histones in general. In contrast, if $H_{5_{\text{total}}}$ was found to be homogenous apart from post-translational modifications, then this would be evidence of genetic variation or processing of mRNAs or protein that was specifically limited to the $H_1$ group and which in turn would support the notion of functional differences between $H_5$ and the variants of $H_1$ in chickens.

In this chapter, I describe analyses of $H_{5_{\text{total}}}$ recovered from the gel filtration purification of pooled erythrocyte $H_{1+H5}$ described in Chapter 3. Even though slight contamination by polypeptides other than $H_5$ was apparent in some of the fractions pooled for $H_{5_{\text{total}}}$ (Fig 3.1). PAGE analyses revealed a single predominant band for $H_{5_{\text{total}}}$ on SDS gels (Fig. 5.1), whereas two bands, an abundant faster migrating form and a less abundant slower form, were resolved on long AU gels (e.g. Fig. 6.1). $H_{5_{\text{total}}}$ was unexpectedly heterogeneous, giving rise to 2 major peaks, and more than 12 minor peaks, that were well-resolved by CX-HILIC. Most of these CX-HILIC peaks were heterogeneous in secondary RP-HPLC employed to recover the proteins. Electrospray MS of the proteins recovered from the CX-HILIC peaks revealed that the most abundant form was the expected size and that in almost all cases, minor forms were larger than expected. The $M_r$ values determined for the minor forms are consistent with limited degrees of genetic variation and post-translational modification and are in sharp contrast to the finding
described in Chapter 4 that almost all of the H1 proteins from the same sample were smaller than expected.

5.2 Results and Discussion

5.2.1 Chromatographic heterogeneity of H5\textsubscript{total}

Given the unexpectedly high degree of heterogeneity apparent for H1\textsubscript{total} in CX-HILIC, it was of interest to see if similar heterogeneity existed within H5\textsubscript{total} obtained from the same Bio-Gel P100 fractionation of pooled erythrocyte H1+H5. Two major peaks and numerous minor peaks were resolved when H5\textsubscript{total} was chromatographed on an analytical PolyCAT A column using CX-HILIC conditions as described in Methods (data not shown). To obtain sufficient amounts of these proteins for further characterization, 2.5 mg of H5\textsubscript{total} was separated on a semi-preparative column of PolyCAT A (9.4 x 200 mm) as shown in Fig. 5.1A. Fractions comprising the peaks labeled A-M were pooled and a Speed-Vac apparatus was used to evaporate most of the acetonitrile. Protein in each peak was then desalted by RP-HPLC using volatile solvents (TFA/CH\textsubscript{3}CN) and recovered by drying in a Speed-Vac. Minor heterogeneity was apparent for many of the H5\textsubscript{total} CX-HILIC peaks in the RP-HPLC step as shown in the Fig. 5.1B. Most of the H5\textsubscript{total} CX-HILIC peaks contained one or two minor species which eluted just before, or just after, the major peak in RP-HPLC (this was not observed when CX-HILIC fractions of H1\textsubscript{total} were chromatographed under the same conditions, data not shown). These RP-HPLC forms were pooled prior to recovery in all cases except for peaks E, F and K (it was expected that different forms would be detected in MS regardless of whether introduced as separate RP-HPLC peaks or as a mixture).

5.2.2 PAGE identification of CX-HILIC fractions

The majority of the H5 CX-HILIC fractions appeared to be homogenous and possessed mobility equivalent to bulk H5 in the starting material on SDS gels (Fig. 5.1C). Of the three polypeptides resolved for peak A, the two most abundant species migrated more rapidly than bulk H5 and most likely represent degraded forms although it is possible that they represent structural variants or post-translationally modified forms. In all of the remaining fractions, the most
Figure 5.1. Chromatographic heterogeneity of pooled erythrocyte H5\textsubscript{total}.
(A) H5\textsubscript{total} (approx. 2.5 mg) was separated on a semi-preparative column (9.4 x 200 mm) of PolyCAT A (30 nm pore) with a four hour linear gradient from 500 mM to 680 mM NaClO\textsubscript{4} in 70% CH\textsubscript{3}CN, 10 mM propionic acid-NaOH pH 6.5 and a flow rate of 2.5 ml/min. Fractions comprising peaks A-M were pooled and most of the acetonitrile evaporated using a Speed-Vac.

(B) Partially dried fractions from (A) were desalted using RP-HPLC. Fractions were injected onto an analytical C18 column and after sufficient washing to purge all salt from the column, proteins were eluted with a linear one hour gradient from 5% to 60% CH\textsubscript{3}CN in 0.1% TFA at 0.8 ml/min. Two peaks were evident for almost all the CX-HILIC peaks as described in the text. Proteins were recovered by vacuum drying in a Speed-Vac.

(C) Aliquots (approx. 500 ng) of the fractions recovered in (B) were analysed on a 12% SDS gel. Lane S = 100 ng pooled H5\textsubscript{total} starting material, 1/5 = crude H1+H5 from chicken erythrocytes. A-M = corresponds to the peaks in (A). For peaks E, F and K, the two peaks resolved during recovery by RP-HPLC were kept separated for analysis. In all other instances, they were pooled. Proteins detected by silver staining.
abundant species possessed the same SDS gel mobility as bulk H5_total recovered from the P100 column. Note also that the different RP-HPLC forms of peaks E (E-1 and E-2), F (F-1 and F-2) and K (K1 and K2) had mobility equivalent to that of bulk H5_total.

5.2.3 Electrospray MS characterization of H5_total CX-HILIC fractions

Electrospray MS revealed that most of the H5 CX-HILIC peaks were slightly heterogeneous (Table 5.1). Minor species were detected in most peaks, however these were generally less abundant than the secondary forms detected by electrospray MS in CX-HILIC peaks of H1_total (see appendix Figs. A1 and A3). In most instances, the small peaks resolved by RP-HPLC during recovery of CX-HILIC peaks were included with the major peak when pooling fractions, so these minor forms may correspond to the smaller RP-HPLC peaks in some instances. However, note that the M_r values determined for F-1 and F-2, the two RP-HPLC peaks resolved during recovery of CX-HILIC peak F are essentially identical (Table 5.1). The basis for the resolution of any of the H5_total CX-HILIC peaks into two or more components by RP-HPLC is unknown. It does not seem likely that this is due to artifact since others have separated two peaks for bulk H5 using RP-HPLC (Helliger et al., 1988) and as shown in Fig. 3.2, as many as four components of bulk H5 can be separated using a different RP-HPLC method. The fact that the mass spectra of F-1 and F-2 were so similar suggests the possibility that these proteins could be allelic variants of H5 containing substitutions (e.g. leu to ile) or other differences (short sequence inversions) that do not alter the M_r of the protein. Alternatively, because deamidation of Asn or Gln results in the loss of only 1 mass unit, these may correspond to forms of the same protein with differences in amide content that are too small to be distinguished by the MS procedure employed.

In general, structural analyses have determined that in higher eukaryotes, the α-amino terminus of linker histones (and also histones H2A and H4) is usually acetylated (e.g., Von Holt et al., 1989; van Holde, 1989). Known exceptions to this are H5 in birds and fish (Von Holt et al., 1989) and the related protein H1° expressed in some mammalian cells (Lindner et al., 1998). The M_r 20606 observed for the most abundant species in both F-1 and F-2 matches the M_r predicted for
non-α-N-acetylated H5, indicating that the sequence of the majority of the H5 molecules in the pooled erythrocyte sample match that predicted from the gene sequence and, in addition, are not α-N-acetylated. The absence of acetylation is consistent with the finding that the N-terminus of chicken H5 can be sequenced directly by the Edman degradation method (Sautiere et al., 1976) and is further corroborated by the M, 20606 observed for bulk H5 (i.e. the P100 fraction) by ion spray MS (Table 4.3). The major species detected in peaks F', H and I also match the predicted M, for non-α-N-acetylated H5 while the M, for the major species in peaks D and G fall between the values predicted for acetylated and non-acetylated H5. Interestingly, the M, 20645 of the major species in peak C, the second-most abundant CX-HILIC component, coincides with the value predicted for acetylated H5. The likelihood that peak C may correspond to α-N-acetylated H5 is strengthened by the inability of investigators using metabolic labeling with 3H- or 14C-acetate to monitor α-N-acetylation to detect labeling of H5 (or any linker histone). Although the function of α-N-acetylation is unknown, a special role in the metabolism of H5 and related proteins is suggested by the fact that aside from the data reported here, evidence for incomplete histone α-N-acetylation has only been described for the related variant H1° in human and mouse tissues (Lindner et al., 1998) (see Discussion).

Another remarkable finding is that the forms of H5 observed within CX-HILIC peaks typically differed by increments of 100 mass units (e.g. the 20645, 20744 and 20844 species observed for peak C or the 20606, 20705 and 20807 species observed for peak F-1) (Table 5.1). This distribution of masses is consistent with phosphorylation and/or sulphation of H5 since each occurrence of these modifications results in an increase of 80 mass units. Sulphation of histones in vivo has not been reported but H5 is known to be multiply phosphorylated in chicken erythroblasts (Sung, 1977). The minor forms of H5 detected may reflect the presence of immature erythroid cells in the sample or residual phosphorylation remaining after the pronounced dephosphorylation of H5 which accompanies transcriptional down-regulation and chromatin condensation in the final
<table>
<thead>
<tr>
<th>HPLC peak</th>
<th>predicted length</th>
<th>predicted $M_r$</th>
<th>measured $M_r$</th>
<th>difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>189</td>
<td>20602.7</td>
<td>20677 ± 4</td>
<td>+74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(20645.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>20645 ± 10</td>
<td>+42</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[20744 ± 12]</td>
<td>+141</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[20844 ± 7]</td>
<td>+241</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td>20633 ± 9</td>
<td>+30</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[20730 ± 11]</td>
<td>+127</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[20828 ± 5]</td>
<td>+225</td>
<td></td>
</tr>
<tr>
<td>E-1</td>
<td></td>
<td>20660 ± 8</td>
<td>+57</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[20764 ± 10]</td>
<td>+161</td>
<td></td>
</tr>
<tr>
<td>E-2</td>
<td></td>
<td>20648 ± 11</td>
<td>+45</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[18332 ± 5]</td>
<td>-227</td>
<td></td>
</tr>
<tr>
<td>F-1</td>
<td></td>
<td>20606 ± 8</td>
<td>+3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[20705 ± 5]</td>
<td>+102</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[20807 ± 6]</td>
<td>+204</td>
<td></td>
</tr>
<tr>
<td>F-2</td>
<td></td>
<td>20605 ± 8</td>
<td>+2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[20700 ± 4]</td>
<td>+97</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td>20598 ± 10</td>
<td>-5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[20694 ± 12]</td>
<td>+91</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td>20613 ± 6</td>
<td>+10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[20713 ± 10]</td>
<td>+110</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[20813 ± 7]</td>
<td>+210</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td></td>
<td>20603 ± 7</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[20700 ± 8]</td>
<td>+97</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[20800 ± 9]</td>
<td>+197</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td></td>
<td>20606 ± 8</td>
<td>+3</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td></td>
<td>20716 ± 8</td>
<td>+113</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td></td>
<td>20683 ± 7</td>
<td>+80</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[20779 ± 6]</td>
<td>+176</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td></td>
<td>20696 ± 9</td>
<td>+93</td>
<td></td>
</tr>
</tbody>
</table>
peaks are labeled according to elution order as in Fig. 5.1. The amount of peaks A, L and K-1 recovered was not sufficient for MS analysis.

Length (in residues) and $M_r$ for non-$\alpha$-N-acetylated H5 as predicted from the protein sequence of the variant containing Arg at position 15 (Briand et al., 1980). The $M_r$ predicted for the $\alpha$-N-acetylated form is shown in brackets.

$\pm$ std. dev. Listed in order of apparent abundance in MS. The $M_r$ of relatively minor forms are shown in square brackets.

$(\text{measured } M_r) - (\text{predicted } M_r, \text{ non-$\alpha$-N-acetylated form})$, integer value.
stages of avian erythrocyte maturation (Sung, 1977). However, the fact that these species differed by multiples of 100 units rather than multiples of 80 units suggests that additional modifications (e.g. lysine methylation or hydroxylation, each of which add 15 mass units) were present or that other factors such as amino acid substitutions were involved.

5.2.4 Chromatographic heterogeneity of H5 from a single chicken

To examine the extent to which genetic variation contributed to the chromatographic heterogeneity of H5 total from pooled erythrocytes, approx. 2.5 mg of Bio-Gel P100 purified H5, prepared from the blood of a single chicken as described in Methods section 2.12, was chromatographed on the same column used in Fig. 5.1. The CX-HILIC profiles of H5 total from pooled blood and H5 from a single chicken are shown aligned in Fig. 5.2. While the two traces are similar overall, note that readily discernable differences exist. The relative proportion of peak C, the second largest peak of the pooled sample (and putatively N-acetylated) is diminished in the single bird sample while that of peak G in the single bird sample is markedly increased relative to the pooled sample. Peak H also appears to less abundant in the single bird sample compared to the pooled sample. However, aside from these changes, the abundance of the remaining CX-HILIC peaks appear to be equivalent in the two samples. Most of the single bird CX-HILIC peaks were separated into 2-3 forms by RP-HPLC during recovery as described for the CX-HILIC peaks from the pooled sample (data not shown). The CX-HILIC fractions of single chicken H5 were not characterized further but the overall similarities of the two CX-HILIC profiles (Fig. 5.2) suggests that only a limited degree of genetic polymorphism (that can be resolved by CX-HILIC) exists in H5. These data also support the notion, as suggested by the MS characterization of the pooled blood H5 CX-HILIC peaks, that many of the minor CX-HILIC peaks are attributable to post-translational modification rather than genetic variation.
Figure 5.2. Chromatographic heterogeneity of H5 from a single chicken.

The CX-HILIC profile of H5 from a single chicken (A) is aligned to a portion of the separation of pooled blood H5\textsubscript{total} (B) shown in Fig. 5.1 to facilitate comparison. Single chicken H5 was prepared from crude H1+H5 by Bio-Gel P100 chromatography as described in Methods. Approx. 2.5 mg was separated on the same column employed for the separation of pooled blood H5\textsubscript{total} shown in Fig. 5.1 using the same conditions. A 14 minute shift in time (out of a total gradient duration of 4 hours) between the two traces was required to align the two separations performed six weeks apart. Peaks in (A) and (B) are labeled as in Fig. 5.1 but the identifications in (A) are assumed on the basis of the alignment of the traces and were not confirmed by ancillary analyses. Apart from increases and decreases, respectively, in the relative amount of peaks G and H in the single chicken compared to the pooled sample, the two profiles appear to be very similar overall.
Chapter 6. Investigations of the origin of the heterogeneity of pooled erythrocyte $H_1_{total}$
6.1 Introduction

The resolution of more than six peaks for pooled blood H\textsubscript{1\,total} by CX-HILIC was unexpected and in this chapter I describe a series of experiments performed to elucidate the basis of the heterogeneity that was discovered. Prior to initiating MS analyses of CX-HILIC peaks of H\textsubscript{1\,total} and H\textsubscript{5\,total}, I performed chemical and enzymatic treatments to test the possibility that post-translational modifications or artifacts related to methods of extraction contributed to the CX-HILIC heterogeneity of H\textsubscript{1\,total}. As described here in section 6.2.1, these analyses suggest that detectable amounts of known (phosphorylation and poly(ADP-ribosylation)) and unknown (glycosylation) modifications were not present in H\textsubscript{1\,total}.

When the MS data for the CX-HILIC peaks of H\textsubscript{1\,total} was obtained (Tables 4.2 and 4.3), it became apparent that the high degree of chromatographic heterogeneity was probably related to the discrepancies between the expected and observed size of the proteins in CX-HILIC peaks of H\textsubscript{1\,total}. The possibility that non-specific proteolysis was the cause of the smaller than expected size and heterogeneity of the H\textsubscript{1} proteins became highly unlikely when MS analyses revealed that virtually all the H\textsubscript{5} proteins from the same sample of pooled blood were the expected size or larger (Table 5.1). Additionally, the similarity of the chromatographic profiles of H\textsubscript{5} from the single bird and the pooled sample was consistent with only a limited degree of genetic variation. Taken together, the data of Chapters 4 and 5 suggested two possible explanations for the smaller than expected size and chromatographic heterogeneity of the pooled blood H\textsubscript{1} proteins. The first potential explanation ensues from the fact that immature avian erythroid cells (erythroblasts) contain more H\textsubscript{1} than H\textsubscript{5} whereas mature erythrocytes contain more H\textsubscript{5} than H\textsubscript{1} (Weintraub, 1975; Urban et al., 1980). Although there is disagreement over whether H\textsubscript{5} replaces H\textsubscript{1} to some extent or simply accumulates without compensatory loss of H\textsubscript{1} (Bates and Thomas, 1981), it seemed possible that partial proteolysis of H\textsubscript{1} could accompany the deposition of H\textsubscript{5} in erythroblast chromatin and that this event might occur only in erythroid cells since H\textsubscript{5} expression is restricted to these cells in chickens (Neelin et al., 1964; van Holde, 1989). To investigate whether erythroid-specific proteolysis of H\textsubscript{1} occurred, CX-HILIC fractions of chicken liver H\textsubscript{1} were prepared and characterized by MS to see if the sizes of liver H\textsubscript{1} variants were closer to the
size predicted by the published gene sequences. The data described here in sections 6.2.2-6.2.4 do not support the existence of erythroid-specific proteolysis of H1. CX-HILIC variants of liver H1 were found to be the same size in most cases, as their counterparts from pooled erythrocytes.

A second potential explanation was that genetic variation in H1 among chickens, which is more extensive than that suggested for H5 in Chapter 5, was responsible for the size discrepancies and the heterogeneity and multiplicity of CX-HILIC peaks for pooled blood H1\textsubscript{total}. Polymorphisms in H1 genes could give rise to the multiple chromatographic forms observed for some variants. Moreover, because the gene sequence reported for each of the non-allelic variants of chicken H1 (Coles et al., 1987) actually represents a specific allele cloned during the course of the work, the single sequence reported for each “non-allelic” variant may be one of several, or possibly many, alleles that exist in chickens. Thus the differences in size observed between the “expected” and actual H1 proteins described in Chapter 4 may reflect the extent of allelic variation in chicken H1 more than presumed proteolysis. Unequivocal proof of this hypothesis requires the cloning and sequencing of several alleles for each “non-allelic” H1 variant gene in conjunction with protein analyses to demonstrate the existence of the corresponding protein in the same individuals in which specific alleles were identified. This Herculean task is clearly outside the scope of this thesis but in sections 6.2.5 and 6.2.6, I describe evidence which demonstrates that genetic variation is a significant cause of the chromatographic heterogeneity of pooled blood H1\textsubscript{total} and which suggests that this variation is likely to be the cause of the discrepancies in the expected and observed sizes of the chicken H1 proteins.

6.2 Results and Discussion

6.2.1 Apparent absence of common post-translational modifications.

Because the resolution of more than six peaks for pooled blood H1\textsubscript{total} by CX-HILIC was unexpected, the involvement of common and potential post-translational modifications of H1 was examined prior to pursuing MS characterization. Glycosylation, phosphorylation and poly(ADP-ribosylation) did not appear to be involved because treatment of H1\textsubscript{total} with glycopeptidase F, alkaline phosphatase or dilute sodium hydroxide, as described in Methods section 2.13, prior to
chromatography did not alter the CX-HILIC profile of H1total (data not shown). None of the heterogeneity appeared to be attributable to acid extraction since the AU-PAGE and CX-HILIC profiles of H1total prepared by salt extraction were identical to those prepared by acidic extraction (data not shown). Because analyses of histone phosphorylation and acetylation have revealed that modified forms differing by one unit of net charge can be resolved on AU gels (Lennox and Cohen, 1989; Talasz et al., 1996), the observation that the multiple H1total CX-HILIC peaks for H1a had identical mobilities in AU-PAGE, as did the multiple peaks for H1b and H1c, strongly suggested that these forms were not differentially acetylated or phosphorylated.

6.2.2 CX-HILIC heterogeneity of chicken liver H1.

To prepare liver H1 fractions whose MS spectra could be compared to those of pooled blood H1 total, two chickens were anesthetized and perfused with heparinized saline as described in Methods section 2.10 prior to tissue collection in order to minimize contamination of liver H1 by erythrocyte H1. Crude H1 was prepared from liver and other organs by direct extraction of tissues with 5% PCA. SDS and AU gel analyses revealed that the content of H5 in the crude liver extract was very low (Fig. 6.1), indicating that the vast majority of H1 in the sample was from liver cells rather than erythrocytes. Comparison of the liver extract and pooled erythrocyte H1 on AU gels revealed noticeable differences in the non-allelic variant composition of the two tissues. Liver H1 contained much less H1a' and H1c and much more H1c' relative to erythrocyte H1 (Fig. 6.1). These same differences appeared to be common to heart, kidney and spleen H1 also. Analytical CX-HILIC chromatograms using equivalent sample loads and identical separation conditions revealed that crude liver H1 was less heterogenous than pooled erythrocyte H1. Nine peaks were resolved for liver H1 under conditions that resolved thirteen major peaks for pooled erythrocyte H1 (Fig. 6.2). Some of this decreased heterogeneity may have been attributable to the diminished content of subtypes H1a' and H1c in the liver sample (Fig. 6.1). However, because the liver sample represented only two animals whereas over 100 birds contributed to the pooled erythrocyte sample, a decreased content of allelic variants in the liver sample may also have contributed to the simpler nature of the CX-HILIC profile. Note, as suggested by the AU gel analysis in Fig. 6.1, that the relative proportions of peaks containing subtypes H1a, H1b, H1c' and H1d, (identified in
Figure 6.1. PAGE analyses of chicken organ linker histone content.

(A) Aliquots (approx. 5 and 20 ug) of 5% PCA soluble protein prepared by direct extraction of organs from saline-perfused chickens were analysed on a 12% SDS gel. H1t and H5t = pooled blood H1_total (2 ug) and H5_total (1 ug), respectively. Proteins detected by Coomassie blue staining. The relative positions of the six H1 variants and H5 are shown on the right side. Note that the levels of H5 in the organ extracts are extremely low, indicating that the perfusion flushed the majority of erythrocytes from the tissues.

(B) Analysis of the same samples in (A) on a 50 cm long AU gel. H1t and H5t = pooled blood H1_total and H5_total, 1 ug of each, respectively. Aliquots representing approx. 2.5 ug total protein were loaded for the organ extracts. Proteins were detected by silver staining. The relative positions of the H1 variants and H5 are shown on the right side. The asterisk indicates the putative monophosphorylated form of H5 (see Chapter 5).
Figure 6.2. Comparison of the CX-HILIC profiles of pooled blood and liver H1.

Aliquots representing approx. 100 µg of H1 prepared from liver tissue pooled from two chickens (A) and pooled chicken erythrocyte H1 total (B) were chromatographed on an analytical column (4.6 x 200 mm) of PolyCAT A (100 nm diam. pore) using a four hour linear gradient from 420 mM to 620 mM NaClO4 in 70% CH3CN, 10 mM propionic acid-NaOH pH 6.5 and a flow rate of 0.8 ml/min. Lowercase letters indicate the predominant H1 subtype in each peak based on AU gel analyses.
Figure 6.3. CX-HILIC and electrophoretic heterogeneity of chicken liver H1.

(A) Approx. 2.0 mg of crude liver PCA extract was separated on a semi-preparative column (9.4 x 200 mm, 100 nm diam. pore) of PolyCAT A using a four hour linear gradient from 410 mM to 620 mM NaClO₄ in 70% CH₃CN, 10 mM propionic acid-NaOH pH 6.5, using a flow rate of 2.5 ml/min. Fractions comprising peaks A-L from several such runs were pooled and in some cases were rechromatographed to reduce cross-contamination between peaks. Fractions representing peak F were accidentally lost during recovery.

(B) Aliquots representing approx. 500 ng of the peaks recovered in (A) were analysed on a 12% SDS gel. M = molecular weight markers (Pharmacia). H5, = approx. 100 ng of pooled blood H5_total. H1, = approx. 1 ug of crude liver PCA soluble protein. Proteins were detected by silver staining. The relative positions of the six H1 variants and H5 are shown on the right side.

(C) Aliquots representing approx. 500 ng of the peaks recovered in (A) were analysed on a 50 cm long AU gel. H1, and H5, = pooled blood H1_total (2 ug) and H5_total (500 ng), respectively. H1, = approx. 2.5 ug crude liver PCA soluble protein. Proteins were detected by silver staining. The relative positions of the six H1 variants and H5 are shown on the right side. The asterisk indicates a putative monophosphorylated form of H5.
the liver chromatogram as described below) were markedly different in the two chromatograms (Fig. 6.2).

6.2.3 PAGE identification of liver H1 CX-HILIC peaks

To prepare sufficient amounts of the liver H1 CX-HILIC peaks for subsequent characterization, fractions from several separations of 2 mg of crude liver H1 on a semi-preparative column of PolyCAT A (9.4 x 200 mm) like that shown in Fig. 6.3 were pooled. Rechromatography on Poly CAT A to diminish cross-contamination was performed for some, but not all, fractions. Analyses of the pooled proteins after recovery by TCA precipitation on SDS and AU gels is shown in Fig. 6.3B and C. AU gel analyses revealed that peaks A-D contained H1b, peaks E and I contained H1a, peaks J and K contained H1c' and peak L contained H1d. Although peaks A, B, C and L appeared to be homogenous on both SDS and AU gels, some cross-contamination among variants was apparent in peaks D, E, I, J and K.

6.2.4 Ion spray MS characterization of liver H1 CX-HILIC peaks

As was found for the CX-HILIC peaks of pooled blood H1total, the M_r of many of the liver H1 proteins resolved by CX-HILIC was smaller than predicted by the corresponding gene sequences (Table 6.1). The largest discrepancies in this case were found for peaks A-D. The M_r of these proteins suggested they contained approx. 5 to 10 fewer residues than the predicted H1b protein. The major species detected in peaks E-L matched the M_r predicted for the corresponding subtypes H1a, H1c' and H1d much more closely. The major M_r of peak E matched that predicted for H1a while the major species in peak L, which also resembled H1a on AU gels, was 26 mass units larger than predicted for H1a. Similarly, the major M_r of peak J matched the value predicted for H1c' whereas the major species in peak K was 29 mass units smaller (Table 6.1). Thus, the major species in peaks E and J are likely to be identical to the predicted H1a and H1c' proteins, respectively, while the major species in peaks I and K are likely to be allelic variants of H1a and H1c' that differ from the predicted sequences at a limited number of positions since they differed from the predicted M_r by less than the mass of one residue.
All of the chicken liver CX-HILIC peaks except A were heterogenous and contained a second molecular species that was 413-432 mass units smaller than the major form detected in the same peak. As discussed in Chapter 7 for the NBS fragments of \( H_{1\text{mol}} \) CX-HILIC peaks 10 and 12, these may have been generated from the major species by removal of the first four N-terminal residues (see Appendix Table A1). Note that this putative loss of four residues appeared to be the primary cause of heterogeneity within the liver \( H_{1} \) CX-HILIC fractions. Although this trend can be discerned when the observed \( M_{r} \) values within the entries for peaks 10 and 12 in Table 4.2 are compared (discussed in Chapter 7) the greater variation in the \( M_{r} \) differences separating the observed sizes of the pooled blood CX-HILIC peaks from the expected sizes (Tables 4.2 and 4.3) suggests that mechanisms other than this putative N-terminal clipping, such as allelic variation, contribute to the many of the size discrepancies in this case.

Comparison of the molecular masses determined by ion spray MS for liver and pooled erythrocyte \( H_{1} \) CX-HILIC fractions did not support the notion that erythroid-specific proteolysis was responsible for the discrepancies between the predicted and observed masses for the pooled erythrocyte \( H_{1} \) fractions. As shown in Table 6.2, the masses for liver CX-HILIC peaks A (\( H_{1b} \)), B (\( H_{1b} \)), I (\( H_{1a} \)), J (\( H_{1c'} \)) and L (\( H_{1d} \)) were essentially identical to those determined for pooled erythrocyte peaks 1 (\( H_{1b} \)), 4 (\( H_{1b} \)), 10 (\( H_{1a} \)), 13 (\( H_{1c'} \)) and 14 (\( H_{1d} \)), respectively. However, close matches for the masses of alternate forms of \( H_{1b} \) (peaks C and D), \( H_{1a} \) (peak E) and \( H_{1c'} \) (peak K) in liver were not observed in the subset of pooled erythrocyte fractions selected for analysis by ion spray MS. Comparison to masses obtained by electrospray MS (Table 4.2) suggests that liver peak B is probably better matched to pooled erythrocyte peaks 2 or 3 and that liver peak E may be comparable to pooled erythrocyte peak 7. These putative alignments are consistent with the identification of these proteins as \( H_{1b} \) and \( H_{1a} \), respectively, on AU gels. Notably, there did not appear to be a match for liver peak K in the pooled blood sample. Liver peaks J and K both migrated like \( H_{1c'} \) whereas pooled erythrocyte \( H_{1} \) contained only one chromatographic form of \( H_{1c'} \) (peak 13), whose molecular mass was best matched by liver peak J (Table 6.2). This observation suggests the possibility that peak K represents an \( H_{1c'} \) sequence expressed in liver but not erythrocytes or that peak K is an allelic variant of \( H_{1c'} \) that was not
Table 6.1. Ion spray MS of chicken liver H1 CX-HILIC fractions.

<table>
<thead>
<tr>
<th>HPLC peak</th>
<th>AU gel subtype</th>
<th>predicted length</th>
<th>predicted $M_r$</th>
<th>measured $M_r$</th>
<th>difference $^5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>b</td>
<td>223</td>
<td>22437.3</td>
<td>21741.8 ± 4.8</td>
<td>-695</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(22395.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>b</td>
<td>223</td>
<td>22437.3</td>
<td>21711.3 ± 5.0</td>
<td>-726</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(22395.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>b</td>
<td>223</td>
<td>22437.3</td>
<td>21742.9 ± 4.0</td>
<td>-694</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(22395.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>b</td>
<td>223</td>
<td>22437.3</td>
<td>21911.0 ± 4.8</td>
<td>-526</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(22395.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>a</td>
<td>224</td>
<td>22439.2</td>
<td>22437.5 ± 4.0</td>
<td>-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(22397.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>a</td>
<td>224</td>
<td>22439.2</td>
<td>22465.7 ± 3.9</td>
<td>+26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(22397.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>c'</td>
<td>218</td>
<td>21955.5</td>
<td>21954.2 ± 4.8</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(21913.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[22006.6 ± 4.5]</td>
<td></td>
<td>+51</td>
</tr>
<tr>
<td>K</td>
<td>c'</td>
<td>218</td>
<td>21955.5</td>
<td>21926.2 ± 4.1</td>
<td>-29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(21913.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>d</td>
<td>217</td>
<td>21850.4</td>
<td>21822.1 ± 4.0</td>
<td>-28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(21808.4)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

$^1$ peaks are labeled according to elution order as in Fig. 6.3

$^2$ H1 subtype based on AU gel mobility according to Shannon et al., (1987)

$^3$ length (in residues) and $M_r$ ($\alpha$-N-acetylated form) predicted from the corresponding gene sequences (Coles et al., 1987). The $M_r$ predicted for non-$\alpha$-N-acetylated forms are shown in brackets

$^4$ ± std. dev. Listed in order of apparent abundance in MS. The $M_r$ of relatively minor forms are shown in square brackets.

$^5$ (measured $M_r$) - (predicted $M_r$, $\alpha$-N-acetylated form), integer value
### Table 6.2. Comparison of selected chicken liver and erythrocyte H1 CX-HILIC fractions by ion spray MS.

<table>
<thead>
<tr>
<th>HPLC peak&lt;sup&gt;1&lt;/sup&gt;</th>
<th>AU gel subtype&lt;sup&gt;2&lt;/sup&gt;</th>
<th>predicted M&lt;sub&gt;r&lt;/sub&gt;&lt;sup&gt;3&lt;/sup&gt; (liver)</th>
<th>measured M&lt;sub&gt;r&lt;/sub&gt;&lt;sup&gt;4&lt;/sup&gt; (liver)</th>
<th>Δ&lt;sup&gt;5&lt;/sup&gt;</th>
<th>measured M&lt;sub&gt;r&lt;/sub&gt;&lt;sup&gt;4&lt;/sup&gt; (erythrocyte)</th>
<th>Δ&lt;sup&gt;5&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/1</td>
<td>b</td>
<td>22437.3 (22395.3)</td>
<td>21741.8 ± 4.8</td>
<td>-695</td>
<td>21743.5 ± 2.1</td>
<td>-694</td>
</tr>
<tr>
<td>B/4&lt;sup&gt;6&lt;/sup&gt;</td>
<td>b</td>
<td>22437.3 (22395.3)</td>
<td>21711.3 ± 5.0</td>
<td>-726</td>
<td>21686.4 ± 3.2</td>
<td>-751</td>
</tr>
<tr>
<td>I/10</td>
<td>a</td>
<td>22439.2 (22397.2)</td>
<td>22465.7 ± 3.9</td>
<td>+26</td>
<td>22468.3 ± 1.8</td>
<td>+29</td>
</tr>
<tr>
<td>J/13</td>
<td>c'</td>
<td>21955.5 (21913.5)</td>
<td>21954.2 ± 4.8</td>
<td>-1</td>
<td>21956.7 ± 2.0</td>
<td>+1</td>
</tr>
<tr>
<td>L/14</td>
<td>d</td>
<td>21850.4 (21808.4)</td>
<td>21822.1 ± 4.0</td>
<td>-28</td>
<td>21827.6 ± 1.6</td>
<td>-23</td>
</tr>
</tbody>
</table>

<sup>1</sup> peaks are labeled according to elution order (liver/erythrocyte) as in Fig. 6.3 and Fig. 4.1. Pooled erythrocyte H1 peaks 5 and 11 described in Table 4.3 have been omitted since these proteins (H1c and H1a', respectively) were not detected in CX-HILIC of chicken liver H1 (see text). Liver H1 peaks C, D, E and K described in Table 6.1 have been omitted because putative matches were not contained in the limited set of pooled erythrocyte H1 peaks selected for analysis by ISMS.

<sup>2</sup> H1 subtype based on AU gel mobility according to Shannon et al., (1987)

<sup>3</sup> M<sub>r</sub> (α-N-acetylated form) predicted from the corresponding gene sequences (Coles et al., 1987). The M<sub>r</sub> predicted for non-α-N-acetylated forms are shown in brackets.

<sup>4</sup> ± std. dev. Listed in order of apparent abundance in MS. The M<sub>r</sub> of relatively minor forms are shown in square brackets.

<sup>5</sup> (measured M<sub>r</sub>) - (predicted M<sub>r</sub>, α-N-acetylated form), integer value

<sup>6</sup> alignment of liver peak B with pooled erythrocyte peak 4 is somewhat arbitrary since pooled erythrocyte peaks 2 and 3 were not analysed by ion spray MS. The major form detected by electrospray MS (Table 4.2) in these peaks (21300) is a better match for the minor form of liver peak B (21298) than is the 21271 species detected by ion spray MS in erythrocyte peak 4. Thus, it is possible that liver peak B should be aligned with peaks 2 or 3 rather than peak 4.
detected in the pooled erythrocyte sample due to low relative abundance among the >100 birds represented in this sample.

6.2.5 Evidence for allelic polymorphism in chicken erythrocyte \( H1 \)

The MS data in Tables 4.2 and 6.1 suggest that limited proteolysis (e.g. loss of residues #1-4) was responsible for the heterogeneity detected within most of the CX-HILIC peaks of liver \( H1 \) and some peaks of pooled erythrocyte \( H1 \). However, since the difference in size between the major species in the multiple peaks resolved by CX-HILIC for \( H1a \) (peaks 7,9,10 and 12), \( H1b \) (peaks 1-4), and \( H1c \) (peaks 5,6 and 8) of pooled erythrocyte \( H1 \) (Table 4.2), and also between the multiple peaks for \( H1a \) (peaks E and I), \( H1b \), (peaks A-D) and \( H1c' \) (peaks J and K) of liver \( H1 \) (Table 6.1), differed by less than the mass of a single residue in virtually all instances, it is unlikely that proteolysis was responsible for these multiple forms. The same logic applies to the major forms of pooled erythrocyte \( H5 \) resolved by CX-HILIC (Table 5.1) and in this case, no role for proteolytic processing is suggested since the \( M_r \) of all of the major forms closely matched or exceeded the predicted value and the differences in mass between the major forms are less than one amino acid residue in almost all cases (Table 5.1). Because the \( H1 \) used to develop the CX-HILIC method was prepared from blood pooled from many (>100) chickens, I examined the possibility that allomorphic forms of \( H1 \) were present in this sample as described below.

Purified \( H1 \) was prepared from crude \( H1+H5 \) from erythrocyte nuclei of a single chicken by Bio-Gel P100 chromatography as described in Methods section 2.11. This single chicken blood \( H1 \) was then compared to pooled blood \( H1_{\text{total}} \) using CX-HILIC. Fig. 6.4 shows a typical CX-HILIC comparison of the two samples on an analytical column of 100 nm diam. pore PolyCAT A. Comparisons were also performed on a column packed with 30 nm diam. pore PolyCAT A (data not shown). The peak assignments shown for the single bird sample were confirmed by AU-PAGE (see below) and rechromatography experiments (individual variants recovered from the single bird sample were co-chromatographed after mixing with aliquots of the pooled sample). These comparisons demonstrated that allelic polymorphism was responsible for the multiple peaks for \( H1a \), \( H1b \) and \( H1c \) in the pooled sample. Specific allomorphs of \( H1a \)
Figure 6.4. Comparison of the CX-HILIC profiles of erythrocyte H1 from pooled blood and from a single chicken.

Approximately 100 μg of erythrocyte H1 from a single chicken (A), and 100 μg of pooled blood H1_total (B), were chromatographed on the same PolyCAT A analytical column (4.6 x 200 mm, 100 nm diam. pores) using a 4 hour linear gradient from 380 mM to 590 mM NaClO₄ in 10 mM propionic acid-NaOH pH 6.5 containing 70% CH₃CN and a flow rate of 0.8 ml/min. Only the central portion of the chromatograms, aligned according to peak identification by PAGE and rechromatography of individual peaks following mixing with the heterologous sample, are shown.
Figure 6.5. Identification of CX-HILIC peaks for single chicken erythrocyte H1.

(A) Approx. 2 mg of erythrocyte H1 purified from blood collected from a single chicken by Bio-Gel P100 chromatography was separated on a semi-preparative column of PolyCAT A (9.4 x 200 mm, 100 nm diam. pores) using a four hour linear gradient from 410 mM to 620 mM NaClO₄ in 70% CH₃CN, 10 mM propionic acid-NaOH pH 6.5 and a flow rate of 2.5 ml/min. Peaks are numbered to correspond to the equivalent peaks of pooled blood H₁ total as determined by PAGE (see B and C) chromatographic analyses (Fig. 6.4) and MS analyses (see text and Tables 6.3 and 6.4). Only the portion of the chromatogram in which H₁ proteins eluted is shown.

(B) SDS gel analysis of the peaks shown in (A) after rechromatography to reduce cross-contamination between variants. Approx. 500 ng of protein was loaded for each peak. The duplicate lanes for peaks 4 and 14 represent portions recovered separately. S = 500 ng starting material (single chicken Bio-Gel P100 purified H1), M = molecular weight markers. Proteins detected by silver staining. The relative positions of the six H1 variants based on AU gel mobility (determined in (C)) and approximate molecular weight (x10³ daltons) are shown on the left and right sides, respectively.

(C) Aliquots representing approx. 500 ng of the recovered peaks were analysed on a 50 cm long AU gel. S = 500 ng starting material (single chicken Bio-Gel P100 purified H1). The duplicate lanes for peak 14 represent portions recovered separately. Proteins detected by silver staining. The six H1 variants based on AU gel mobility in the nomenclature of Shannon and Wells (1987) are shown on the right side.
(pooled blood peaks 9 and 12), H1b (pooled blood peaks 1, 2 and 3) and H1c (pooled blood peak 5) were not detected in the blood of this chicken (Fig. 6.4). Excluding late-eluting minor species that were not characterized, the H1 set of this bird consisted of two forms of H1a (corresponding to peaks 7 and 10 of the pooled sample), one form of H1a' (peak 11), two forms of H1c (peaks 6 and 8), one form of H1c' (peak 13) and one form of H1d (peak 14). Although the majority of erythrocyte H1b in this chicken displayed CX-HILIC retention similar to peak 4 of the pooled sample, a second form labeled 4' in Fig. 6.4 was present that had not been detected in the pooled sample. The 4' form of H1b co-eluted with pooled erythrocyte peak 5 on 30 nm diam. pore PolyCAT A (data not shown) but co-eluted with peak 6 on 100 nm diam. pore PolyCAT A (Fig. 6.4).

6.2.6 PAGE identification of single chicken H1 CX-HILIC peaks

Larger amounts of erythrocyte H1 variants from the single bird sample shown in Fig. 6.4 were prepared in two separations of 2 mg of H1 on a semi-preparative column packed with 30 nm pore diam. Poly CAT A as shown in Fig. 6.5A. Pooled fractions from this first chromatographic step were further purified by rechromatography on the same column or a column packed with 100 nm diam. pore Poly CAT A. Eight fractions, which appeared to be homogenous or containing residual cross-contamination from adjacent peaks in SDS and AU gels were recovered, as shown in Fig. 6.5B and C. The numerical peak identities shown for the single chicken sample in Figs. 6.4 and 6.5A were assigned based on the AU gel shown in Fig. 6.5C and experiments in which aliquots of the recovered single chicken blood H1 peaks were co-chromatographed with samples of pooled blood H1 on analytical PolyCAT A columns.

6.2.7 Electrospray MS characterization of erythrocyte H1 CX-HILIC peaks from a single chicken.

As was found for the CX-HILIC peaks of H1 from pooled erythrocytes (Table 4.2) and liver (Table 6.1), most of the species detected in the CX-HILIC peaks of erythrocyte H1 from the single chicken were smaller than predicted by the corresponding gene sequences (Table 6.3). The differences between the predicted and observed sizes were remarkably similar to those found for
the peaks from pooled erythrocyte H1. The molecular masses determined for the single bird fraction are shown aligned to those of the pooled erythrocyte fractions in Table 6.4. With the exception of the novel 4’ component, the sizes determined for the single bird H1 peaks are well matched to those of the pooled erythrocyte H1 peaks they are aligned to on the basis of CX-HILIC retention and AU gel mobility in Table 6.4. As had been found for the pooled sample, the forms of H1a’ in peak 11 were the only H1 proteins recovered from the single bird sample whose M, substantially exceeded that predicted by the corresponding gene sequence. The masses observed within peak 11 from the single bird are similar, but not identical, to those observed in the pooled sample (the M, 21356 species observed in the single bird sample is likely to be due to contamination from peak 10). Given the similarities in M, values, the amino acid sequences peak 11 proteins in the single bird probably resemble that of H1b rather than H1a’, as was shown for peak 11 from pooled erythrocytes (Fig. 4.6), but this was not confirmed experimentally. The finding that the molecular masses of the single bird erythrocyte H1 proteins represent a subset of those in pooled erythrocyte H1 strongly supports the chromatographic evidence (Fig. 6.4) that the greater CX-HILIC chromatographic heterogeneity displayed by the pooled sample is due to the presence of a greater number of allelic forms of H1.
Table 6.3.  Electrospray MS of erythrocyte H1 CX-HILIC fractions isolated from a single chicken.

<table>
<thead>
<tr>
<th>HPLC peak</th>
<th>gel band</th>
<th>predicted length</th>
<th>predicted $M_r$</th>
<th>measured $M_r$</th>
<th>difference $^5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>b</td>
<td>223</td>
<td>22437.3 (22395.3)</td>
<td>21269 ± 11</td>
<td>-1168</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20634 ± 8</td>
<td>-1803</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20947 ± 7</td>
<td>-1490</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21694 ± 3</td>
<td>-743</td>
</tr>
<tr>
<td>4'</td>
<td>b</td>
<td>223</td>
<td>22437.3 (22395.3)</td>
<td>21479 ± 7</td>
<td>-958</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21693 ± 6</td>
<td>-744</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21921 ± 5</td>
<td>-516</td>
</tr>
<tr>
<td>7</td>
<td>a</td>
<td>224</td>
<td>22439.2 (22397.2)</td>
<td>22017 ± 8</td>
<td>-422</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21666 ± 12</td>
<td>-773</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22447 ± 6</td>
<td>+8</td>
</tr>
<tr>
<td>8</td>
<td>c</td>
<td>219</td>
<td>21843.4 (21801.4)</td>
<td>20955 ± 9</td>
<td>-888</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21267 ± 8</td>
<td>-576</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21646 ± 7</td>
<td>-197</td>
</tr>
<tr>
<td>10</td>
<td>a</td>
<td>224</td>
<td>22439.2 (22397.2)</td>
<td>22032 ± 6</td>
<td>-407</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21349 ± 5</td>
<td>-1090</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22467 ± 5</td>
<td>+28</td>
</tr>
<tr>
<td>11</td>
<td>a'</td>
<td>218</td>
<td>21714.3 (21672.3)</td>
<td>21356 ± 8</td>
<td>-358</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22037 ± 12</td>
<td>+323</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22487 ± 10</td>
<td>+773</td>
</tr>
<tr>
<td>13</td>
<td>c'</td>
<td>218</td>
<td>21955.5 (21913.5)</td>
<td>21514 ± 7</td>
<td>-441</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21199 ± 8</td>
<td>-756</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21959 ± 8</td>
<td>+3</td>
</tr>
<tr>
<td>14</td>
<td>d</td>
<td>217</td>
<td>21850.4 (21808.4)</td>
<td>21391 ± 9</td>
<td>-459</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21051 ± 7</td>
<td>-799</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21831 ± 3</td>
<td>-19</td>
</tr>
</tbody>
</table>

$^1$ peaks are labeled according to elution order as in Fig. 4.1. The peak designated 4' coeluted with peaks 5 and 6 of the pooled sample on 30 nm and 100 nm diam. Pore PolyCAT A media, respectively. This component was not apparent in chromatograms of pooled blood H1.

$^2$ H1 subtype based on AU gel mobility according to Shannon et al., (1987)

$^3$ length (in residues) and $M_r$ of the $\alpha$-N-acetylated form predicted from the corresponding gene sequences (Coles et al., 1987). The $M_r$ predicted for non-$\alpha$-N-acetylated forms are shown in brackets.

$^4$ ± std. dev. Listed in order of apparent abundance in MS. The $M_r$ of relatively minor forms are shown in square brackets.

$^5$ (measured $M_r$) - (predicted $M_r$, $\alpha$-N-acetylated form), integer value.
Table 6.4. Comparison of the masses of erythrocyte H1 CX-HILIC fractions from a single chicken with those from pooled blood.

<table>
<thead>
<tr>
<th>HPLC peak(^1)</th>
<th>AU gel subtype(^2)</th>
<th>predicted (M_r)^(^3)</th>
<th>measured (M_r)^(^4) pooled</th>
<th>(\Delta^5)</th>
<th>measured (M_r)^(^4) single</th>
<th>(\Delta^5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>b</td>
<td>22437.3 (22395.3)</td>
<td>21318 ± 6</td>
<td>-1119</td>
<td>20947 ± 7</td>
<td>-1490</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>21004 ± 8</td>
<td>-1433</td>
<td>21269 ± 11</td>
<td>-1168</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[20693 ± 8]</td>
<td>-1744</td>
<td>20634 ± 8</td>
<td>-1803</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[21740 ± 10]</td>
<td>-697</td>
<td>21694 ± 3</td>
<td>-743</td>
</tr>
<tr>
<td>2</td>
<td>b</td>
<td>22437.3 (22395.3)</td>
<td>21300 ± 13</td>
<td>-1137</td>
<td>21479 ± 7</td>
<td>-958</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20978 ± 7</td>
<td>-1459</td>
<td>[21693 ± 6]</td>
<td>-744</td>
</tr>
<tr>
<td>3</td>
<td>b</td>
<td>22437.3 (22395.3)</td>
<td>21300 ± 15</td>
<td>-1137</td>
<td>20956 ± 12</td>
<td>-1481</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20975 ± 7</td>
<td>-1462</td>
<td>21263 ± 8</td>
<td>-1174</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20634 ± 8</td>
<td>-1803</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21694 ± 3</td>
<td>-743</td>
</tr>
<tr>
<td>4</td>
<td>b</td>
<td>22437.3 (22395.3)</td>
<td>20956 ± 12</td>
<td>-1481</td>
<td>20947 ± 7</td>
<td>-1490</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>21263 ± 8</td>
<td>-1174</td>
<td>21269 ± 11</td>
<td>-1168</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20634 ± 8</td>
<td>-1803</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21694 ± 3</td>
<td>-743</td>
</tr>
<tr>
<td>4'</td>
<td>b</td>
<td>22437.3 (22395.3)</td>
<td></td>
<td></td>
<td>21479 ± 7</td>
<td>-958</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[21693 ± 6]</td>
<td>-744</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[21921 ± 5]</td>
<td>-516</td>
</tr>
<tr>
<td>5</td>
<td>c</td>
<td>21843.4 (21801.4)</td>
<td>20991 ± 6</td>
<td>-852</td>
<td>22017 ± 9</td>
<td>-422</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[22447 ± 6]</td>
<td>+8</td>
</tr>
<tr>
<td>6</td>
<td>c</td>
<td>21843.4 (21801.4)</td>
<td>20979 ± 9</td>
<td>-864</td>
<td>21666 ± 11</td>
<td>-773</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[22447 ± 6]</td>
<td>+8</td>
</tr>
<tr>
<td>7</td>
<td>a</td>
<td>22439.2 (22397.2)</td>
<td>22017 ± 9</td>
<td>-422</td>
<td>22017 ± 8</td>
<td>-422</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>21666 ± 11</td>
<td>-773</td>
<td>[21666 ± 12]</td>
<td>-773</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[22447 ± 6]</td>
<td>+8</td>
</tr>
<tr>
<td>8</td>
<td>c</td>
<td>21843.4 (21801.4)</td>
<td>20954 ± 8</td>
<td>-889</td>
<td>20955 ± 9</td>
<td>-888</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[20637 ± 8]</td>
<td>-1206</td>
<td>21267 ± 8</td>
<td>-576</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[21270 ± 11]</td>
<td>-573</td>
<td>21646 ± 7</td>
<td>-197</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[21644 ± 6]</td>
<td>-199</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>a</td>
<td>22439.2 (22397.2)</td>
<td>22064 ± 8</td>
<td>-375</td>
<td>21977 ± 4</td>
<td>-462</td>
</tr>
</tbody>
</table>
Peaks are labeled according to the elution order of the pooled sample as in Fig. 4.1. Entries for the single chicken sample are aligned with the pooled sample on the basis of co-chromatography. M<sub>r</sub> matching gives the same alignment. Peak 4' of the single chicken sample coeluted with peaks 5 and 6 of the pooled sample on 30 nm and 100 nm pore diam. PolyCAT A, respectively. This component was not apparent in chromatograms of pooled blood H1.

H1 subtype based on AU gel mobility according to Shannon et al., (1987)

M<sub>r</sub> for the α-N-acetylated form predicted from the corresponding gene sequences (Coles et al., 1987). The M<sub>r</sub> predicted for non-α-N-acetylated forms are shown in brackets

± std. dev. Values for the pooled blood samples are listed in order of apparent abundance in MS. Masses from the chicken samples are listed according to best match with the pooled sample. Relatively minor forms are shown in square brackets. All measurements were made on a Vestec electrospray single quadrupole mass spectrometer.

(measured M<sub>r</sub>) - (predicted M<sub>r</sub>, α-N-acetylated form), integer value

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>22439.2</td>
<td>21348 ± 7</td>
<td>-1091</td>
<td>21349 ± 5</td>
<td>-1090</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(22397.2)</td>
<td>[22033 ± 6]</td>
<td>-406</td>
<td>22032 ± 6</td>
<td>-407</td>
</tr>
<tr>
<td></td>
<td>a'</td>
<td>21714.3</td>
<td>22015 ± 17</td>
<td>+301</td>
<td>22037 ± 12</td>
<td>+323</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(21672.3)</td>
<td>22400 ± 10</td>
<td>+686</td>
<td>[22487 ± 10]</td>
<td>+773</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[22458 ± 5]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>22439.2</td>
<td>21321 ± 6</td>
<td>-1118</td>
<td>21349 ± 5</td>
<td>-1192</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(22397.2)</td>
<td>[22009 ± 11]</td>
<td>-430</td>
<td>22032 ± 6</td>
<td>-431</td>
</tr>
<tr>
<td></td>
<td>c'</td>
<td>21955.5</td>
<td>21519 ± 6</td>
<td>-436</td>
<td>21514 ± 7</td>
<td>-441</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(21913.5)</td>
<td>21198 ± 15</td>
<td>-757</td>
<td>21199 ± 8</td>
<td>-758</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[21951 ± 7]</td>
<td>-4</td>
<td>[21959 ± 8]</td>
<td>+3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[22049 ± 3]</td>
<td>+93</td>
<td>[22052 ± 3]</td>
<td>+94</td>
</tr>
<tr>
<td></td>
<td>d</td>
<td>21850.4</td>
<td>21383 ± 10</td>
<td>-467</td>
<td>21391 ± 9</td>
<td>-459</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(21808.4)</td>
<td>[21042 ± 4]</td>
<td>-808</td>
<td>21051 ± 7</td>
<td>-799</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[21819 ± 7]</td>
<td>-31</td>
<td>21831 ± 3</td>
<td>-19</td>
</tr>
</tbody>
</table>
Chapter 7. Evidence that the gene for H1a' has been identified incorrectly
7.1 Introduction

Taken together, the data presented in Chapters 4, 5 and 6 suggest that allelic variation, rather than post-translational modification or proteolysis, is responsible for much of the CX-HILIC heterogeneity observed for Hl_{total} and that the sequences of the alleles of the H\textsubscript{1} variant genes described by Coles et al., (1987) are somewhat larger than the majority of alleles expressed in chickens. This raised the question of what were the nature of the differences between the variants resolved by CX-HILIC and the reported sequences. Furthermore, as shown in Tables 4.2 and 4.3, MS analyses indicated that the pooled blood variant in CX-HILIC peak 11 represented an exception to this trend as the M, values determined for this protein exceeded the value for H1a', suggesting the possibility that the identification of this, and possibly other variants according to AU gel mobility was in error. To investigate these issues, CX-HILIC peak 11, which migrated like H1a' on AU gels, together with the flanking peaks 10 and 12 which migrated like H1a on AU gels, but whose M, was smaller than predicted for H1a (Fig. 4.1C and Table 4.2), were selected for further study. Two types of analyses were performed. Comparisons of the amino acid sequences derived from the genes of the six chicken H1 variants (Fig. 4.2) revealed that all six subtypes are predicted to lack tryptophan and possess a single tyrosine residue at equivalent positions in their sequences. MS analyses of the peptides released by cleavage at this tyrosine by N-bromosuccinimide (NBS) were used to determine whether the differences between these CX-HILIC variants and the reported sequences were localized to one or both of these fragments. Additionally, a phenylalanine residue is predicted to occur at an equivalent position in all six variants (residue 109 in H1a, Fig. 4.2). Immediately adjacent to this residue, the predicted sequences of the six variants diverge such that amino acid sequencing of the C-terminal fragments could be used to confirm AU gel based variant identifications. This strategy was used to confirm the identification of peaks 10 and 11. Whereas the sequence of the peak 10 C-terminal peptide matched that of H1a, as expected, the sequence of the peak 11 C-terminal peptide was identical to that of H1b rather than H1a' as expected. Together with MS analyses of the fragments recovered from NBS cleavages of peaks 10, 11 and 12, the results confirm that peaks 10 and 12 are most closely related to H1a while peak 11 is most closely related to H1b.
7.2 Results and Discussion

7.2.1 Characterization of NBS peptides from Hl_total CX-HILIC peaks 10, 11 and 12

Aliquots of CX-HILIC peaks 10, 11 and 12 were incubated with NBS to cleave the proteins at the tyrosine residue (residues 75 and 72 in H1α and H1α', respectively) as described in Methods, section 2.8. The proposed reaction mechanism for polypeptide cleavage by NBS, including conversion of tyrosine to a dibromodienone spirolactone is shown in Fig. 7.1. The characteristic absorbance of this moiety permitted direct identification of N-terminal fragments from cleavage reactions in HPLC chromatograms using absorbance detection at 260 nm as shown in Fig. 7.2.

N and C-terminal fragments recovered from RP-HPLC eluents were characterized by ion spray MS. The results of these analyses are summarized in Table 7.1 where the Mᵋ observed for the various fragments are aligned with the values predicted for the corresponding fragments of H1α and H1α'. In addition to assuming complete α-N-acetylation, the predicted Mᵋ values for N-terminal fragments were augmented by 172.8 mass units to reflect the conversion of tyrosine as shown in Fig. 7.1. Identical N-terminal fractions were recovered following cleavage of peaks 10 and 12. Each contained a Mᵋ 7329 peptide and a less abundant Mᵋ 6899 peptide. Neither peptide matched the Mᵋ 7304 predicted for the N-terminal fragment of H1α. However, assuming that the structure proposed for the converted tyrosine product is correct, the fact that the Mᵋ 7329 peptide exceeds this by only 25 units, less than the mass of a single amino acid residue, suggests that the sequence of this peptide could be identical to that predicted for H1α but that it is modified at one or more sites to give rise to the 25 unit difference (e.g. lysine methylation or hydroxylation would augment the observed Mᵋ by 14 or 16 units, respectively). Alternatively, this peptide could differ in sequence from that predicted for residues 1-75 of H1α but any differences are likely to be extremely limited since the similar Mᵋ values are consistent with only one or two amino acid substitutions.
Figure 7.1. Cleavage of polypeptides at tyrosine by NBS.

Mechanism of polypeptide cleavage by NBS at tyrosine residues as proposed by Fontana and Gross, (1986). Bromination of the tyrosyl side-chain and subsequent hydrolysis of the C-terminal peptide bond converts the tyrosine to a dibromodienone spirolactone (IV) with characteristic absorbance that permits specific detection of modified fragments at 260 nm as shown in Fig. 7.2. (Figure reproduced from Fontana and Gross, 1986, with permission)
Figure 7.2. RP-HPLC peptide maps of CX-HILIC peaks 10 and 11 after NBS cleavage.

Aliquots representing approx. 20 µg of pooled blood H1 total CX-HILIC peaks 10 (H1a) and 11 (H1a') were cleaved with NBS as described in Methods. The reaction mixture was separated on an analytical C18 column as described in Methods and peptides corresponding to the N- and C-terminal fragments of peak 10 (A) and peak 11 (B) were collected and recovered by vacuum drying. Peptide elution was monitored at 214 nm (solid line) and 260 nm (dotted line). Note the characteristic absorbancy of the converted tyrosine residue of the N-terminal fragment at 260 nm. The peak labeled H1* in (A) and (B) is presumed to represent H1 which reacted with NBS but was not cleaved.
Table 7.1.  Ion spray MS of N-bromosuccinimide fragments of pooled chicken erythrocyte H1 CX-HILIC peaks 10, 11 and 12.

<table>
<thead>
<tr>
<th>Fragment (residues)</th>
<th>AU gel subtype</th>
<th>predicted $M_r$</th>
<th>measured $M_r$</th>
<th>difference $^5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10N (1-75)</td>
<td>a</td>
<td>7304.2</td>
<td>7329.3 ± 0.5</td>
<td>+25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[6898.6 ± 1.8]</td>
<td>-406</td>
</tr>
<tr>
<td>10C (76-224)</td>
<td></td>
<td>15283.8</td>
<td>15312.0 ± 1.8</td>
<td>+28</td>
</tr>
<tr>
<td>11N (1-72)</td>
<td>a'</td>
<td>6990.9</td>
<td>7272.4 ± 0.6</td>
<td>+281</td>
</tr>
<tr>
<td>11C (73-218)</td>
<td></td>
<td>14872.3</td>
<td>15298.4 ± 1.6</td>
<td>+426</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[15397.9 ± 4.0]</td>
<td>+526</td>
</tr>
<tr>
<td>12N (1-75)</td>
<td>a</td>
<td>7304.2</td>
<td>7329.6 ± 0.9</td>
<td>+25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[6898.8 ± 2.8]</td>
<td>-406</td>
</tr>
<tr>
<td>12C (76-224)</td>
<td></td>
<td>15283.8</td>
<td>15283.9 ± 2.5</td>
<td>0</td>
</tr>
</tbody>
</table>

$^1$ protein recovered from CX-HILIC peaks 10, 11 and 12 of pooled erythrocyte H1 (Fig. 4.1) were cleaved with NBS as described in methods.

$^2$ H1 subtype based on AU gel mobility according to Shannon et al., (1987)

$^3$ the $M_r$ of the $\alpha$-N-acetylated protein predicted from the corresponding gene sequence (Coles et al., 1987) was used to estimate the $M_r$ of the N and C terminal fragments released following cleavage after the unique tyrosine residue of each sequence. The $M_r$ predicted for each N-terminal fragment (7131.4 for peaks 10 and 12 and 6818.1 for peak 11) has been augmented by 172.8 to reflect the conversion of tyrosine to a dibromodienone spirolactone according to the mechanism proposed for NBS cleavage (Fontana and Gross, 1986)

$^4$ ± std. dev. Listed in order of apparent abundance in MS. The $M_r$ of relatively minor forms are shown in square brackets.

$^5$ (measured $M_r$) - (predicted $M_r$, $\alpha$-N-acetylated form), integer value
Table 7.2. Putative identification of pooled erythrocyte H1 CX-HILIC peaks 10, 11 and 12 based on the M, of NBS fragments

<table>
<thead>
<tr>
<th>H1 subtype</th>
<th>N-terminal</th>
<th>C-terminal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>predicted</td>
<td>observed</td>
</tr>
<tr>
<td>a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>peak 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[6898.6]</td>
</tr>
<tr>
<td>peak 12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[6898.8]</td>
</tr>
<tr>
<td>a'</td>
<td>6990.9</td>
<td>14872.3</td>
</tr>
<tr>
<td>b</td>
<td>7289.2</td>
<td>15296.8</td>
</tr>
<tr>
<td>peak 11</td>
<td>7272.4</td>
<td>15298.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[15397.9]</td>
</tr>
<tr>
<td>c</td>
<td>6866.7</td>
<td>15125.5</td>
</tr>
<tr>
<td>c'</td>
<td>6965.8</td>
<td>15138.5</td>
</tr>
<tr>
<td>d</td>
<td>6993.9</td>
<td>15005.4</td>
</tr>
</tbody>
</table>

1 H1 protein recovered from CX-HILIC peaks 10, 11 and 12 of pooled erythrocyte H1 (Fig 4.1) were cleaved with NBS and peptides recovered by RP-HPLC as described in Methods.

2 H1 subtypes a, a', b, c, c' and d are listed according to AU gel mobility as described by Shannon et al., (1987). CX-HILIC peaks 10, 11 and 12 are listed immediately following the H1 subtype entry they best match according to M, of NBS fragments.

3 the M, of the α-N-acetylated protein predicted from the corresponding gene sequence (Coles et al., 1987) was used to estimate the M, of the N and C terminal fragments released following cleavage after the unique tyrosine residue of each sequence. The M, predicted for each N-terminal fragment has been augmented by 172.8 to reflect the conversion of tyrosine to a dibromodienone spiro lactone according to the mechanism proposed for NBS cleavage (Fontana and Gross, 1986). Measurements were made on a Sciex API III triple quadrupole MS. The M, of relatively minor forms are shown in square brackets.
The Mr 6899 peptide observed in the N-terminal fractions for peaks 10 and 12 is 431 mass units smaller than the Mr 7329 peptide. Examination of the predicted H1a sequence reveals that removal of the first four residues (SETA) would give a similar loss (430 units), suggesting that the presence or absence of these residues distinguishes the Mr 7329 and 6899 peptides, respectively. Note that this difference can be traced back to the intact proteins observed in peak 10 and in peak 12. Using the values for peak 10 in Table 4.3, the difference in size between the two most abundant ions detected was 22468.3 - 22036.5 = 431.8. Similarly, using the values for peak 12 in Table 4.2, the difference between the second most abundant ion detected and the predicted molecule was 22009 - 22439.2 = -430.2. These differences are close matches to the 430.4 unit change predicted upon the loss of the first four residues, SETA, of the predicted H1a sequence (Coles et al., 1987; see Fig. 4.2 and Table A1). Interestingly, loss of the same four N-terminal residues has been proposed to account for heterogeneity detected by time-of-flight MS within fractions of mouse liver H1 subtypes H1d and H1e prepared by RP-HPLC (Wang et al., 1997), suggesting that this may represent a specific proteolytic event with potential regulatory significance (see Discussion). Inspection of the “difference” column in Table 4.1 suggests that this putative “N-terminal clipping” may also be a characteristic of variants in H1 total peaks 9, 13 and 14.

In contrast, the C-terminal fractions from NBS cleavages of peaks 10 and 12 each appeared to be homogenous in ion spray MS but contained peptides that differed in Mr (Table 7.1). The Mr of the peak 12 C-terminal peptide matched that predicted for H1a, strongly suggesting that the sequences of these peptides are identical. Because only one type of peptide was observed, this supports the notion that most, if not all, of the proteins detected by electrospray MS of peak 12 (Table 4.1) were forms of H1a varying in sequence or post-translational modification in the N-terminal, but not the C-terminal, domain flanking the sole tyrosine residue. The same hypothesis can be proposed regarding different forms of H1a within peak 10. Because the N-terminal fragments observed for peaks 10 and 12 were identical, it appears that peaks 10 and 12 were resolved in CX-HILIC due to the difference(s) in their C-terminal domains. The Mr observed for the C-terminal peptide of peak 10 exceeded the Mr value predicted for H1a by only 28 mass units, suggesting that a limited number of small post-translational modifications or amino acid
substitutions found in peak 10 H1a, but not peak 12 H1a, contribute to their separation by CX-HILIC. Taken together, these data strongly suggest that both peaks 10 and 12 represent forms of H1a that differ from the predicted protein by only one or two amino acid substitutions or post-translational modifications.

All of the peptides recovered for the N and C-terminal fragments of peak 11 were larger than predicted for the corresponding fragments of H1a' (Table 7.1). The N-terminal fraction contained only an Mr 7272 peptide that is 281 units larger than expected while the C-terminal fraction contained an abundant Mr 15298 peptide and a less abundant Mr 15398 peptide that are 426 and 526 units larger than expected, respectively. In the absence of post-translational modifications, plausible explanations for these discrepancies are that peak 11 H1 is actually related to a subtype other than H1a' or that there are errors in the predicted H1a' sequence. As shown in Table 7.2, the Mr observed for the N and C peptides of peak 11 are better matched with those predicted for H1b than they are with those of H1a', despite the fact that the AU gel mobility of the H1 variants in peak 11 most closely resembles that of H1a' as described by Shannon et al. (1987) (Fig 4.1). The Mr of the abundant C-terminal peptide of peak 11 (15298) is an exact match with the Mr predicted for the corresponding peptide of H1b. The Mr of the less abundant C-terminal peptide (15398) is also best matched with that predicted for H1b, and may be a modified form of the Mr 15298 peptide. Similarly, the Mr observed for the N-terminal fragment of peak 11 more closely resembles that predicted for H1b than it does any other subtype, even though it is slightly smaller than that predicted for H1b (-18.8 units). Thus, analyses of N- and C-terminal fragments support the identification of peaks 10 and 12 as H1a subtypes but suggest, in contrast to AU gel characterization, that peak 11 is more likely to be related to H1b than it is to H1a'.

7.2.2 MS and microsequence analyses of α-chymotrypsin peptides of H1total CX-HILIC peaks 10 and 11.

To resolve the ambiguity regarding the identity of H1total CX-HILIC peak 11, an experimental approach based on microsequencing was devised. Comparison of the gene-derived
H1 variant sequences revealed that all six proteins are predicted to contain a single phenylalanine residue situated at the position equivalent to residue 109 in H1a (Fig. 4.2). If these predicted sequences are correct, then sequencing the twenty residues of any chicken H1 protein immediately C-terminal to this Phe residue can be used to establish its variant identity because a series of substitutions occur in this region in combinations that are unique to each of variant. The proteases pepsin and α-chymotrypsin cleave after phenylalanine preferentially under appropriate conditions and initial experiments led to the use of α-chymotrypsin as the preferred reagent using the conditions described in Methods, section 2.9. To prepare fragments for sequencing, aliquots of CX-HILIC peaks 10 and 11 were incubated with α-chymotrypsin and the C-terminal fragments isolated by RP-HPLC as described in Methods. Figure 7.3 shows RP-HPLC chromatograms of typical α-chymotrypsin digestions of H1 variants in CX-HILIC peaks 10 and 11. Ion spray MS (Table 7.3) and SDS gels (not shown) were used to confirm the identity of the fragments indicated in Fig. 7.3 and assess their purity prior to sequencing. The M₅s observed for the C-terminal fragment of peak 10 recovered from NBS cleavage of peak 10 (Table 7.1) and the fragment recovered from α-chymotrypsin digestion (Table 7.3) both exceeded the predicted value by approximately 30 mass units, indicating that C-terminal differences between the sequence predicted for H1a and the actual sequence of peak 10 H1 are localized to the segment spanning residues 110-224.

The α-chymotrypsin C-terminal fragments of peaks 10 and 11 were sequenced at the University of Michigan Protein and Carbohydrate Structure Facility in collaboration with Dr. P. C. Andrews (Director). As expected from AU gel mobility and ionspray MS analyses of the N and C-terminal fragments generated by NBS cleavage, microsequence analysis of the α-chymotrypsin C-terminal fragment from peak 10 (α10C) confirmed that peak 10 was related to H1a (Fig. 7.4). H1a differs from the other five variants by the presence of the unique combination of Arg, Ser.
Pro, Glu, Val, Lys, Pro, Lys, Ala and Ser at positions 110, 112, 115, 117, 118, 119, 123, 124, 126, 127 and 128, respectively (Fig. 7.4). To facilitate comparison, these positions are underlined in the sequences obtained for the peak 10 and peak 11 peptides in Fig. 7.4. Although most positions are identical among all six variants in this region of sequence, the identities of at least two residues could be used to unequivocally distinguish the sequenced portion of _10C from any of the predicted H1 subtype sequences.

Even though the AU gel mobility of peak 11 was more similar to that described for H1a' than it was to that of any other subtype (Shannon et al., 1987), the sequence of α11C was identical to that predicted for the corresponding segment of H1b (Fig. 7.4). Together with the finding that the M, of both the N and C-terminal NBS fragments of peak 11 were more similar to those predicted for H1b than they were to those of any other H1 subtype (Table 7.2), the data reveal that peak 11 is actually a form of H1b. This finding raises several questions including which, if any, of the CX-HILIC peaks contain proteins whose sequence matches the gene sequence assigned to H1a' (Shannon et al., 1987), and are there any other CX-HILIC peaks that have been misidentified based on AU gel data. These issues and possible explanations for the erroneous identification of peak 11 as H1a' are discussed in Chapter 8.
Aliquots representing approx. 40 ug of pooled blood CX-HILIC peaks 10 (H1a) and 11 (H1a') were digested with α-chymotrypsin as described in Methods. The digestion mixture was separated on an analytical C18 column as described in Methods and peptides corresponding to the N- and C-terminal fragments of peak 10 (A) and peak 11 (B) were collected and recovered by vacuum drying.
Table 7.3. Ion spray MS of the C-terminal fragments of pooled erythrocyte H1 CX-HILIC peaks 10 and 11 released by α-chymotrypsin

<table>
<thead>
<tr>
<th>H1 subtype²</th>
<th>C-terminal fragment³ predicted</th>
<th>observed</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a'/peak10</td>
<td>11780.8</td>
<td>11810.9±1.2</td>
<td>+ 30.1</td>
</tr>
<tr>
<td>(residues 110-224)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a'/peak 11</td>
<td>11369.2</td>
<td>11823.9±1.9</td>
<td>+ 454.7</td>
</tr>
<tr>
<td>(residues 107-218)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹ H1 protein recovered from CX-HILIC peaks 10 and 11 of pooled erythrocyte H1 (Fig. 4.1) was digested with α-chymotrypsin as described in Methods and the C-terminal fragments released upon cleavage after the unique phenylalanine were isolated by RP-HPLC as shown in Fig. 7.3 as described in Methods.

² CX-HILIC peaks 10 and 11 are designated as H1 subtypes a and a’, respectively, since they best match the AU gel mobility of these variants as described by Shannon et al., (1987).

³ the M₁ of the C-terminal fragments released from H1a and H1a’ following cleavage after the unique phenylalanine residue of each protein was predicted from the corresponding gene sequences (Coles et al., 1987). Measurements were made on a Sciex API III triple quadrupole MS.
Figure 7.4. Microsequence analyses of the α-chymotrypsin C-terminal fragments of CX-HILIC peaks 10 and 11.

Protein from CX-HILIC peaks 10 and 11 were incubated with α-chymotrypsin as described in Methods. The C-terminal fragment from each reaction (α10C and α11C) was isolated by RP-HPLC and microsequenced as described in Methods. The sequences of these peptides are shown aligned to the first 35 residues predicted from the gene sequences (Coles et al., 1987) for the corresponding C-terminal fragments released by α-chymotrypsin from variants H1a, a', b, c, c' and d according to the identification of the corresponding proteins described previously (Shannon and Wells, 1987). Sequences are aligned to obtain maximum homology of the intact proteins. Residues identical to those in H1a are indicated by *, gaps introduced to maximize homology are indicated by -. Numbers indicate the positions of the first and last residues within each predicted sequence. The positions of residues that vary among the six predicted protein sequences are underlined in the α10C and α11C peptide sequences.
Chapter 8. Discussion
8.1 Resolution of allelic and non-allelic variants of chicken H1 by CX-HILIC compared to other methods

Taken together, the chromatographic and electrophoretic data presented in chapters 4, 5 and 6 demonstrate that the CX-HILIC method described here resolves all six of the non-allelic variants of chicken H1 described previously (Shannon et al., 1987) and in addition, resolves allelic variants of at least three of these proteins that have not been described previously. Species-specific and tissue-specific heterogeneity of H1 attributable to non-allelic variation was first discovered using Bio-Rex 70 chromatography (Luck et al., 1958; Kinkade and Cole, 1966; Bustin and Cole 1968; Kinkade, 1969). Compared to CX-HILIC, Bio-Rex 70 chromatography is time-consuming and provides less resolution of chicken H1 variants. Only five components were partially resolved for H1 extracted from pooled chicken erythrocytes (i.e. similar to that employed here) in Bio-Rex 70 fractionations lasting 6 or 20 days (Fig. 1.10A and B; Kinkade, 1969; Dupressoir et al., 1984).

Furthermore, guanidine hydrochloride, the gradient-forming material typically employed in histone fractionations on Bio-Rex 70, absorbs light strongly at wavelengths less than 280 nm, limiting the sensitivity of detection by UV absorbance of proteins like histones which are deficient in aromatic residues. Adaptation of the Bio-Rex 70 methodology for use with a Mono S column permitted four components of chicken erythrocyte H1 to be resolved within 30 minutes (Fig 1.10C: Shannon and Wells, 1987). Even though optimal resolution of pooled erythrocyte H1 by CX-HILIC required gradients lasting 3-4 hours, the greater degree of resolution achieved outweighs the benefit of the rapidity of the method described by Shannon et al., (1987). In total, 14 major and several minor components were resolved during a four hour CX-HILIC fractionation of pooled erythrocyte H1. The buffers are compatible with UV detection at 214 nm and chromatograms with 14 major peaks can be obtained from as little as 20 µg of pooled total erythrocyte H1 on an analytical (4.6 mm i.d. x 200 mm) column. This sensitivity can be enhanced on smaller format columns (i.e. 2.1 mm i.d. x 100 mm) with only a slight loss in resolution (data not shown) and shorter separation times (1-2 hours, e.g Fig. 3.9) can be employed if maximal resolution is not required.
Comparison of published reports suggests that the efficacy of RP-HPLC for resolving non-allelic H1 variants may be species-dependent. Mouse (Wurtz, 1985; Lindner et al., 1988; Helliger et al., 1992; Lindner et al., 1992; Giancotti et al., 1993), rat (Kurokawa and MacLeod, 1985; Karhu, et al., 1988; Tchouatcha-Tchouassom et al., 1989; Lindner et al., 1990) and human H1 (Ohe et al., 1986) have been resolved into four or more components (excluding H1°) by different RP-HPLC methods. However, calf thymus H1 manifesting 6 components in Bio-Rex 70 chromatography (Cole, 1987; Smerdon and Isenberg, 1976), was resolved into only 2 (Lindner et al., 1986a, 1986b) or 3 (Certa and von Ehrenstein, 1981) components by RP-HPLC. Similarly, pooled chicken H1 eluted as a single peak in an optimized RP-HPLC method that resolved two forms of H5 (Helliger et al., 1988). As shown here, the resolution of chicken H1 and H5 by RP-HPLC can be significantly improved by the appropriate choice of column and chromatographic conditions (Fig. 3.2), but even the best RP-HPLC conditions did not resolve as many components of chicken H1 as did CX-HILIC. Comparison of the published predicted sequences of the non-allelic variants of mouse and chicken H1 (Figs. 1.12 and 4.2) reveals that greater numbers of amino acid substitutions distinguish the mouse variants from each other than is found for the chicken variants. Taken together, these observations suggest that RP-HPLC is unable to resolve H1 variants of some species, possibly those in which amino acid sequences are most highly conserved between variants. I have found that with appropriate adjustment of the acetonitrile and sodium perchlorate concentrations employed, CX-HILIC resolves more variants than RP-HPLC for human, mouse and rat H1 (data not shown).

8.2 Factors in the separation of H1 variants by CX-HILIC

The six gene-derived sequences of the non-allelic variants of chicken H1 reported by Coles et al., (1987) are shown aligned for maximum homology in Fig. 4.2. Even though the data presented here indicate that the sequences of the majority of the H1 proteins analysed here differ somewhat from these predictions, I assume that the general properties of the predicted proteins can be used to guide an intuitive discussion of the chromatographic mechanism. The sequences are relatively highly conserved and virtually all of the sequence variation occurs in the amino- and carboxyl-terminal domains. Non-conservative changes in ionizable residues are found at only 5
places in the entire sequence collection, corresponding to residues 10, 15, 119, 183 and 224 of the H1a sequence. Consequently, at the nearly neutral pH typically employed for cation-exchange chromatography of H1, these six proteins comprise only three values of predicted net charge, +58, +59, and +60 (Table 4.1). Three variants, H1a, H1c and H1c' (whose gene identification is not disputed), are predicted to differ slightly in size (224, 219 and 218 residues, respectively, Table 4.1), but possess identical net charge at pH 6.5 (+59, Table 4.1). Because all but two of the ionizable residues are found at equivalent positions in these sequences (Fig. 4.2), it is likely that factors other than electrostatic interactions contributed to the resolution of these non-allelic variants. Similar considerations apply for the separation of the allomorphs of H1a, H1b and H1c. Although the MS data (Table 4.2) are consistent with as much as a 6-7 residue difference in length between certain allomorphs (e.g. between the major forms of H1a in peaks 9 and 10, Mr 22064 and 21348, respectively), in most cases the differences in Mr between allomorphs are smaller and are consistent with the possibility that many allomorphs are identical lengths but differ by a limited number of amino acid substitutions. Moreover, the fact that the respective allomorphs possess identical AU gel mobilities (Fig. 4.1), suggests they have identical charge densities under these conditions and contain similar, if not identical, numbers of residues. These data suggest that the resolution of allelic and non-allelic variants of chicken H1 by CX-HILIC on PolyCAT A involves a complex, mixed-mode mechanism.

Evidence that interactions involving both ionic and nonionic polar groups such as lysine and serine, respectively, (i.e. electrostatic and hydrophilic interactions) contribute to the retention of solutes on PolyCAT A and related supports under the largely non-aqueous conditions employed here is discussed elsewhere (Alpert, 1990; Zhu et al., 1992; Jeno et al., 1993). As shown in Fig. 3.7, increasing levels of acetonitrile resulted in a biphasic curve for H1 retention. The reduction in retention between 0-40% (v/v) acetonitrile is probably not due solely to suppression of hydrophobic interactions, since less than 15% (v/v) acetonitrile was required to suppress such interactions between peptides and a similar stationary phase (Zhu et al., 1992). Increased α-helical structure may also have contributed to the decreased retention of chicken H1 at 40% (v/v)
acetonitrile but this was not confirmed experimentally. Chicken H1 possessed considerable α-helical content under CX-HILIC conditions (e.g. 70% CH₃CN + 0.6 M NaClO₄, Fig. 3.8), consistent with reports that trifluoroethanol (TFE) and NaClO₄ promote folding of the carboxyl terminal domain of H1 (Clark et al., 1988; Hill et al., 1989). Even though retention of such an ordered structure by electrostatic interactions is expected to decrease because some of the basic residues formerly accessible to the stationary phase would be oriented away from it, these effects appear to have been overshadowed by increases in hydrophilic interaction since H1 retention was greater in 70% (v/v) CH₃CN than in buffer containing less CH₃CN (Fig. 3.6). It has been proposed that the conformation of the C-terminal domain of H1 in 50-65% (v/v) TFE mimics that of H1 bound to DNA (Clark et al., 1988; Hill et al., 1989). This suggests the possibility that the resolution of H1 variants by CX-HILIC is attributable, in part, to features of the variants that may contribute to the differences in their interactions with DNA or chromatin that have been reported previously (reviewed in section 1.3.5).

8.3 Two of the six genes for the non-allelic variants of chicken H1 have been incorrectly identified

The nomenclature employed in the literature and used here for the non-allelic variants of chicken H1 is based on AU gel mobility as described by Shannon and Wells (1987). These authors separated five peaks for H1 prepared from a sample of pooled rooster blood by cation-exchange chromatography as shown in Fig. 1.10C. AU gels, which resolved six bands, designated H1a, H1a', H1b, H1c, H1c' and H1d, for the starting material, were used to identify the variants in each peak and revealed that peaks 1, 3, 4 and 5 contained H1a, H1c, H1d and H1c', respectively, in pure form but that peak 2 contained a mixture of H1a, H1a' and H1b (Shannon and Wells, 1987). Protein sequencing of peptides recovered from V8 protease digestions of peaks 1, 3, 4 and 5 was used to correlate the H1a, H1c, H1d and H1c' proteins, respectively, with the 11L, 0.10, 02 and 01 genes described by Coles et al., (1987). However, assignment of gene sequences to peak 2 was complicated by the fact that protease digestion liberated fragments from the three types of molecules, H1a, H1a' and H1b, present in this peak. Comparison of the RP-
HPLC profiles of V8 digests of peak 1 (containing pure H1a) and peak 2 was used to identify unique peptides that were assumed to be derived from H1b since H1a' was a relatively minor component of peak 2 (Shannon and Wells, 1987). Thus, when peptide sequences were obtained from peak 2 that matched the 03 gene, these were assumed to correspond to H1b, and by elimination, H1a' was identified as the product of the 11R gene (Shannon and Wells, 1987).

Several lines of evidence presented in this thesis demonstrate that the amino acid sequence of H1a' is not equivalent to the sequence predicted from the 11R gene. This protein was highly enriched in peak 11 of CX-HILIC purifications of pooled erythrocyte H1 and H1 from a single chicken although slight contamination by H1a was apparent in both cases (Figs. 4.1 and 6.5). Contamination by H1b was not apparent. In each case, MS analyses were consistent with the presence of two forms of the protein (M, 22015 and 22400 in Table 4.2) which were larger than the size predicted for the product of the 11R gene (M, 21714) by approximately three or six residues (Tables 4.2, 4.3 and 6.3). Given that the Ms determined for essentially all of the other CX-HILIC peaks were smaller than predicted by the genes assigned according to the published scheme (Shannon and Wells, 1987), the basis for this discrepancy was investigated. In section 7.2, it was shown directly and unequivocally that the sequence of the first 22 residues of the α1IC peptide recovered following cleavage of CX-HILIC peak 11 H1a' by α-chymotrypsin is identical to that predicted for the product of the 03 gene, identified by Shannon and Wells to correspond to H1b. Moreover, the molecular masses of both the N- and C-terminal fragments released upon cleavage by NBS resemble those predicted for the 03 gene much more than the fragments predicted for any other variant. Thus, I conclude that H1a' actually corresponds to the 03 gene rather than the 11R gene as assumed by Shannon and Wells, (1987). This raises the questions of which H1 variant actually corresponds to the 11R gene and are other H1 protein-gene assignments in error? These conundrums are discussed below.

Because the original analyses performed to align the H1a, H1c, H1d and H1c' proteins with the 11L, 0.10, 02 and 01 genes, respectively, utilized pure fractions of these proteins (peaks 181
1, 3, 4 and 5 of Shannon and Wells, 1987), these assignments are unlikely to be in error. Moreover, I have confirmed the correlation of H1a with the 11L gene directly by MS analyses of NBS fragments from CX-HILIC peaks 10 and 12 (Table 7.1) and also by amino acid sequencing the C-terminal α-chymotrypsin fragment from CX-HILIC peak 10 (Fig. 7.4). Together with the above, these data suggest that only the gene assignments for H1a’ and H1b described by Shannon and Wells (1987) are in error. Based on the lesser abundance of H1a’ in peak 2, they assumed that peptide sequences matching that of the 03 gene which were obtained by sequencing of RP-HPLC fractions recovered from the V8 digest of peak 2 derived from H1b. However, it is apparent from their published chromatograms that the rapid (20 min.) RP-HPLC peptide mapping procedure employed was not adequate to resolve all of the unique peptides generated from H1a, H1a’ and H1b when peak 2 was digested with V8 protease (9 are predicted, Fig. 4.2). Thus, I propose that the sequences they assumed to derived from H1b actually derived from H1a’ present in the digest. Based on the data presented here in Table 7.2 and Fig. 7.4, I conclude that H1a’ is the product of the 03 gene, and based on the arguments detailed above and by elimination, propose that H1b is the product of the 11R gene.

Ideally, direct confirmation that the H1b protein corresponds to the 11R gene by protein sequencing is required to finalize this amended identification but this was not performed due to time and financial constraints. However, considered in whole, the amended identification scheme fits the available data better than the scheme of Shannon and Wells (1987) as discussed below. Alternate explanations such as the possibility that the 11R gene may not correspond to any of the H1 variant proteins are unlikely since mRNAs corresponding to each of the six genes identified by Coles et al., (1987) have been detected using variant-specific antisense RNA probes in RNase protection assays (Takami et al., 2000). Furthermore, while the existence of more than six H1 genes in chickens remains a formal possibility, current data on H1 gene copy number suggests that this is unlikely (Ruiz-Carillo et al., 1983; D’Andrea et al., 1985; Coles et al., 1987; Nakayama et al., 1993; Takami and Nakayama, 1997; Takami et al., 2000).
Revising the identity of the genes corresponding to the H1a' and H1b proteins has a significant impact on the magnitude of the discrepancies between the Mₛ's measured for these proteins and the predicted values. Using Table 4.2 as an example of the impact of the old identification scheme, note that the Mₛ's determined for CX-HILIC peaks 1-4 (all forms of H1b) were generally smaller than the predicted value by 1130-1450 mass units and that the Mₛ's determined for CX-HILIC peak 11 (H1a') were 301 and 686 mass units larger than expected. Using the revised gene identifications to derive the predicted Mᵢ values (Table 8.1), note that the discrepancies for the forms of H1b now generally range from 410 to 730 mass units smaller than predicted and that the forms of H1a' are now 37 and 422 mass units smaller than predicted. The magnitude of these discrepancies is much more similar to that observed for the remaining variants H1a, H1c, H1c' and H1d, for whom the gene identities are not disputed (Table 4.2).

The identification scheme for all six non-allelic H1 variants originally proposed by Shannon and Wells (1987) is compared to the amended version in Table 8.2 where the proteins are listed according to observed AU gel mobility (slowest to fastest). According to the original scheme, the migration of the product of the 11R gene, predicted to contain 218 residues, was highly anomalous since the H1a' protein actually migrates more slowly than H1b, assumed to be the product of the 03 gene and predicted to be five residues larger (Table 8.2). Due to uncertainty regarding the relative ionization of aspartic and glutamic residues in these proteins under the conditions of AU gel electrophoresis, I have not calculated the net charge of the predicted H1 variants under these conditions. However, inspection of the predicted sequences (Figure 4.2) suggests that the net charge of the 11R and 03 products is similar under these conditions and thus the difference in charge density is not likely to be consistent with the observed order of migration of the proteins as identified by Shannon and Wells. In contrast, according to the revised identification scheme, the slower migration observed for the H1a' protein is consistent with the identification that this is the product of the 03 gene, predicted to be 5 residues larger than that of the 11R gene, now identified to correspond to the faster migrating H1b protein (Table 8.2).
Table 8.1. Comparisons of electrospray MS data for pooled chicken erythrocyte H1a' and H1b after revising the identity of the corresponding genes.

<table>
<thead>
<tr>
<th>HPLC peak#</th>
<th>AU gel band</th>
<th>gene</th>
<th>predicted length</th>
<th>predicted $M_r$</th>
<th>measured $M_r$</th>
<th>difference $^6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>b</td>
<td>11R</td>
<td>223</td>
<td>21714.3 (21672.3)</td>
<td>21318 ± 6</td>
<td>-396</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21004 ± 8</td>
<td>-710</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[20693 ± 8]</td>
<td>-1021</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[21740 ± 10]</td>
<td>+26</td>
</tr>
<tr>
<td>2</td>
<td>b</td>
<td>11R</td>
<td>223</td>
<td>21714.3 (21672.3)</td>
<td>21300 ± 13</td>
<td>-414</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20978 ± 7</td>
<td>-736</td>
</tr>
<tr>
<td>3</td>
<td>b</td>
<td>11R</td>
<td>223</td>
<td>21714.3 (21672.3)</td>
<td>21300 ± 15</td>
<td>-414</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20975 ± 7</td>
<td>-739</td>
</tr>
<tr>
<td>4</td>
<td>b</td>
<td>11R</td>
<td>223</td>
<td>21714.3 (21672.3)</td>
<td>20956 ± 12</td>
<td>-758</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21263 ± 8</td>
<td>-451</td>
</tr>
<tr>
<td>11</td>
<td>a'</td>
<td>03</td>
<td>218</td>
<td>22437.3 (22395.3)</td>
<td>22015 ± 17</td>
<td>-422</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22400 ± 10</td>
<td>-37</td>
</tr>
</tbody>
</table>

1 peaks are labeled according to elution order as in Fig. 4.1

2 H1 variant based on AU gel mobility (Shannon et al., 1987)

3 revised identity of the corresponding gene according to the data presented in this thesis

4 length (in residues) and $M_r$ of the $\alpha$-N-acetylated form predicted from the gene sequences (Coles et al., 1987). The $M_r$ predicted for non-$\alpha$-N-acetylated forms are shown in brackets

5 ± std. dev. Listed in order of apparent abundance in MS. The $M_r$ of relatively minor forms are shown in square brackets.

6 $(\text{measured } M_r) - (\text{predicted } M_r, \alpha$-N-acetylated form), integer value
Table 8.2. Existing and revised alignments of chicken H1 variant proteins with their respective genes

<table>
<thead>
<tr>
<th>Existing scheme&lt;sup&gt;1&lt;/sup&gt;</th>
<th>AU gel band&lt;sup&gt;2&lt;/sup&gt;</th>
<th>Revised scheme&lt;sup&gt;3&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>gene length</td>
<td>gene length</td>
<td></td>
</tr>
<tr>
<td>11L</td>
<td>224</td>
<td>11L</td>
</tr>
<tr>
<td>11R</td>
<td>218</td>
<td>03</td>
</tr>
<tr>
<td>03</td>
<td>223</td>
<td>11R</td>
</tr>
<tr>
<td>0.10</td>
<td>219</td>
<td>0.10</td>
</tr>
<tr>
<td>01</td>
<td>218</td>
<td>01</td>
</tr>
<tr>
<td>02</td>
<td>217</td>
<td>02</td>
</tr>
</tbody>
</table>

<sup>1</sup> Names of the genes described previously (Coles et al., 1987) and predicted protein lengths aligned to the AU gel mobility forms of H1 according to Shannon and Wells (1987)

<sup>2</sup> H1 variant protein designation according to AU gel mobility as described previously (Shannon and Wells, 1987) and used throughout this thesis

<sup>3</sup> Names of the genes described previously (Coles et al., 1987) and predicted protein lengths aligned to the AU gel mobility forms of H1 as described previously (Shannon and Wells, 1987) but revised as described in Chapter 7.3 of this thesis
8.4 Characteristics of the allelic and non-allelic variants of chicken H1 resolved by CX-HILIC

The marked difference in the number of peaks resolved by CX-HILIC for erythrocyte H1 from the single bird compared to the pooled sample (Fig. 6.4), strongly suggests that allelic variants of H1a, b and c were present in the pooled sample. Allelic variation in a histone was first demonstrated with the partial resolution on Bio-Rex 70 of two chicken H5 variants reported to differ by a single amino acid substitution (Gln to Arg) (Greenaway and Murray, 1971), and it has been suggested that polymorphisms encountered during the sequencing of Bio-Rex 70 fractions of calf and rabbit H1 were due to allelic variation (Cole, 1987). As reviewed in section 1.3.4, allelic variants of linker histones have been detected by electrophoretic techniques in a wide variety of species. In these reports, only allomorphs that differed enough in size and/or net charge to be resolved as distinct bands were detected. Thus, the allomorphs described here that are resolved only by CX-HILIC would not have been detected in electrophoretic surveys and appear to represent a class of allelic variants of H1 that have not been described previously. The number and apparent abundance of these allomorphs suggests that allelic variants of H1 which are not resolved by electrophoretic methods are common in chickens.

Comparison of the $M_s$ observed for chicken H1 CX-HILIC fractions prepared from pooled blood, liver, and blood from a single bird (Tables 4.2, 6.1 and 6.3), to the $M_s$ predicted by the respective genes (using the revised assignment of H1a' and H1b to the 03 and 11R genes, respectively, as in Table 8.1) reveals that almost all of the H1 proteins detected in the CX-HILIC fractions were smaller than predicted by the corresponding gene sequences. As discussed below, these discrepancies may originate due to factors affecting the accuracy of the gene sequences these comparisons are based on, proteolysis of H1 proteins, or both.

Complete sequences for chicken H1 proteins determined solely by amino acid sequencing have not been reported and the six H1 variant sequences used as "standards" here were derived from nucleic acid sequences. It is possible that the gene sequences reported contain errors attributable to artifacts encountered during cloning or DNA sequencing that led to errors in the
predicted protein sequences. Five of the six gene sequences were determined by one group of investigators and careful inspection of the published literature and molecular biology databases reveals that independent confirmation has not been reported for any of the six sequences. However, the high degree of identity predicted for the H1a, H1a', H1b, H1c and H1c' proteins, (corresponding to the H1.11L, H1.11R, H1.03, H1.10 and H1.01 genes reported by Coles et al., 1987), and the high degree of identity these sequences share with that predicted for the H1d protein (corresponding to the H1.02 gene reported by Sugarman et al., 1983), both suggest it is unlikely that many errors are present in any of the reported sequences.

All six chicken H1 genes lack introns (Coles et al., 1987; Nakayama et al., 1993) and thus, alternative mRNA splicing is unlikely to be the basis for the discrepancies in predicted and observed M_r's. Alternate 3' processing of transcripts from the 10 and 01 genes that results in the production of both polyadenylated and nonpolyadenylated mRNAs in chicken tissues and cultured cells has been described (Kirsh et al., 1989). However, the differences in these messages are limited to the 3' untranslated region and do not affect the protein sequences. Thus, it appears that allelic variation in H1 gene sequences is the only potential mechanism operating prior to protein translation that could possibly contribute to discrepancies between the predicted and observed size of the H1 proteins. Even though amino acid sequence differences attributable to allelic variation were not identified directly here, the simpler nature of the CX-HILIC profiles observed for erythrocyte H1 obtained from a single bird and the two pooled livers strongly suggest that allelic variation is the most likely explanation for much of the multiplicity of CX-HILIC peaks for the H1a, H1b and H1c proteins in the pooled blood sample. The fact that the respective presumed allomorphs of H1a, H1b and H1c possessed identical AU gel mobilities suggests that differences between allomorphs are extremely limited and are unlikely to involve non-conservative substitutions involving ionizable residues (or at least those whose effect on net charge is not compensated for by other changes). For example, the difference of approximately 18 mass units determined for the major forms of H1b in CX-HILIC peak 1 (M_r 21318) and peak 2 (M_r 21300, Table 4.2), and the difference of approximately 12 mass units between the major forms of H1c detected in peak 5 (M_r 20991) and peak 6 (M_r 20979), are both consistent with the existence of a
single conservative amino acid substitution between the respective allomorphs (e.g. L/I>M = +18, S>V = +12; a table compiling possible mass shifts due to amino acid substitution is available on the World Wide Web at http://prospector.ucsf.edu). Small differences like these distinguish the major forms observed by MS for pooled erythrocyte allomorphs of H1b (peaks 1, 2 and 3), H1c (peaks 5, 6 and 8) and H1a (peaks 10 and 12) (Table 4.2).

However, greater differences in M, were observed for other presumed pooled erythrocyte allomorphs, such as the approximately 362 mass unit difference between the major forms of H1b in peak 1 (M, 21318) and peak 4 (M, 20956) and the approximately 716 mass unit difference between the major forms of H1a in peak 9 (M, 22064) and peak 10 (M, 21348) (Table 4.2). These larger differences are more consistent with partial proteolysis than with limited amino acid substitution due to allelic variation. The chromatographic and electrophoretic data presented in Fig. 4.1 suggest that these differences are not artifacts arising from comparison of non-homologous variants contaminating these fractions. Moreover, differences of similar magnitude were observed between the forms detected within individual CX-HILIC peaks and also between the observed masses and those predicted from gene sequences (Table 4.2, see Table 8.1 for the appropriate values for CX-HILIC peaks 1-4 and 11). Thus, even though the majority of pooled erythrocyte CX-HILIC peaks were found to be electrophoretically homogenous (Fig. 4.1), the MS data strongly suggests that differences due to proteolytic processing were present in addition to those potentially due to allelic variation described above.

The clearest evidence for both allelic variation and in vivo proteolysis of chicken H1 proteins is provided by the M,s determined for the NBS fragments of pooled blood peaks 10 and 12. This data, shown previously in Table 7.1 is amended in Table 8.3 to reflect the revised assignment of H1a’ (peak 11) as the product of the H1.03 gene. As discussed in Chapter 7, it appears that the major form of H1a eluting in peak 10 was identical to that eluting in peak 12 except that an amino acid substitution (or possibly a modification) specific to peak 10 H1a made the C-terminal NBS fragment of this form 28 mass units larger (Table 8.3). However, the difference in M, between the two N-terminal fragments recovered for peak 10, and also for the two fragments
Table 8.3. Revised comparison of N-bromosuccinimide fragments of pooled chicken erythrocyte H1 CX-HILIC peaks 10, 11 and 12.

<table>
<thead>
<tr>
<th>Fragment (residues)</th>
<th>AU gel subtype</th>
<th>predicted $M_r$</th>
<th>measured $M_r$</th>
<th>difference $^5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10N (1-75)</td>
<td>a</td>
<td>7304.2</td>
<td>7329.3 ± 0.5</td>
<td>+25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[6898.6 ± 1.8]</td>
<td>-406</td>
</tr>
<tr>
<td>10C (76-224)</td>
<td></td>
<td>15283.8</td>
<td>15312.0 ± 1.8</td>
<td>+28</td>
</tr>
<tr>
<td>11N (1-72)</td>
<td>a'</td>
<td>7289.2</td>
<td>7272.4 ± 0.6</td>
<td>-17</td>
</tr>
<tr>
<td>11C (73-218)</td>
<td></td>
<td>15296.8</td>
<td>15298.4 ± 1.6</td>
<td>+2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[15397.9 ± 4.0]</td>
<td>+101</td>
</tr>
<tr>
<td>12N (1-75)</td>
<td>a</td>
<td>7304.2</td>
<td>7329.6 ± 0.9</td>
<td>+25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[6898.8 ± 2.8]</td>
<td>-406</td>
</tr>
<tr>
<td>12C (76-224)</td>
<td></td>
<td>15283.8</td>
<td>15283.9 ± 2.5</td>
<td>0</td>
</tr>
</tbody>
</table>

$^1$ Protein recovered from CX-HILIC peaks 10, 11 and 12 of pooled erythrocyte H1 (Fig. 4.1) were cleaved with NBS as described in methods.

$^2$ H1 subtype based on AU gel mobility according to Shannon et al., (1987)

$^3$ The sequences of the H1.11L and the H1.03 genes described by Coles et al., (1987) were used for the H1a and H1a' proteins according to the revised protein-gene assignment shown in Table 8.2. The $M_r$ of the α-N-acetylated protein was used to estimate the $M_r$ of the N and C terminal fragments released following cleavage after the unique tyrosine residue of each sequence. The $M_r$ predicted for each N-terminal fragment (7131.4 for peaks 10 and 12 and 7116.4 for peak 11) has been augmented by 172.8 to reflect the conversion of tyrosine to a dibromodienone spirolactone according to the mechanism proposed for NBS cleavage (Fontana and Gross, 1986)

$^4$ ± std. dev. Listed in order of apparent abundance in MS. The $M_r$ of relatively minor forms are shown in square brackets.

$^5$ (measured $M_r$) - (predicted $M_r$, α-N-acetylated form), integer value
recovered for peak 12, coincide precisely with the change in molecular mass predicted for the loss of residues 1-4 of H1a (7329.3 - 6898.6 = 430.7, predicted value = 430.4 in Table A1). This suggests that a form of H1a lacking residues 1-4 co-eluted with H1a containing residues 1-4 in both CX-HILIC peaks 10 and 12. Taken together the data suggest that in this particular case, the CX-HILIC procedure resolved allelic variants of H1a that differed by a single substitution in the C-terminal domains but did not resolve N-terminally clipped forms of these putative allelic variants of H1a. Note that several of the M,s determined for CX-HILIC fractions of H1 from liver (Table 6.1) and blood from a single chicken (Table 6.3) were smaller than expected by values approximating 430 mass units and are consistent with the loss of residues 1-4. Larger differences (e.g. the 888 mass unit difference for erythrocyte H1a in peak 8 of the single chicken, Table 6.3) suggest that the loss of up to 10 residues may occur in some instances. The fact that no evidence of proteolysis of H5 was obtained, and that a discrete number of truncated forms of H1 were observed suggests the existence of a mechanism involving regulated proteolysis.

Taken together, the data presented in this thesis demonstrate that chicken H1 is highly heterogeneous and suggest that both allelic variation and proteolytic processing in vivo contribute to this heterogeneity.

8.5 If allelic variation in chicken H1 is so extensive, why is it not better known?

Three factors have prevented the allelic variation described here from being detected previously. Two of these factors apply to work performed at the protein level. First, the high degree of resolution required to resolve the allelic variants described here has not been attained in previous chromatographic and electrophoretic analyses. The practical impact of this limitation is exemplified by the mis-identification of the gene for H1a′ as described above.

Secondly, it appears that this thesis is the first instance in which MS analyses have been performed on chicken linker histones. I am not aware of any reports in which MS characterization was performed on crude or purified chicken linker histones. At an early stage in this work, electrospray MS of pooled blood H1-total was attempted. Accurate molecular masses could not be
determined for any components in this extremely heterogeneous sample due to the limits of resolution of the spectrometer. However, it was apparent from the complexity of the spectra that the number of different forms of H1 greatly exceeded the six sequences expected (see below). Moreover, given the state of knowledge at that time, many workers would have ascribed this heterogeneity to post-translational modification, as I did until accurate molecular masses were obtained for individual CX-HILIC peaks. I anticipate that allelic variation will be increasingly recognized as a cause of heterogeneity in histones and other proteins, especially in those prepared from wild rather than inbred laboratory strains of organisms, as MS are used more routinely. Heterogeneity in mass spectra that is consistent with allelic variation has been reported for electrophoretically homogenous RP-HPLC fractions of H1 from inbred strains of mice (Giancotti et al., 1993) and RP-HPLC fractions of bovine H1 (Berger, 1995).

The third factor applies to work performed at the level of DNA. Apart from the single copy of the H5 gene (Ruiz-Carillo et al., 1983), it is estimated that there are only six H1 genes, each corresponding to a separate non-allelic variant, in the chicken genome (Ruiz-Carillo et al., 1983; D'Andrea et al., 1985; Coles et al., 1987; reviewed in Nakayama et al., 1983) based on Southern blotting analyses and genomic cosmid library screening. The fact that hybridization with a probe corresponding to one H1 variant sequence was used to detect all six variant genes in these studies demonstrates that hybridization-based methods such as Southern blotting can be insensitive to the degree of nucleotide sequence variation associated with allelic polymorphism. Thus, allelic variation has not been detected in earlier efforts utilizing these methods to identify and map histone genes in chickens and other species. However, as more information becomes available through genome sequencing projects and genome-based phylogenetic studies, a clearer picture of histone allelic polymorphism is beginning to form. Examples in chickens are not available at present, but data from other organisms illustrate the point that histone gene sequences show considerable polymorphism. A genome-based phylogenetic study of the copepod Tigriopus californicus revealed the existence of 22 different nucleotide sequences among wild populations for a segment of DNA including approx. 150 bases of coding sequence and 350 bases of the 5' flanking region (Burton and Lee, 1994). One change in the protein sequence was observed to occur in the short
stretch of H1 coding region examined and this change distinguished one major population from another. Similarly, genomic analyses revealed extensive polymorphisms in H2A genes of different strains of *Trypanosoma cruzi* (Thomas et al., 2000) and the H3 gene of different strains of *Fusarium subglutinans* (Steenkamp et al., 1999) that were used to develop PCR-RFLP assays to distinguish strains of these organisms. No changes in amino acid sequence were observed in these latter two cases and they serve to illustrate the point that extensive polymorphisms can accumulate in histone genes in the absence of amino acid changes. Similarly, 10 genes with different nucleotide sequences but encoding a single H4 protein sequence have been found in the human genome (Doenecke et al., 1997). Presumably, somewhat greater nucleotide sequence divergence has occurred in H1 genes of higher eukaryotes, and lower degrees of constraint on H1 function, relative to that of H4, have enabled allelic variants with amino acid sequence differences to be perpetuated. Because of limitations to resolution inherent to protein analyses, full appreciation of the extent of allelic variation in H1, or any protein, is likely to depend on systematic analyses of genomic polymorphisms.

8.6 If proteolytic processing of chicken H1 is so extensive, why is it not better known?

An unexpected feature of the MS data presented in this thesis that is difficult to rationalize is that multiple species, some of which appear to be missing 4 (or in some cases more) amino acids compared to their counterparts, were detected when fractions that appeared to be homogenous in AU and SDS gels were analysed. Although in some cases, this may be attributable to the detection of minor components in MS that were not detected by the gel staining procedures employed, this does not seem to be applicable to all cases. Moreover, this cannot be the explanation for those instances in which the $M_r$ of the major forms eluting in separate CX-HILIC peaks were found to differ by a large amount, and yet possessed identical AU gel mobilities, such as the earliest and the latest eluting forms of H1b from pooled erythrocytes (peaks 1 and 4 differ by approximately 362 mass units, Table 4.2), and the corresponding peaks from liver (peaks A and D differ by approximately 170 mass units, Table 6.1). This suggests that the electrophoretic procedures employed were not capable of resolving such differences. It is well known that histone proteins,
in general, migrate anomalously in SDS gels, with apparent \( M_s \)s that significantly exceed the true values (e.g. Hamana and Iwai, 1974), so it is difficult to estimate the limits of resolution by this technique. However, the inability of AU gels to resolve such large differences was unexpected since these gels are routinely used to resolve acetylated and phosphorylated forms of core histones that differ by single units of net charge (e.g. Cheung et al., 2000). However, the fact that lysine or arginine are not found within the first 16 positions of any of the predicted chicken H1 sequences (Fig. 4.2), suggests the possibility that AU gels may be relatively insensitive to N-terminal proteolysis since this does not result in the loss of ionized residues, assuming that the aspartic and glutamic acid residues found in this portion of the proteins remain protonated during AU gel electrophoresis. This possibility may extend to other species since lysine and arginine are not found in the first 15 positions of non-allelic variants of mouse H1 (Fig. 1.12). Note that this hypothesis is not applicable to the loss of residues from the lysine rich C-terminal portions of the molecules (Fig. 4.2). All 6 of the predicted chicken H1 sequences terminate in a lysine and it is anticipated that species resulting from the loss of this residue or multiple lysines upon more extensive C-terminal proteolysis, would be readily resolved from forms containing these residues by AU gel electrophoresis.

Taken together, the above suggests, in particular, that limited N-terminal proteolysis of H1 may not be detected by standard electrophoretic techniques, suggesting, in turn, the possibility that such proteolytic processing of H1 may be more widespread than is currently appreciated. MS data consistent with the loss of the first four residues in two different RP-HPLC fractions of mouse H1 has been reported previously (Wang et al., 1997). I anticipate that future MS analyses of H1 in other species will confirm both the occurrence of this event and the limitations of current electrophoretic methods to resolve these forms.
8.7 Potential significance of the molecular heterogeneity in chicken linker histones

The bulk of the data presented in this thesis is concerned with characterizing the molecular basis for the heterogeneity resolved by CX-HILIC for chicken H1 and data addressing the issue of whether functional differences exist for any of these forms is not presented. Nonetheless, aspects of both the proposed N-terminal proteolysis and allelic variation that contribute to the heterogeneity of chicken H1 warrant further investigation.

Gurley and colleagues have presented evidence from analyses of synchronized populations of CHO cells that the serine found at position 1 of all non-allelic variants of mouse H1 apart from H1⁰ (Fig. 1.12), four of the six non-allelic variants of chicken H1 (Fig. 4.2), and many non-allelic variants of H1 in other eukaryotes (Makalowska et al., 1999), differs from other known sites of phosphorylation in H1 in that the serine 1 site is phosphorylated only during mitosis (Gurley et al., 1995). This suggests a regulatory role for phosphorylation at this site that is precluded upon proteolysis of the N-termini of H1 molecules. Thus, regulated N-terminal proteolysis could conceivably represent an epigenetic mechanism for determining the distribution of serine 1-phosphorylated H1 in mitotic chromatin.

Similarly, N-terminal proteolysis may affect putative regulatory functions of α-N-acetylation at the N-termini of H1 histones. Lindner and colleagues have recently reported that only portions of the H1⁰ and H5 molecules recovered from rat and chicken samples, respectively are α-N-acetylated (Lindner et al., 1998; Sarg et al., 1999). This is in contrast to evidence that somatic H1 variants are completely α-N-acetylated and that core histones are either completely α-N-acetylated or completely non-α-N-acetylated depending on the type (see Sarg et al., 1999 and references therein). This suggests the possibility that α-N-acetylation performs a regulatory function, although the nature of such a role is unknown at present. This notion is further
supported by the observation that the extent of α-N-acetylation of H1\(^\circ\), but not H5, increases with age (Sarg et al., 1999). Thus, as suggested above for serine 1 phosphorylation, N-terminal proteolysis may represent an epigenetic mechanism for determining the distribution of α-N-acetylated forms of H1 in chromatin.

The characterization of the CX-HILIC fractions described here provides an initial framework for better understanding the physical basis of the physical heterogeneity of chicken H1 and demonstrates that the technique provides greater resolution of H1 than common electrophoretic and chromatographic techniques. Thus the technique appears to be well-suited for the purpose which originally led me to devise it, that is, to determine if the distributions of non-allelic variants of H1 or various covalently-modified forms of these variants differ between fractions enriched in transcriptionally active and inactive chromatin. Because the technique appears to resolve allelic variants of H1 in some cases, it is possible that the technique may be useful in at least one other way. RP-HPLC peptide mapping of CX-HILIC fractions employing MS/MS analyses represents a direct approach capable of identifying the differences between allelic forms of a given H1 variant. Cataloging the nature and sequence locations of such differences may identify regions of the H1 molecule where variations in sequence are, or are not, compatible with function. Such information could ultimately be used to guide mutational analyses of H1 function.

Despite more than twenty years of intensive study, the function of H1 remains obscure to a large extent. The unexpected degree of heterogeneity of chicken H1 uncovered in the work presented here, and heterogeneity uncovered in MS analyses of H1 from other higher eukaryotes (discussed above), may suggest that this situation is hopelessly complex, or alternatively, that minimal constraints on H1 function are compatible with this large degree of molecular heterogeneity. Clearly, genomic analyses will provide the best information into evolutionary aspects of H1 function. However, because much of the heterogeneity described here appears to arise via post-translational mechanisms, molecular genetic approaches alone cannot provide all the information necessary to understand the biological significance, if any, of this heterogeneity.
Overall, the data presented here demonstrates that the metabolism of H1 is more complicated than previously appreciated and suggests the possibility that this may be relevant to understanding the function of H1. Appropriate use of CX-HILIC and other methodologies described in this thesis should facilitate better understanding of these factors.
Appendices
Figure A1. Electrospray MS analysis of pooled chicken erythrocyte H1 CX-HILIC peak 10

Ions corresponding to each of the molecular species detected, as determined by the instrument operator, are indicated by symbols. The m/z value and charge for the most abundant ion of each species is shown.
Figure A2. Ionspray MS analysis of pooled chicken erythrocyte H1 CX-HILIC peak 10 ions corresponding to each of the molecular species detected, as determined by the instrument operator, are indicated by symbols. The m/z value and charge for the most abundant ion used in the Mr calculation for each species is shown.
Figure A3. Electrospray MS analysis of pooled chicken erythrocyte H5 CX-HILIC peak F-1

Ions corresponding to each of the molecular species detected, as determined by the instrument operator, are indicated by symbols. The m/z value and charge for the most abundant ion of each species is shown.
Table A1. Decrements in predicted H1 variant M<sub>c</sub> values after N-terminal proteolysis<sup>1</sup>

<table>
<thead>
<tr>
<th>Hla</th>
<th>AcS</th>
<th>E</th>
<th>T</th>
<th>A</th>
<th>P</th>
<th>A</th>
<th>A</th>
<th>A</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>129.1</td>
<td>129.1</td>
<td>101.1</td>
<td>71.1</td>
<td>97.1</td>
<td>71.1</td>
<td>97.1</td>
<td>71.1</td>
<td>129.1</td>
</tr>
<tr>
<td>△M&lt;sub&gt;c&lt;/sub&gt;</td>
<td>129.1</td>
<td>258.2</td>
<td>359.3</td>
<td>410.4</td>
<td>527.5</td>
<td>598.6</td>
<td>695.7</td>
<td>766.8</td>
<td>837.9</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>A</td>
<td>P</td>
<td>A</td>
<td>A</td>
<td>P</td>
<td>A</td>
<td>A</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>71.1</td>
<td>71.1</td>
<td>97.1</td>
<td>71.1</td>
<td>71.1</td>
<td>97.1</td>
<td>71.1</td>
<td>97.1</td>
<td>71.1</td>
</tr>
<tr>
<td>△M&lt;sub&gt;c&lt;/sub&gt;</td>
<td>1038.1</td>
<td>1109.2</td>
<td>1206.3</td>
<td>1277.4</td>
<td>1348.5</td>
<td>1419.6</td>
<td>1516.7</td>
<td>1587.8</td>
<td>1684.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hla'</th>
<th>AcA</th>
<th>E</th>
<th>T</th>
<th>A</th>
<th>P</th>
<th>A</th>
<th>A</th>
<th>A</th>
<th>P</th>
<th>A</th>
<th>A</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>113.1</td>
<td>129.1</td>
<td>101.1</td>
<td>71.1</td>
<td>97.1</td>
<td>71.1</td>
<td>97.1</td>
<td>71.1</td>
<td>128.2</td>
<td>71.1</td>
<td>71.1</td>
<td></td>
</tr>
<tr>
<td>△M&lt;sub&gt;c&lt;/sub&gt;</td>
<td>113.1</td>
<td>242.2</td>
<td>343.3</td>
<td>414.4</td>
<td>511.5</td>
<td>582.6</td>
<td>653.7</td>
<td>724.8</td>
<td>821.9</td>
<td>993</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>A</td>
<td>P</td>
<td>A</td>
<td>A</td>
<td>K</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>71.1</td>
<td>71.1</td>
<td>97.1</td>
<td>71.1</td>
<td>71.1</td>
<td>128.2</td>
<td>71.1</td>
<td>71.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>△M&lt;sub&gt;c&lt;/sub&gt;</td>
<td>964.1</td>
<td>1035.2</td>
<td>1132.3</td>
<td>1203.4</td>
<td>1300.5</td>
<td>1371.6</td>
<td>1442.7</td>
<td>1570.9</td>
<td>1642</td>
<td>1713.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hlb</th>
<th>AcA</th>
<th>E</th>
<th>T</th>
<th>A</th>
<th>P</th>
<th>V</th>
<th>A</th>
<th>A</th>
<th>A</th>
<th>P</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>113.1</td>
<td>129.1</td>
<td>101.1</td>
<td>71.1</td>
<td>97.1</td>
<td>99.1</td>
<td>71.1</td>
<td>71.1</td>
<td>97.1</td>
<td>71.1</td>
<td></td>
</tr>
<tr>
<td>△M&lt;sub&gt;c&lt;/sub&gt;</td>
<td>113.1</td>
<td>242.2</td>
<td>343.3</td>
<td>414.4</td>
<td>611.5</td>
<td>682.6</td>
<td>753.7</td>
<td>824.8</td>
<td>921.9</td>
<td>995</td>
<td></td>
</tr>
<tr>
<td></td>
<td>V</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>P</td>
<td>T</td>
<td>P</td>
<td>A</td>
<td>K</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>99.1</td>
<td>71.1</td>
<td>71.1</td>
<td>71.1</td>
<td>97.1</td>
<td>101.1</td>
<td>97.1</td>
<td>71.1</td>
<td>128.2</td>
<td>71.1</td>
<td></td>
</tr>
<tr>
<td>△M&lt;sub&gt;c&lt;/sub&gt;</td>
<td>1064.1</td>
<td>1135.2</td>
<td>1206.3</td>
<td>1277.4</td>
<td>1374.5</td>
<td>1475.6</td>
<td>1572.7</td>
<td>1643.8</td>
<td>1772</td>
<td>1843.1</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hlc</th>
<th>AcS</th>
<th>E</th>
<th>T</th>
<th>A</th>
<th>P</th>
<th>A</th>
<th>A</th>
<th>A</th>
<th>P</th>
<th>A</th>
<th>A</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>129.1</td>
<td>129.1</td>
<td>101.1</td>
<td>71.1</td>
<td>97.1</td>
<td>71.1</td>
<td>71.1</td>
<td>97.1</td>
<td>71.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>△M&lt;sub&gt;c&lt;/sub&gt;</td>
<td>129.1</td>
<td>258.2</td>
<td>359.3</td>
<td>410.4</td>
<td>527.5</td>
<td>598.6</td>
<td>669.7</td>
<td>740.8</td>
<td>837.9</td>
<td>909</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>V</td>
<td>A</td>
<td>A</td>
<td>P</td>
<td>A</td>
<td>A</td>
<td>K</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>99.1</td>
<td>71.1</td>
<td>71.1</td>
<td>97.1</td>
<td>71.1</td>
<td>71.1</td>
<td>128.2</td>
<td>71.1</td>
<td>71.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>△M&lt;sub&gt;c&lt;/sub&gt;</td>
<td>1008.1</td>
<td>1079.2</td>
<td>1150.3</td>
<td>1247.4</td>
<td>1318.5</td>
<td>1389.6</td>
<td>1517.8</td>
<td>1588.9</td>
<td>1660</td>
<td>1731.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hlc'</th>
<th>AcS</th>
<th>E</th>
<th>T</th>
<th>A</th>
<th>P</th>
<th>A</th>
<th>A</th>
<th>A</th>
<th>P</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>129.1</td>
<td>129.1</td>
<td>101.1</td>
<td>71.1</td>
<td>97.1</td>
<td>71.1</td>
<td>71.1</td>
<td>97.1</td>
<td>71.1</td>
<td></td>
</tr>
<tr>
<td>△M&lt;sub&gt;c&lt;/sub&gt;</td>
<td>129.1</td>
<td>258.2</td>
<td>359.3</td>
<td>410.4</td>
<td>527.5</td>
<td>598.6</td>
<td>669.7</td>
<td>740.8</td>
<td>837.9</td>
<td>909</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>P</td>
<td>A</td>
<td>P</td>
<td>G</td>
<td>A</td>
<td>K</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>71.1</td>
<td>97.1</td>
<td>71.1</td>
<td>97.1</td>
<td>57</td>
<td>71.1</td>
<td>128.2</td>
<td>71.1</td>
<td>71.1</td>
<td>71.1</td>
</tr>
<tr>
<td>△M&lt;sub&gt;c&lt;/sub&gt;</td>
<td>1024.1</td>
<td>1121.2</td>
<td>1192.3</td>
<td>1289.4</td>
<td>1346.4</td>
<td>1417.5</td>
<td>1545.7</td>
<td>1616.8</td>
<td>1687.9</td>
<td>1759</td>
</tr>
<tr>
<td>H1d</td>
<td>AcS</td>
<td>E</td>
<td>T</td>
<td>A</td>
<td>P</td>
<td>V</td>
<td>A</td>
<td>A</td>
<td>P</td>
<td>A</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>129.1</td>
<td>129.1</td>
<td>101.1</td>
<td>71.1</td>
<td>97.1</td>
<td>99.1</td>
<td>71.1</td>
<td>71.1</td>
<td>97.1</td>
<td>71.1</td>
</tr>
<tr>
<td>ΔM,</td>
<td>129.1</td>
<td>258.2</td>
<td>359.3</td>
<td><strong>412.4</strong></td>
<td>527.5</td>
<td>626.6</td>
<td>697.7</td>
<td>768.8</td>
<td>865.9</td>
<td>937</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>V</th>
<th>S</th>
<th>A</th>
<th>P</th>
<th>G</th>
<th>A</th>
<th>K</th>
<th>A</th>
<th>A</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9.1</td>
<td>87.1</td>
<td>71.1</td>
<td>97.1</td>
<td>57</td>
<td>71.1</td>
<td>128.2</td>
<td>71.1</td>
<td>71.1</td>
<td>71.1</td>
</tr>
<tr>
<td>ΔM,</td>
<td>1036.1</td>
<td>1123.2</td>
<td>1194.3</td>
<td>1291.4</td>
<td>1348.4</td>
<td>1415.5</td>
<td>1547.7</td>
<td>1618.8</td>
<td>1689.9</td>
<td>1761</td>
</tr>
</tbody>
</table>

1 The first twenty residues predicted for each chicken H1 subtype by the corresponding genes sequences (Coles et al., 1987) is shown in one letter code, assuming complete α-N-acetylation on the first line (Ac = α-N-acetyl). The molecular mass of each residue is listed immediately below it on the second line. The cumulative changes in molecular mass resulting from sequential loss of residues from the N-terminus are shown below the dashes on the third line (ΔM,). The M, shifts resulting from the loss of the first four residues, AcSETA, of H1a, H1c, H1c' and H1d (Δ M, = -430.4), and AcAETA of H1a' and H1b (Δ M, = -414.4), that appeared to be common in some samples are underlined.
Table A2. Decrements in predicted H1 variant $M_i$ values after C-terminal proteolysis$^1$

<table>
<thead>
<tr>
<th>H1a/a'</th>
<th>$K$</th>
<th>$P$</th>
<th>$K$</th>
<th>$A$</th>
<th>$A$</th>
<th>$K$</th>
<th>$P$</th>
<th>$K$</th>
<th>$A$</th>
<th>$A$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>128.2</td>
<td>87.1</td>
<td>128.2</td>
<td>71.1</td>
<td>128.2</td>
<td>97.1</td>
<td>128.2</td>
<td>71.1</td>
<td>71.1</td>
<td></td>
</tr>
<tr>
<td>$\Delta M_i$</td>
<td>1987.9</td>
<td>1859.7</td>
<td>1762.6</td>
<td>1634.4</td>
<td>1563.3</td>
<td>1492.2</td>
<td>1364</td>
<td>1266.9</td>
<td>1138.7</td>
<td>1067.6</td>
</tr>
<tr>
<td></td>
<td>128.2</td>
<td>71.1</td>
<td>128.2</td>
<td>71.1</td>
<td>128.2</td>
<td>71.1</td>
<td>128.2</td>
<td>128.2</td>
<td>128.2</td>
<td></td>
</tr>
<tr>
<td>$\Delta M_i$</td>
<td>996.5</td>
<td>868.3</td>
<td>797.2</td>
<td>726.1</td>
<td>597.9</td>
<td>526.8</td>
<td>455.7</td>
<td>184.6</td>
<td>256.4</td>
<td>128.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>H1b</th>
<th>$K$</th>
<th>$P$</th>
<th>$K$</th>
<th>$A$</th>
<th>$T$</th>
<th>$K$</th>
<th>$P$</th>
<th>$K$</th>
<th>$A$</th>
<th>$A$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>128.2</td>
<td>97.1</td>
<td>128.2</td>
<td>71.1</td>
<td>101.1</td>
<td>128.2</td>
<td>97.1</td>
<td>128.2</td>
<td>71.1</td>
<td>71.1</td>
</tr>
<tr>
<td>$\Delta M_i$</td>
<td>2043.9</td>
<td>1915.7</td>
<td>1818.6</td>
<td>1690.4</td>
<td>1619.3</td>
<td>1518.2</td>
<td>1390</td>
<td>1292.9</td>
<td>1164.7</td>
<td>1093.6</td>
</tr>
<tr>
<td></td>
<td>128.2</td>
<td>71.1</td>
<td>128.2</td>
<td>71.1</td>
<td>128.2</td>
<td>71.1</td>
<td>91.1</td>
<td>128.2</td>
<td>128.2</td>
<td>128.2</td>
</tr>
<tr>
<td>$\Delta M_i$</td>
<td>1022.5</td>
<td>894.3</td>
<td>823.2</td>
<td>752.1</td>
<td>623.9</td>
<td>552.8</td>
<td>481.7</td>
<td>384.6</td>
<td>256.4</td>
<td>128.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>H1c</th>
<th>$K$</th>
<th>$P$</th>
<th>$K$</th>
<th>$A$</th>
<th>$A$</th>
<th>$K$</th>
<th>$P$</th>
<th>$K$</th>
<th>$A$</th>
<th>$A$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>71.1</td>
<td>128.2</td>
<td>97.1</td>
<td>128.2</td>
<td>71.1</td>
<td>128.2</td>
<td>97.1</td>
<td>128.2</td>
<td>71.1</td>
<td></td>
</tr>
<tr>
<td>$\Delta M_i$</td>
<td>1930.8</td>
<td>1859.7</td>
<td>1731.5</td>
<td>1614.4</td>
<td>1506.2</td>
<td>1435.1</td>
<td>1364</td>
<td>1235.8</td>
<td>1138.7</td>
<td>1017.5</td>
</tr>
<tr>
<td></td>
<td>71.1</td>
<td>128.2</td>
<td>71.1</td>
<td>71.1</td>
<td>128.2</td>
<td>71.1</td>
<td>71.1</td>
<td>71.1</td>
<td>128.2</td>
<td>128.2</td>
</tr>
<tr>
<td>$\Delta M_i$</td>
<td>939.4</td>
<td>868.3</td>
<td>740.1</td>
<td>669</td>
<td>597.9</td>
<td>469.7</td>
<td>398.6</td>
<td>327.5</td>
<td>256.4</td>
<td>128.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>H1c'</th>
<th>$K$</th>
<th>$P$</th>
<th>$K$</th>
<th>$A$</th>
<th>$T$</th>
<th>$K$</th>
<th>$P$</th>
<th>$K$</th>
<th>$A$</th>
<th>$A$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>128.2</td>
<td>97.1</td>
<td>128.2</td>
<td>71.1</td>
<td>101.1</td>
<td>128.2</td>
<td>97.1</td>
<td>128.2</td>
<td>71.1</td>
<td></td>
</tr>
<tr>
<td>$\Delta M_i$</td>
<td>2047.9</td>
<td>1919.7</td>
<td>1822.6</td>
<td>1694.4</td>
<td>1623.3</td>
<td>1522.2</td>
<td>1394</td>
<td>1296.9</td>
<td>1168.7</td>
<td>1097.4</td>
</tr>
<tr>
<td></td>
<td>128.2</td>
<td>71.1</td>
<td>128.2</td>
<td>71.1</td>
<td>128.2</td>
<td>71.1</td>
<td>91.1</td>
<td>128.2</td>
<td>128.2</td>
<td>128.2</td>
</tr>
<tr>
<td>$\Delta M_i$</td>
<td>1026.5</td>
<td>898.3</td>
<td>827.2</td>
<td>756.1</td>
<td>627.9</td>
<td>526.3</td>
<td>455.7</td>
<td>384.6</td>
<td>256.4</td>
<td>128.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>H1d</th>
<th>$K$</th>
<th>$S$</th>
<th>$K$</th>
<th>$A$</th>
<th>$A$</th>
<th>$K$</th>
<th>$P$</th>
<th>$K$</th>
<th>$A$</th>
<th>$A$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>128.2</td>
<td>87.1</td>
<td>128.2</td>
<td>71.1</td>
<td>128.2</td>
<td>97.1</td>
<td>128.2</td>
<td>71.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta M_i$</td>
<td>2007.9</td>
<td>1879.7</td>
<td>1792.6</td>
<td>1664.4</td>
<td>1593.3</td>
<td>1522.2</td>
<td>1394</td>
<td>1296.9</td>
<td>1168.7</td>
<td>1097.4</td>
</tr>
<tr>
<td></td>
<td>128.2</td>
<td>71.1</td>
<td>128.2</td>
<td>71.1</td>
<td>128.2</td>
<td>71.1</td>
<td>101.1</td>
<td>128.2</td>
<td>128.2</td>
<td>128.2</td>
</tr>
<tr>
<td>$\Delta M_i$</td>
<td>1026.5</td>
<td>898.3</td>
<td>827.2</td>
<td>756.1</td>
<td>627.9</td>
<td>556.8</td>
<td>485.7</td>
<td>384.6</td>
<td>256.4</td>
<td>128.2</td>
</tr>
</tbody>
</table>

$^1$ The last twenty residues predicted for each chicken H1 subtype by the corresponding gene sequences (Coles et al., 1987) are shown on the first line. The molecular mass of each residue is listed immediately below it on the second line. The cumulative changes in molecular mass resulting from sequential loss of residues from the C-terminus (reading right to left) are shown below the dashes on the third line ($\Delta M_i$).
Appendix B. Publications from work performed during the degree residency

(15 publications of 34 total)


Biard-Roche, J., Gorka, C., and Lawrence, J.J. (1982). The structural role of histone H1: properties of reconstituted chromatin with various H1 subfractions (H1-1, H1-2, and H1o). EMBO J. 1, 1487-1492.


Georgakopoulos, T. and Thireos, G. (1992). Two distinct yeast transcriptional activators require the function of the GCN5 protein to promote normal levels of transcription. EMBO J. 11, 4145-4152.


226


227


