# Compilation analysis of histones and histone genes

Dan E. Wells

Department of Medicine, Stanford University and Veterans Administration Medical Center, 3801 Miranda Avenue, Palo Alto, CA 94304, USA

This compilation of histones and histone genes is an attempt to gather, align and briefly analyze all histone sequences available in the NIH-GENbank and EMBL nucleic acid sequence databases, and the NRBF protein data base as of late 1985. In addition, an attempt was made to include all recently published sequence data not currently available in these data bases. The facilities at IntelliGenetics were used to collect sequence data using the QUEST program. In addition new sequence data was entered from original manuscripts using the GENED program. QUEST and GENED are products of IntelliGenetics (Palo Alto). Any information regarding missing sequence data or erroneous presentation is encouraged.

## PROTEIN SEQUENCES

Data from direct amino acid sequence analysis along with data from translated gene sequences are presented for each of the histone subtypes (Figures 1-5). Consensus sequences were generated for each histone subgroup and numbering is based on the consensus sequence. For the H1-H5 alignments (Figure 1) only the central conserved hydrophobic region (Levy et al., 1982) could be aligned with any degree of certainty. A dash (-) indicates unsequenced or ambiguous regions compared to the consensus. Blank spaces in the protein alignments indicate deletions with respect to the consensus. An asterisk (*) in the consensus indicates the absence of a consensus amino acid for that position. A blank space in the consensus indicates an insertion in one or more of the protein sequences at that point. A dot (·) below a consensus amino acid indicates identity with that amino acid.

## GENE SEQUENCES

Histone gene alignments are displayed for each of the five subtypes with the coding, upstream, and downstream regions displayed separately. In an effort to draw some functional conclusions from these alignments, genes known or presumed to be nonfunctional have been excluded from this compilation.

### Coding regions

The coding regions of the histone genes were aligned using the alignment-analysis program MULTAN (Bains,1986). MULTAN was run on the SUMEX-AIM DEC 2060 computer at Stanford University. In each alignment the consensus (cos) sequence is displayed on the top line followed by each individual histone gene sequence (Figures 6-9). The numbering begins at the first base of the consensus and only bases within the consensus are numbered. A dot (·) under a consensus base indicates identity with the consensus. A dash (-) indicates an ambiguous or unsequenced base. A blank space under the consensus indicates that a deletion in the sequence was introduced to maximize alignment. A gap in the consensus indicates an insertion was introduced into at least one of the sequences to maximize the alignment. Consensus sequences that are underlined indicate codon positions that are largely invariant despite the vast evolutionary distances represented. These coincident codon positions have been observed previously for H4 genes (Turner and Woodland, 1982) and for H3 genes (Wells et al.,1986). These observations have been extended to H2A and H2B genes and updated for the H3 and H4 genes. Although conserved regions within the coding sequence could be of great interest, it should be pointed out that the formal demonstration of the statistical significance of these of these regions has yet to be demonstrated.

### Flanking regions

Where available, sequences upstream and downstream of the coding region were compiled (Figures 10-14). The upstream sequences include the 5' untranslated portion of the mRNA and the proximal promoter regions up to about 220 bases upstream of the ATG initiation codon. The downstream sequences include the 3' untranslated portion of the mRNA and other downstream sequences up to about 150 bases downstream of the termination codon were included. No precise alignments were attempted for these flanking regions and no consensus sequences are shown. The upstream sequences are numbered backward away from the ATG and downstream sequences are numbered beginning with the termination codon. The initiation and termination codons are displayed in bold type.

## CONSERVED FLANKING REGIONS

Conserved upstream and downstream sequence blocks of all histone subtypes have been aligned and displayed together in figures 15 an 16. The numbers between the conserved blocks indicate the number of bases between individual blocks. Where possible, the conserved regions suggested by the original authors were used in figures 15 and 16.

### Upstream sequences

The upstream consensus sequences for each subtype is shown in italics below each group. These conserved blocks have been for the most part described elsewhere. The CCAAT, GATCC, TATAA, and CATTC upstream homology blocks have been reviewed (Hentschel and Birnstiel, 1981). The H2B specific sequence (CTCATTTGCATAC) has been previously identified (Harvey et al., 1982). The H1 specific sequence (AAACACA) has been previously identified (Coles and Wells, 1985). In addition, there is a purine-rich region, specific to vertebrates, present in H4 genes immediately upstream from the CCAAT sequence. As can be seen in figure 15, the degree of similarity within the upstream consensus blocks varies greatly, and the actual significance of most of these regions has yet to be demonstrated.

### Downstream Sequences

The downstream regions of most histone genes are unique in that they lack the canonical AATAAA transcription termination motif usually present in polymerase II transcribed genes. Instead they contain a highly conserved 23 nucleotide sequence at the 3' end of their mRNA containing a hyphenated dyad symmetry motif (figure 16). The 3' end of the mature mRNA occurs immediately following this sequence and immediately prior to a purine rich region. These conserved sequence blocks have been previously reviewed (Hentschel and Birnstiel, 1981) and are updated here. Interestingly, the non-cell cycle regulated vertebrate histone gene variants (22, 62, 66) along with the fungal histone genes ( 27, 28, 51, 52, 78, 79, 80, 100, 101, 103) are conspicuously missing this highly conserved motif. In addition the 3' untranslated region of the H4 mRNAs is pyrimidine-rich, and the 3' untranslated region of the H1 genes contain a spatially conserved TCCPy motif.

## Acknowledgements

I would like to thank Dr. William Bains and Dr. Larry Kedes for their assistance and helpful discussions regarding this compilation. In addition, Doug Hoffman and Lena Arcinas provided excellent editoral assistance. Computer facilities were provided through the National Institute of Health supported SUMEX-AIM facility and through Intelligenetics.

## Table 1. Sequence List

### HISTONE H1

1. Chicken H1-1A1 (Sugarman et al., 1983)
2. Chicken H1-11 (Coles and Wells, 1985)
3. Newt H1 (Stephenson et al., 1981)
4. Xenopus H1-C8 (Turner et al., 1983)
5. Xenopus H1-C2 (Turner et al., 1983)
6. Sea Urchin H1-Sp (Levy et al., 1982)
7. Sea Urchin H1-Pm (Busslinger et al., 1979)
8. Human H1 (Carozzi et al., 1984)
9. Drosophila H1 (Goldberg Ph.D. Thesis, 1979)
10. Rabbit H1-3 (Jones et al., 1974)
11. Rabbit H1-4 (Rall and Cole., 1971)
12. Bovine H1-1 (Liao et al. 1981)
13. Trout H1 (McLeod et al., 1977)
14. Echinolampus H1 (Strickland et al, 1982)
15. Parachinus H1 (Strickland et al., 1980)
16. Goose H5 (Yaguchi et al., 1979)
17. Chick H5 (Krieg et al., 1983)

### HISTONE H2A

18. Human H2A (Zhong et al., 1983)
19. Mouse H2A (Sittman et al., 1983)
20. Sea Urchin H2A-PB (Birnstiel et al., 1979)
21. Chicken H2A (D'Andrea et al., 1981)
22. Chicken H2A-F (Harvey et al., 1983)
23. Newt H2A (Stephenson et al., 1981)
24. Sea Urchin H2A-E3 (Sures et al., 1978)
25. Sea Urchin H2A-PA (Schaffner et al., 1978)
26. Xenopus H2A-L (Moorman et al., 1982)
27. Yeast H2A-1 (Choe et al., 1982)
28. Yeast H2A-2 (Choe et al., 1982)
29. Drosophila H2B (Goldberg Ph.D. Thesis, 1979)
30. Trout H2A (Connor et al., 1984)
31. Bovine H2A (Sautiere et al., 1974)
32. Rat H2A-1 ( Laine et al., 1976)
33. Rat H2A-2 ( Laine et al., 1976)
34. Chicken erythrocyte H2A (Laine et al., 1978)
35. Sea Urchin gonadal H2A-Pm (Wouters et al., 1978)
36. Tetrahymena H2A-1 (Fusauchi et al., 1983)
37. Tetrahymena H2A-2 (Fusauchi et al., 1983)
38. Wheat germ H2A (Rodrigues et al,. 1979)
39. Cuttlefish H2A (Wouters-Tyrou et al., 1982)
40. Starfish H2A (Mertinage et al., 1983)
41. Sipunculus H2A (Kmiecik et al., 1983)

### HISTONE H2B

42. Human H2B (Zhong et al., 1983)
43. Mouse H2B (Sittman et al., 1983)
44. Chicken H2B-A2B (Harvey et al., 1982)
45. Chicken H2B-B (Harvey et al., 1982)
46. Chicken H2B-2BA (Grandy et al., 1982)
47. Newt H2B (Stephenson et al., 1981)
48. Sea Urchin H2B-E3 (Sures et al., 1978)
49. Sea Urchin H2B-PA (Schaffner et al., 1978)
50. Xenopus H2B-L (Moorman et al., 1982)
51. Yeast H2B-1 (Wallis et al., 1980)
52. Yeast H2B-2 (Wallis et al., 1983)
53. Drosophila H2B (Goldberg Ph.D. Thesis)
54. Bovine H2B (Iwai et al., 1972)
55. Crocodile H2B (Van Helden et al., 1978)
56. Trout H2B (Kootstra et al., 1978)
57. Limpet H2B (Van Helden et al., 1979)
58. Starfish Sperm H2B (Strickland et al., 1980)
59. Sea Urchin Sperm H2B-Pa (Strickland et al., 1977)

### HISTONE H3

60. Human H3-5B (Zhong et al., 1983)
61. Human H3-26H (Clark et al., 1981)
62. Human H3-B2 (Wells and Kedes, 1985)
63. Mouse H3-1 (Sittman et al., 1983)
64. Mouse H3-2 (Sittman et al., 1983)
65. Chicken H3-3D (Engel et al., 1982)
66. Chicken H3-4A (Engel et al., 1982)
67. Xenopus H3-LA (Ruberti et al., 1982)
68. Xenopus H3-BOR (W. Bains Ph.D. Thesis 1982)
69. Trout H3 (Connor et al., 1984)
70. Sea Urchin H3-E3 (Sures et al., 1978)
71. Sea Urchin H3-PA (Schaffner et al., 1978)
72. Sea Urchin H3-PB (Birnstiel et al., 1979)
73. Sea Urchin H3-PC (Busslinger et al., 1982)
74. Sea Urchin H3-LP19 (Childs et al., 1982)
75. Sea Urchin H3-Lp21 (Roberts et al., 1984)
76. Drosophila H3 (M. Goldberg, PhD. Thesis 1979)
77. Wheat H3 (Tabata et al., 1984)
78. Neurospora H3 (Woudt et al.,1983)
79. Yeast H3-1 (Smith et al., 1984)
80. Yeast H3-2 (Smith et al., 1984)
81. Sea Urchin H3-LpE (Roberts et al., 1984)
82. Bovine H3-1 (DeLange et al., 1973)
83. Bovine H3-2 (Patthy and Smith, 1975)
84. Shark H3 (Brandt et al., 1974)
85. Pea embryo H3 (Patthy et al., 1973)
86. Buffalo Fish H3 (Hooper et al., 1973)

### HISTONE H4

87. Chicken H4 (Sugarman et al., 1983)
88. Drosophila H4 (Goldberg Thesis, 1979)
89. Mouse H4 (Seiler-Tuyns et al., 1981)
90. Newt H4 (Stephenson et al., 1981)
91. Sea urchin H4-Pm (Schaffner et al., 1978)
92. Sea urchin H4-Sp (Grunstein et al., 1981)
93. Xenopus H4-B (Turner and Woodland, 1982)
94. Xenopus H4-L (Turner and Woodland, 1982)
95. Xenopus H4-LB (Moorman et al., 1981)
96. Human H4-A1 (Heintz et al., 1981)
97. Sea Urchin-Lp19 (Roberts et al., 1985)
98. Sea Urchin-Lp21 (Roberts et al., 1985)
99. Wheat H4 (Tabata et al., 1983)
100. Neurospora H4 (Woudt et al., 1983)
101. Yeast H4-Sc (Woudt et al., 1983)
102. Human H4 (Sierra et al., 1983)
103. Yeast (Smith and Andresson, 1983)
104. Xenopus H4 (Clerc et al., 1983)
105. Tetrahymena H4 (Bannon et al., 1984)
106. Xenopus H4-Z (Zernik et al., 1980)
107. Trout H4 (Winkfein et al., 1985)

108 Pig Thymus H4 (Sautiere et al., 1971a)

109. Bovine hepatoma H4 (Wilson et al. 1970)
110. Pea seedling H4 (DeLange et al., 1969)
111. Rat chloroleukemic tumor H4
(Sautiere et al., 1971b)

Figure 1. Histone H1 and H5 Protein Alignments

```
                                                  10          20
COS                                        ASGPPVSELITKAV*ASKER*GVSLAAI
  1      SETAPVAAPAVSAPGAKAAAKKPKKAAGGAKPRKPAGPSVTELITKAVSASKERKGLSLAAL
  2      SETAPAAAPDAPAPGAKAAAKKPKKAAGGAKARKPAGPSVTELITKAVSASKERKGLSLAAL
  3                                        ----------DLSVTASPERKGVSVASI
  4    TETAATETTPAAPPAEPKQKKKQQPKKAAGGAKAKKPSGPSASELIVKSVSASKERGGVSLAAL
  5   AETASTETTPAAPPAEPKQKKKKQQPKKAAGGAKAKKPSGPSASELIVKSVSASKERGGVSLAAL
  6                          AEKNSSKKVTTKKPAAHPPAAEMVATAITELKDRNGSSLQAI
  7                            TDTAKKVTQKKPAAHPPAAEMVTTAIKELKERKGSSRQAI
  8    SETAPAETATPAPVEKSPAKKKATKKAAGAGAAKRIAAGPPVSELITKAVPASKERN-------
  9       AAPPATVEKKVVQKKASGSAGTKAKKASATPSHPRGGSSLLAIKKYITA------------
 10     SEAPAETAAPAPAEKSPAKKKKAAKKPGAGAAKRKAAGPPVSELITKAVAASKERNGLSLAAL
 11     SEAPAETAAPAPAKSPATPVKKARKKKSAGAAKRKASGPPVSELITKAVAASKERSGVSLAAL
 12       SETAPAAPAAAPPAEKTPVKKKAAKKPAGARRKASGPPVSELITKAVAASKERSGVSLAAL
 13               AEVAPAPAAAAPAKAPKKKAAAKPKKSGPAVGELAGKAVAASKERSGVSLAAL
 14     AASPQKRAASPRKSPKKSPRKSPKKKSPRKRKARSAAHPPVIDMITAAIAAQKERRGSSVAKI
 15 PGSPQKRAASPRKSPRKSPKKSPRKASASPRRKAKRARASTHPPVLEMVQAAITAMKERKGSSAAKI
 16                 TDSPIPAPAPAAKPKRARAPRKPASHPTYSEMIAAAIRADKSRGGSSRQSI
 17                  TESLVLSPAPAKPKRVKASRRSASHPTYSEMIAAAIRAEKSRGGSSRQSI
    30         40        50         60         70
COS KKALAAGGYDV*EKNNSR*IKLALKSLV*KGTLVQTKGTGASGSFKLNKK
  1 KKALAAGGYDV EKNNSR IKLGLKSLVSKGTLVQTKGTGASGSFKLNKKPGETKAKATKKKPAAKPKK
  2 KKALAAGGYDV EKNNSR IKLGLKSLVSKGTLVQTKGTGASGSFRLNKKPGEVKEKAPRKRATAAKPK
  3 K CLSAEGYDV DKNNSR VKVALKSLVRKG-------------------
  4 KKALAAGGYNV ERNNSR LKLALKALVTKGTLTQVKGSGASGSFKLNKKQLETKVKAVAKKKLVAPKA
  5 KKALAAGGYDV ERNNSR LKLALKALVTKGTLTQVKGSGASGSFKLNKKQLETKVKAVAKKKLVAPKA
  6 KKYIAT NFDVQMDRQLLFIKRALKSGVEKGKLVQTKGKGASGSFKVNVQAAKAQASEKAKKEKEKAKL
  7 ANYIKA HFDVEIDQQLVFIKKALRSGVAKGTLVQTKGTGASGSIKL-TRLDRTLSKKVTQPK
  9 -------TYKCDAQKLAPFIKKYLKSAVVNGKLIQTKGKGASGSFKLSASAKKEKDPKAKSKVLSAEKK
 10 KKALAAGGYDV EKNNSR IKLGLKSLVSKGTLVETKGTGASGSFKLDKKAASGEAKPKPKKAGAAKPK
 11 KKALAAAGYB---------------------------------------
 12 KKALAAAGYDV EKNNSR IKLGLKSLVSKGTLVQTKGTGASGSF-----
 13 KKSLAAGGYDV EKNNSR VKIAVKSLVTKGTLVETKGTGASGSFKLNKKVEAKKPAKKAAAPKAKKVA
 14 QSY IAAKYRCDINALNPHIRRALKNQVKSGALKQVSGVGATGRFRVGAVKRSAASANKLKATREKARA
 15 KSY MAANYRVDMNVLAPHVRRALRNGVASGALKQVTGTGASGRFRVGAVAKPKKAKKTSAAAKAKKAK
 16 QKY VKSHYKVGQHADLQ IKLAIRRLLTTGVLKQTKGVGASGSFRLAKGDKAKRSPAGRKKKKKAARK
 17 QKY IKSHYKVGHNADLQ IKLSIRRLLAAGVLKQTKGVGASGSFRLAKSDKAKRSPGKKKKAVRRSTS

  1 PAAKKPAAAAKKPKKAAAVKKSPKKAKKPAAAATKKAAKSPKKATKAGRPKKTAKSPAKAKAVKPKAAKS
  2 KPAAKKPAAAAKKPKKAAAVKKSPKKAKKPAAAADQEGGQEPQEGRQGWPPQEGRQEPGQGKGGEAQGCQ
  4 KKPVAAKKKPKSPKKPKKVSAAAAKSPKKAKKPVKAAKSPKKPKAVKPKKVTKSPAKKATKPKAAKAKIP
  5 KKPVTAKKKPKSPKKPKKVSAAAAKSPKKAKKPVKAAKSPKKKAVKSKKVTKSPAKKATKPKAAKAKIAK
  6 LAQREKAKEKGCSEEGETAEGSRPKKVKAAPKKAKKPVKKTTEKKEKKKTPKKAPKKPAAKKSTPKKTPK
  9 VQSKKVASKKIGVSSKKTAVGAADKKPKAKKAVATKKTAENKATKAAKPKAAKPKKAAKSPKKVKKPAAA
 10 KPAGATPKKPKKAAGAKKAVKKTPKKAPKPKAAAKPKVAKPKSPAKVAKSPKKAKAVKPKAAKPKAPKPK
 13 AKKPAAAKKPKKVAAKKAVAAKKSPKKAKKPATPKKAAKSPKKATKAAKPKAAKPKKAAKSPKKVKKPAA
 14 RAKAKKAKAAARRKAAAAKRKAAAAKRRAAKKARKAKAKP
 15 AAAAKKARKAKAAAKRKAALAKKKAAAAKRKAAAKAKKAKKPKKKAAKKAKKPAKKSPKKAKKPAKKSPK
 16 STSPKKAARPRKARSPAKKPKAAARKARKKSRASPKKAKKPKTVKAKSLKTSKPKKARRSKPRAKSGARK
 17 PKKAARPRKRSPAKKPKATARKARKKSRASPKKAKKPKTVKAKSRKASKAKKVKRSKPRAKSGARKSPKKK

  1 KAAKPKAAKAKKAATKKK
  2 AQGDQTQGGQGEEDGSQEEV
  4 AKPKIAKAKAAKGKKAAAKK
  5 AKAAKGKKAAAKK
  6 KAAAKKPKTAKPKKPA-KKAAKSK
  9 KKKTEKAKAKDAKKTGIIKSKPAATKAKVTAAKPKA
 10 AAKAKKTAAKKKK
 13 AKK
 15 KKKAKRSPKKAKKAAGKRKPAAKKARRSPR-KAGKRRSPKKARK
 16 SPKKK
```

Figure 2. Histone H2A protein Alignment

```
                             20                   40                  60
COS   SGRGKQGGKARAKAK     SRSSRAGLQFPVGRVHRLLRKGN YAERVGAGAPVYLAAVLEYLTAEILEL
18    ...............     T....................-- .S.........................
20    ---------------     -------------....F..... .AK...G.....M..............
21    ...............     ....................... ...........................
22  AGGKA..DS...K...V     ...Q.............H.KTRTTSHG....T.A..S..........V...
23    ..... S....T...     T................F..... ..K...G.....M..............
25    ..... S....T...     T................F..... ..K...G.....M..............
26    .........T.....     T...................... ...........................
27    .. .. ....GSA..ASQ  ...AK...T...........R.. ..Q.I.S....................
28    .. .. ....GSA..ASQ  ...AK...T...........R.. ..Q.I.S....................
29    ..... ...VKG...     ...N..........I........ ................M......V...
30    .....T.........     T...................... ...........................
31    ...............     T...................... ...........................
32    ...............     ....................... ...........................
33    ...............     ....................... ...........................
34    ...............     ....................... ...........................
35    ..... .A.GK....     .................F..... ..N.................A......
36    .TT.. ....KG.TASSKQV...A..........IS.F.K..R .S..I.T.............A..V...
37    .TT.. ....KG.TASSKQV...V..........IS.F.K..R .S..I.T.............A..V...
38    A..K  ..D .K..V     T..VK.........IG.Y.K..R ..Q..GS............--------
39    ..... ...VKG.S.     T.............I........ ..Q.............M...A..V...
40    ..... .........     .................F....  ................M...A......
41    ..... ....KG.S.     ..............I........ ....I...........M...A.AV.V.

                  80                 100                 120
COS  AGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGGVTIAQGGVLPNIQAVLLPKKTESHHKAKGK
18   .................................R..............................
19   ---------------------------.................................P...
20   ..........S..........V........----------------------------------
21   .................................K.....................D.. .....
22   ....SK.L.VK..T.........G...LDS.I-KA...G...I.H.HKS..G..GQQKTA
23   ..--...................--....-----------------------------------
24   ..........S..........V.................................AKSS
25   ..........S..........V.................................AKSS
26   .....................V...........R.............F.........SKS..S.
27   .........................D.......N............HQN.....SAKAT..SQEL
28   .........................D.......N............HQN.....SAKAT..SQEL
29   ................................S.......................K
30   .....................V..................................KAKV..
31   .................................K..............................
32   .................................K..............................
33   .................................K..............................
34   .................................K.....................D.. .....
35   .......................................................G.KSSK
36   .....K.......V....L............MANT...D.......NPM...S.SKKTESRGQASQDI
37   .....K.......V.,..L............MANT...D.......NPM...S..KKTSE.EH
39   ..........S.....................S......................QKAA.
40   .............N..................S......................AKAA.
41   ................................S......................QKSK
```

Figure 3. Histone H2B Protein Alignments

```
                                       20                  40
COS                  PEPAKSAPAPKKGSKKALTKAQKKDGKKRKRSRKESYSIYVYKVLKQVHPDTGISS
42                   ...........................SAAH ............I...........
43                   .................L....................V.....I...........
44                   ....................T...GD..............................
45                   ............A.......T...GD..............................
46                   .................V..T...GD....K.........................
47                   -------------------------------------------------.......
48                      APTAQVA.......VKGTKTAG....N.......G..I...............
49                      APTGQVA.......VKPPRASG....H.K.....G..I............V..
50                   .................A..T........SKT.....A..................
51                SAKAEKKPASKAPAEKKPA.KKTSTST.....SKA...T..S.I......T.......Q
52                SSAAEKKPASKAPAEKKPA.KKTSTSV.....SKV...T..S.I......T.......Q
53                      PPKTSGKAAKKAGKAQKNITKTD..K..K.....A..I...............
54                   ......................................V.................
55                   .................V..T........---------------------------
56                   ........  .......V..TAG.G............A..................
57                       PPKVSSKGAKKAGKAKAARSGD.....R........I............V..
58                        PPKSGKGQKKAGKAKGAPRSD...R.K.....G..I...M...........
59 SQKSPTKRSPTKRSPTKRSPQKGGKGGKGAKRGGKAGKRRRGVQV..R.R.R...G..I...............

        60                  80                 100                 120
COS KAMGIMNSFVNDIFERIAGEASRLAHYNKRSTITSREIQTAVRLLLPGELAKHAVSEGTKAVTKYTSSK
42  ...................................................................A.
43  .....................................................................
44  ...............................R.....................................
45  ...............................R.....................................
46  ...............................R.....................................
47  ---N................IF---------......F....---------------------------
48  R..V................S....Q...K.......................................
49  ...T....................TQ...K.....................................V.
50  ............V...............N..R...................................A.
51  .S.S.L............T...K..A...K...SA.........I..............R.....S..TQA
52  .S.S.L............T...K..A...K...SA.........I..............R.....S..TQA
53  ..................A..................................................
54  .....................................................................
55  -----...............................---------------------------------
56  ....................S................................................
57  ..................A..................................................
58  ..................A..................V............................T..
59  R..SV.......V.....A..G..TT..R...VS...V............................T.R
```

Figure 4. Histone H3 Protein Alignments

```
              10        20        30        40        50        60        70
COS   ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTELLIRKLPFQRL
60    ......................................................................
61    ..........................--------------------------------------------
62    ..............................S.......................................
63    ......................................................................
64    ......................................................................
65    ......................................................................
66    ..............................S.......................................
69    ......................................................................
70    ......................................................................
71    ......................................................................
72    -----------------------------------------------------------...........
73    ......................................................................
74    ......................................................................
75    ......................................................................
76    ......................................................................
77    ........................................F...........K.................
78    .....................S........S..........K............................
79    .....................S........S..........K...........F................
80    .....................S........S..........K...........F................
81    ......................................................................
82    ......................................................................
83    ......................................................................
84    ......................................................................
85    ........................................F...........K.................
86    ......................................................................

              80        90       100       110       120       130
COS   VREIAQDFKTDLRFQSSAVMALQEASEAYLVGLFEDTNLCAIHAKRVTIMPKDIQLARRIRGERA
60    .........................C.......................................
61    ................A.IG.............................................
63    .........................C.......................................
64    .................................................................
65    .................................................................
66    ................A.IG.............................................
67    -------.........A.IG......................................-------
68    -------......................................................----
69    .................................................................
70    ..........E.....................---..............................
71    ..........E......................................................
72    ..........E..............--..................................----
73    ..........E......................................................
74    ..........E......................................................
75    ..........E......................................................
76    ............................................---------------------
77    ...................S.....A.......................................
78    ..................IGL...SV.S...S.................QS........L....N
79    ..................IG....SV.....S.......A.........QK...K....L....S
80    ..................IG....SV.....S.......A.........QK...K....L....S
81    ..........E......................................................
82    .................................................................
83    .........................C.......................................
84    .................................................................
85    ...................S.....A.......................................
86    .................................................................
```

Figure 5. Histone H4 Protein Alignments

```
                   .         20         .        40         .
COS       SGRGKGGKGLGKGGAKRH RKVLRDNIQGITKPAIRRLARRGGVKRISGLIY
87        .................. .................................
88        T................. .................................
89        .................. .................................
90        -------------..... .................................
91        .................. ...................--------......
92        .................. .................................
93        .................. .................................
94        ---............... .................................
95        .................. .................................
96        .................. ........H........................
97        .................. .................................
98        .................. .................................
99        .................. .................................
100       T................. ..I..........................AM..
101       .................. ..I..............................
102       .................. .................................
103       .................. ..I..............................
104       .................. .................................
105       A. ......M..V.....S..SNKAS.E....................SF..
106       -------------..... ............................-----
107       .................. .................................
108       .................. .................................
109       .................. .................................
110       .................. .................................
111       .................. .................................

                 60         .        80         .       100
COS       EETRGVLKVFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG
87        ...................................................
88        ....................-------------------------------
89        ...................................................
90        ..................---------------------------------
91        .........---------------------------------.........
92        .....................C.............................
93        ...................................................
94        ...................................................
95        ...................................................
96        ...................................................
97        .....................C.............................
98        .....................C.............................
99        ........I................R.........................
100       ........T......................SL..................
101       ..V.A...S...S....S.............SL..................
102       ...................................................
103       ..V.A...S...S....S.............SL..................
104       ...................................................
105       DDS.D.........V..........R.........................
106       ----------------------------.......................
107       ...................................................
108       ...................................................
109       ...................................................
110       ........I................R.........................
111       ...................................................
```

Figure 6. H2A Coding Sequence Alignments

```
                        20                  40
cos       ATGTCTGGAAGAGGAAAGAAAGGTGGAAAGGCCCGCACCAAGGCAAAG
 18       ........TC.C..C..AC....C..T..A..T...G...........
 21       .....G..GC.C......C.G..C..G.....G...G.......C...
 22 ATGGCAGGTGGGAAGGCT..G...G.CA.C..G......AAGG.G.....GGT.
 24       ........C.........   A..........................
 25       ..............T..A   A.............T...........A
 26       ..............C..AC....C..C...A.T...G.T.........
 27       .....C..T   ..T..A   .....T..A..TG.TT.AGCT..T..A
 28       .....C..T   ..T..A   .....T..A..TG.TT.AGCT..T..A
 29       .........C.T......   .....C..A.TGAAGGGA.........
 30       ...AGC........C..A.CC..A..C......A.GG.G.........

                           60                  80
cos          TCTCGCTCATCCCGCGCCGGTCTGCAGTTCCCAGTGGCCGT
 18          A....G..T..T..T..A...T......T..T........A
 20          --------------------------------------...
 21          ..G.....G..G..G....G...........C........C
 22          ........GCAGA.A....AT..........C........C
 24          A.G...........T..A..G..C....T........A...
 25          .....T...........T.....C.............A...
 26          A..........T..G.....G.............C......
 27 GCTTCTCAA...A.A..TG.TAAG..T...T..ACA.......C..TA.A
 28 GCTTCTCAA...A.A..TG.TAAA..T...T.AACA.......T..TA.T
 29          ..C......AA...T........T..A....T.........
 30          A.A...........T.....G..C.......C.........

      100                 120                 140
cos GTTCACCGGTTGCTGCGCAAGGGCAAC   TACGCAGAGAGGGTCGGCGC
 18 ..G.....CC....C.......----.   ...T.C...C.C.....G..
 20 .....T.....T..C..A.........   ..T...A............G
 21 ..G......C.................   .....G...C....G.....
 22 A.C..T...CAC...AAG.C.C...C.ACGAG.CAT.G.C.......G..
 24 .....T.....T..C..A.........   ..T...A............G
 25 ........A..T..A.....A......   ..T...A............G
 26 .........C.CT..A.G........T   ..T..C...C....G..A..
 27 ..G...A.A.....AA.A.GA..T...   .....CC.A..AA.T..TT.
 28 ..G...A.A.....AA.A.GA..T...   .....CC....AA.T..TT.
 29 A.......T.....C..G.....A...   .........C.T..T..T..
 30 ..G...A..C.......T..A......   .....C...C.T..C.....

                160                 180
cos TGGCGCTCCAGTCTACATGGCTGCCGTCCTGGAGTACCTGACTGCCGAAA
 18 ......C..G..G..TC.C..G..G..G..T...........C.....G.
 20 ...A.....C....................A........C..........
 21 C.....C..GC.G...C.A..G.....G..............G.....G.
 22 CAC...CG.C..G....GC......A....C.....T..C.....T..GG
 24 ...A.....T....................A........C..........
 25 ...G..A..............C..T...T........T............
 26 C..A.....G.....TC....C..A..G..C.....T.....C..T..G.
 27 ...T............T..A....T...T....A..TT..G.C..T....
 28 ...T...........TC..A....T...T.A..A..TT..G....T....
 29 A...........T...C.A........AA....A..T...G.C..T..GG
 30 ......A.....G...C....C..A..G..C..............T..G.
```

Figure 6. (cont)

```
      200                 220                 240
cos TCCTGGAGCTGGCTGGCAACGCTGCCCGCGACAACAAGAAGACTAGCATC
 18 ....A........G.....T..G....................CC.....
 20 ..T....A..C..A..........................AT....G...
 21 ..........A..G........G....................GC.....
 22 ........T....A........CT..AAG...CTG...GT..AGC.....
 23 -----------......------.....G..............CC.....
 24 ..T....A..C..G..........................AT....G...
 25 .T..C.....C..............T..............AT....G...
 26 ..T.....T....C...........T..G..T.....A.....CC.....
 27 .TT.A..AT.A.....T..T.....TA.G..T...........C..A..T
 28 .TT.A..AT.......T..T.....TA.A..T.....A..A..C..A..T
 29 .T..C...T..........T.....T..T.................A..T
 30 ........T....C..A...........T...............C.T...

                260                 280
cos ATCCCCCGCCACCTGCAACTGGCCATCCGCAACGACGAAGAACTCAACAA
 18 .....G............T.............T.....G..G..T..T..
 19 ------------------------------------------..G.....
 20 .....A........T.....C..TG.G..T..T..T..............
 21 .................G....................G..G........
 22 .CT........TT....G.....G......GG.........GT.GG.TTC
 23 ......A.G.....C..G..C.--.......C-.....G..G........
 24 .....A........T.....C..TG.G..T..T..T..............
 25 ..T.....T..T..T.....T...G.G.......................
 26 ......A.G........G..C..TG.G........T..G..G........
 27 ..T..AA.A..TT.....T....T...A.A..T..T..C...T.G.....
 28 ..T..AA.A..TT.A...T........A.A..T..T..T...T.G.....
 29 ..T..G..T..T..........................G..GT.A.....
 30 ........T........G.....AG....T........G..G..G.....

      300                 320                 340
cos GCTGCTGGGAGGGGTGACCATCGCCCAAGGTGGTGTTCTGCCCAACATCC
 18 A..TT....GC.T...........G..G.....C...T.G..T..T..T.
 19 ....T....CC.C...........G..G..C..C..C.............
 21 .........CAA............G..G..C..G..G.............
 22 C..CA.CAA.---.CC.....A..GGGG..A..C..CA.C...C......
 23 .......-------------------------------------------
 24 ...TT....T........G.....T.........................
 25 ...C..C...........G.................C.............
 26 A.....C.....A..C..T.....T..G..C..G..............T.
 27 ...AT....TAAC..T.....T...............T....A.....T.
 28 ...AA....TAAT..T.....................T....A.....T.
 29 ......CTCC..C..C..A..T..A........C..GT....T..T..A.
 30 A.....T..C..C..C........T..G.....................

                360                 380                 400
cos AGGCCGTGCTGCTGCCCAAGAAGACCGACAGCCAC   AAGGCCAAGGCCAAG
 18 ....G...........T.....A..T..G.....TCAT..........G....TGA
 19 ............................G......CAT...C.......A...TAA
 21 ...................................   ...............TGA
 22 .CAAGTCT...A.CGGG......GG.C.GCAGA.A   .CC..GTAG
 24 .A...........C........A...               ..T..AT.A.GCTAG
 25 .A...........T.........GGT               ..A..AT.G...TGA
 26 ..TT..................A.....G...TC.AAGTCG......AG....TGA
 27 .TCAAAACT..T....A......T.T.CC.AGGCTACC.....TTCTCAAG.ATTATAA
 28 .CCAAAACT..T....A......T.T.CC.CGACTGCC.....TTCTCAAG.ACTGTAA
 29 ....T..T...T................G.AG------------------------
 30 ....A........C...........T..G.AGGC.      .T...A......TAA
```

Figure 7. Histone H2B Coding Sequence alignments

```
                                  20
cos             ATGCCTGAGCCGGCCAAGTCCGCTCCCGCCCCCAAGAAG
 42             .....A.....A..G.....T...........G..A...
 43             ..........C...............T.....G......
 44             .....C........T........G........G......
 45             .......................A...............
 46             .......................A...............
 48                      ATG..TC.AA.A....AA.TTG.T.....A
 50             ........A..A..............A...G.A.....A
 51       ATGTCTGCTAAA.CCGAAAAG.AAC.A..CT..AAAG..CCAGCT
 52       ATGTCCTCTG.C.CCGAAAAG.AAC.A...T..AAAG.TCCAGCT

    40                  60                     80
cos  GGCTCCAAGAAGGCGGTCACCAAGACCCAGAAG   AAGGGCGGCAAGAA
 42  .................G..T...G.G......   ..A.A.......
 43  ..............CC.G......G........   ..G.A.........
 44  .....T...........................   .......A......
 45  .C...............................   .......A......
 46  .................................   .......A......
 48  ..............A....AAGGC...A...C.GCC:GC..T........
 49  -------------------------------CT    GC..T........
 50  ...........A....CA........T......   ..A.A...G..A..
 51  .AAAAG..ACCA..C.CT.AA.....TTCC.CTTCC.CT.AT..T.....
 53  .AAAAG...CCA..T.CC.AG..A..ATCA.CCTCCGTC.AT..T.....

                100                 120
cos  GAGCAAGAAGAGCAGCAAGGAGAGCTACTCGATCTACGTGTACAAGGTGC
 42   ...GCAGC.CA.C.............T..C.....T...........T.
 43  .C.....CGC...C.................G.G................
 44  .C...........C....................................
 45  .C...........C....................................
 46  .C...........C....................................
 48  ...G..C.G..AA..G........T..TGGA......A.C.....A..C.
 49  ...GC.T.G..AA..G............GGT......A.C.....A..C.
 50  .C...G.....CA..G........T..TG.C..T................
 51  ...A.GC...GCT..A.....A.CA.....TTCT...A.T.....A..TT
 52  ...ATCT...GTT..A.......C...T..CTCT..TA.C.....A..TT

      140                 160                 180
cos  TGAAGCAGGTGCACCCCGACACCGGCATCTCCTCCAAGGCCATGGGCATC
 42  ..........C.....T...........T..G..................
 43  .......A..........................................
 44  ........C......................G..................
 45  ......................G........G..................
 46  ......................G........G..................
 47  --------------..G...........................AA....
 48  .C........T..T..A..T............AGTCG........T....
 49  .C........T.....T.....T..TG.....AG.CG.......ACA...
 50  ...................T...........G............A.....
 51  .......AACT.....T.....T..T..T...CAA...T.....TCT...
 52  .......AACT.....A.....T..T..T...CAG...T.T...TCT..T
```

Figure 7. (cont)

```
                 200                 220
cos   ATGAACTCGTTCGTCAACGACATCTTCGAGCGCATCGCCGGCGAAGCGTC
 42   .....T.....T..G........T..............A..T..G..T..
 43   ..............G.......................G..A..G.....
 44   ............................................G.....
 45   ............................................G.....
 46   ............................................G.....
 47   .......:C.....G.........--------------------------
 48   ......AGC.--..--................A..T.........T.T..
 49   ......AGC..T........T...........G..............C..
 50   ........C..T........TG.G..T...........A..G.....C..
 51   T.......T.....T.....T.....T..AA.A.....TACT.....T..
 52   T.......T.....T.....T.....T..AA.A..T..TACT.....T..

       240                 260                 280
cos   CCGCCTGGCGCATTACAACAAGCGCTCGACCATCACGTCGCGGGAGATCC
 42   ......T.............................C..CA.........
 43   ......A................................C..........
 44   G...........C.....................................
 45   G...........C.....................................
 46   G...........C.....................................
 47   ----------------------------........:..CA.:...T...
 48   ......C..T..G........AAAG.--..--...GCAGT..C.....T.
 49   ...T..CA.C..G.........AAG..A.......GTAGC........T.
 50   ......A..T...........C.....C........C..C..........
 51   TAAAT....TGCG..T......AAG..T..T...T.TG.TA.A..A..T.
 52   TAAAT....CGC...T......AAA..C..T..TT.TG.TA.A..A....

                 300                 320
cos   AGACCGCCGTGCGCCTGCTGCTGCCCGGCGAGCTGGCCAAGCACGCCGTC
 42   ....G....................T..G...T................G
 43   ....G.......................G....................G
 44   ....A........G................................G...
 45   ...:G........G................................G...
 46   ....A........G................................G...
 47   .......T...:T.G........---------------------------
 48   ..........C.....CA.T..C.....A........A........T..G
 49   ................T..T..C..A..A...T................G
 50   .......G..C..A...........T..G...T.......A........G
 51   .A.....T..TA.AT..A.CT.A..A..T..AT...T.....T..T...
 52   .A..A.....TA.AT..A.CT.A..T..T..AT...T..A..T......

       340                 360                 380
cos   TCCGAGGGTACCAAGGCGGTCACCAAGTACACCAGCTCCAAGTAA
 42   ...........T.....C..................G........
 43   ..G.....C........T....................T....G.
 44   ..G.....C....................................
 45   ........C...................................G
 46   ............................................G
 48   AG...............A..G..G..A.....T.C.........G
 49   AGT.....G.....A..A..G.............C.GT.......
 50   ........C........T..................G........
 51   ..T..A.....T.GA..T..T.........T.TTC...T.CTC...C.TAA
 52   .....A.....T.G...T..T.....A...T..TC...T.CTC..GC.TAA
```

Figure 8. Histone H3 Coding Sequence Alignments.

```
                      20                  40                  60
COS ATGGCCCGTACCAAGCAGACCGCTCGCAAATCCACCGGAGGTAAGGCTCCCCGCAAGCAGCTGGCCACCA
 60 .....T.....T..A.....A.................C.....A..G..A..............T....
 61 .....G.....G........G..G..T..G..G..G..C........G..A..T..............T.
 62 .....T.....A........T..C........G.....T.....A..A...A.G.....A.....T..A.
 63 .....T.....T.................G..T.....C..C.....C..G...................
 64 .....T.....T.................G.....T..C..C.....C..G...................
 65 .....G.....G........G..G..T..G..G..G..C..G.....G......................
 66 .......................C.....G........C..C.....G......................
 69 ......A.A........A....................T..C..A..A...A.G........C.......
 70 .....A..C.......................T..A.....G.......................A....
 71 --...A..A...........G...........T..A........A..C.................A....
 73 .....A..C.......................T..A.....G.......................A....
 74 ................................T........A........A...........C.......
 75 .......................A..T.....T..T.....A..A.....A...........C.......
 76 .....T...........A..T...........G..T..T..A.....G..A.....A..A.....T..T.
 77 ........C...........G..GA.G..G..G.....C..C.....G..GA.G...........G....
 78 ........C..T...........C.....G........T..C.....C.....T........C..TT...
 79 ......A.A..A.....A..A..AA.A..G.....T..T........C..AA.A.....AT.A..TT.T.
 80 ......A.A..T..A..A..A...A.A........T..T.....A..C..AA.A..A..AT.A...T...
 81 .....A..C...........G...........T........G.........A.A......T....A....

            80                 100                 120                 140
COS AGGCTGCCCGCAAGAGCGCCCCCGCCACCGGCGGAGTGAAGAAGCCTCACCGTTACAGGCCTGGTACCGT
 60 .......T...........G..G..T........T.....A...............C.T..G.....T..
 61 ....G..TA.A-----------------------------------------------------------
 62 .A..C..T........T..G...T.T..T..A..G........A.....T.................T..
 63 ....C.................G...........C.................C...C.T..C..C..T..
 64 ....C.................G...........C........A........C...C.T..C..C.....
 65 ....G.....T........G..G.....G.....G...........A........TC....C.....G..
 66 ....G........A.....G...T.T..T.....G...........G.....C...C..........T..
 69 .......G..............G...........C..........................C..C.....
 70 .A......A.A.....T...........T..A.....C...........T..A...........C..A..
 71 ........A.A.....T.....A........A.....A...........T..C.....A........A..
 73 .A..A...A.A.....T...........T..A.....C...........T..A...........C..A..
 74 ....A..T......TC......A.....T........C...........T..............A.....
 75 .............ATC......A..............C........C..T...........C..A.....
 76 ....C..T........T..T..AC.......A..T...........C.....C..TC.C.....A.....
 77 ....C..T......TC......G...........C...........C.....C..CC.C..C..C.....
 78 ..............TCG......T..........T..C........C..........A...C........
 79 ........A....ATC......AT.T.....T..T..T............A.A..T.A...A........
 80 ........A.A..ATC......AT.T.....T..T..T............A.A..T.A...A.....T..
 81 ....G...A.G.....T.....A.....A..G.....T..A.....A..TA.G...........A.....
```

Figure 8. (cont)

```
                     160                 180                 200
COS GGCTCTCCGTGAGATCCGTCGCTACCAGAAGTCCACCGAGCTTCTGATCCGCAAGCTGCCCTTCCAGCGT
 60 ......G...........C........A.....G......T.G.....T..G........G........C
 62 ...G........A..TA.A..T..T...........T..A........T.....A..T............
 63 ...A..G..C........G..............G........G.................G........C
 64 ...G..G..C........G..............G........G.................G........C
 65 ...G..G..C..A.....G.................G..A..G...........A...........A..C
 66 ...C........A...A.G.....T........A.....A...T.A...........T............
 69 ......TA.A...........T..............T.....G...........A.....T........C
 70 C..C..GA.A.....T..C............AG...T........C.....A..A.....A.........
 71 C..AT.GA.A.....T...............AG......A.....T.....A.....A...........C
 72 ----------------------------------------.....C.....A..A.....A.........
 73 C..C..GA.A.....T..C............AG......A.....C.....A..A.....A.........
 74 C..............................AG............C...........C............
 75 C..............................AG............C...........C............
 76 ...CT.G.....A..T...........A...AG............A..............T.........
 77 ...G.....C........CAAG.........AG...G.....G..C...........C...........C
 78 C..............T....................T....................C............
 79 T...T.GA.A..A...A.AA.A.T...A..A..T..T..A..GT.....A.A...T....T.....AA.A
 80 T..CT.GA.A..A..TA.AA.A.T...A..A..T..T..A..GT.....A.A...T.A..T.....AA.A
 81 C..C..GA.A.....T..C............AG..................A..A..C..A.........

           220                 240                 260                 280
COS CTGGTGCGTGAGATCGCTCAGGACTTCAAGACCGACCTGCGCTTCCAGAGCTCCGCTGTCATGGCTCTGC
 60 ........A..A.....C..A..............T..T..............T..G..A.....G....
 62 ........A..A..T...........T..A..A..T...............G.A...A..GGT...T...
 63 T.......C........G...................................G..C.............
 64 T.......C........G...................................G..C.............
 65 ........C........G::............T..T.................G..C........G..C.
 66 ...........A..T..............A..A..T...............G.T...A..GGT...T...
 67 ----------------------..........T...T..A.G........TG.A...A.TGGT.....T.
 68 ----------------------.T..T...........A..G...........G..C.............
 69 ........A..A..T..C........T.....T.................T.....A..G.....C....
 70 ..A...........T..A..............A..G..A..T........T........G.....C..T.
 71 .....A.................T...........G..A..T.................T.....C..C.
 72 ..A...........T..A..............A..G..A..T........T........G.....C..T.
 73 ..A...........T..A..............A..G..A..T........T........G.....C..T.
 74 .....C.......................A.....G..C.................C........C..C.
 75 .....C.....A.......................G..C.................C...........C.
 76 ...........A..............T.....G...T....A...........G..G..T..........
 77 ..A.....C........C..............................TC......C...TCC..C....
 78 ..C..C........T..C............T.......C..............T..CA..GGCCTC..C.
 79 T....CA.A..A........A..T............T..A.A..T..ATCT..T..CA..GGT..CT...
 80 T....CA.A..A........A..T............T..A.A..T..ATCT..T...A..GGT...T...
 81 ..A...A.G.....T..C...........A..A..G..C...........T.....C..G.....C..T.
```

Figure 8. (cont)

```
                      300                 320                 340
COS AGGAGGCCAGCGAGGCCTACCTGGTCGGTCTCTTTGAGGACACCAACCTGTGTGCCATCCACGCCAAGCG
 60 .......TT.T.........T....A..G..............A.....T..C........T..T.....
 62 .......A..T........T.....T..C..T.....A.................T.....T.....A..
 63 ........T.T...........C..G.....G....................C.................
 64 .......G..............T..G.....G....................C.................
 65 .......G..............T..G..C.....C........T..T..C....................
 66 .......A..T..A......T....T..C..G.....A..T..............T.....T.....A..
 67 ....A..A........T..T.....T..C........A..T.................T..T......A.
 68 ..........T.....T..T.........T.G....................C...............A.
 69 .......A........T...........C..G..C.................C...............A.
 70 .A..A...........A.....A..T..C.------GA..............................A.
 71 .A.....A.......................G..C...........T.....C...............A.
 72 .A..A...:::::...A........T..C...........T...........................A.
 73 .A..A...........A........T..C...........T...........................A.
 74 .A........G..A..T...T.............C..............T..C.................
 75 .............A..T...T..........T..C..............................T.....
 76 ....A..T.....A..T...T....T........C..A..T......T..........T..T......--
 77 ........G.............C..G..C.....C..............C..C.................
 78 .....T..GT....T.T.....C...TC......C..............C..C..T........T.....
 79 .A..AT.TGT...A......T.A...TC.T.A.....A..T......T..GC......T...........
 80 .A..AT..GT...A..A...T.A...TC.T.G.....A.....T..T...GC...T..T.....T.....
 81 .A..A..T.....A.....T........G..A....................C........T..T...A.

            360                 380                 400
COS TGTCACCATCATGCCCAAGGACATCCAGCTCGCCCGCCGCATCCGCGGAGAGCGTGCTTAA
 60 A..G..T..T........A..............T........T........AA.A..G...
 62 ...A..A..T.....A..A...........A..A........A..T.....A.........
 63 ..............................G.....T...........G...A.G......
 64 ..............................G........T........C.....G......
 65 C........T...........T...........T.....T.....T........C..C.G.
 66 ......A........A..A..T........A..A........T..T..G........C...
 67 A..A.....T..................T.A...A.AA.----------------------
 68 G.............................G......A.G.....G---------------
 69 G..G..........................G.....T..T..T...........C..A...
 70 G..T..............A.................T..A...........A..C..C..G
 71 G..A.................T.....A.....T..T..A...........A.....C..G
 72 G..T..............A.................T..A........:::::--------
 73 G..T..............A.................T..A...........A..C..C..G
 74 ..............................T........T.....T..C............
 75 ...............A..............T........T.....T..C........C...
 77 C............................................T..C...A.G..C..G
 78 ..........CA.AG.........................C.......T.....CAAC...
 79 ......T...CAAAAG.....T...A..T.G..TA.AA.AT.AA.A..T..AA.AT.A..G
 80 ...T..T...CAAAAG.....T...A.AT.G...A.AA.AC.AA.A..T..AA.AT.A.G.
 81 G..T...........G..A..T........G.....T..A.....T.....A..C..C..G
```

Figure 9. H4 Coding Sequence Alignments

```
                        20                  40
cos   ATGTCTGGACGAGGCAAAGGAGGAAAGGGACTGGGAAAAGGAGGCGCCAAGCGCCA
 87   ........CA.......G..C..G.....G..C..C.....G..T...........
 88   ---G....T..T..T........C..A..CT.......G..T...........T..
 89   ........CA....A..G..T........T..A..C..G..T..............
 90   ------------------------------------------..G..T.....G..
 91   .....A..C..T..T........C.....G..C.....G..............T..
 92   -....A..T.....A.................C.....G..T..T.....A..T..
 93   .........A.......G...........T.....G...........T........
 94   -------A.A.......G..C........T.....C...........T........
 95   ...........C..............A........G.................G..
 96   ........C..C........C..G.....C..T..C.....C.....T........
 97   ...........T...........T..A.....C.....G..T...........T..
 98   ...........T....................C.....G..............T..
 99   .....C..G..C.....G.....C.....C..A..C..G..C..............
100   ...A.....G.C.....G..C..C.....C..C.....G..C..T...........
101   .....C..TA....T.....T..T..A..T..A........T..T........T..
102   .....C..CT.T..A..G..C........CT.A..C.....T.....T........
103   .....C..TA....T.....T..T..A..T..A..T.....T..T........T..
104   ...........C..............A........G.................G..
105   ...G..   G.T..T.....T..T..A..TA....T....TC..A......A.A..
106   ---------------------........T.....C....................
107   .........G....T.....C..C........C....................T..

           60                  80                   100
cos   C   CGCAAGGTGCTGCGAGATAACATCCAGGGCATCACCAAGCCTGCCATCC
 87   .   ..............C..C..T....................G.....T.
 88   T   .....A........T...........A..T.....G........T....
 89   T   .....A..CT....T..C...........T...........C.......
 90   .   A.G........C..:..C..........................T...G
 91   T   ........C..A.....C..........................A....
 92   T   ........T..A..............A.................A....
 93   .   A.G...........G...........A.C......T.....C.......
 94   .   A.G...........G..........................C.......
 95   .   A.G........TA.G..C...............................
 96   .   .CT..A..A.....C..C..T.....T...........A..A.....T.
 97   T   ........CT....T...........A..A...........A.....T.
 98   T   .........T....T...........A..A...........A.......
 99   .   ..G.....C..C..C..........................G..G....
100   T   ......A.T..T..T..C...........T...........C..T....
101   .   A.A...A.T..AA..........T..A..T.....T.....A..T...A
102   .   ........CT..A....C.....T.......................T.
103   .   A.A...A.T..AA.............A..T..T..T.....A..T...A
104   .   A.G........CA.G..C...............................
105   .TCCA.A...TCTAACAAG.C.TC...TG.A..T..T..T.....C..T...A
106   .   A.G...........G..........................C.......
107   .   ..T.....T..T..C..............A...........C.....T.
```

Figure 9. (cont)

```
                 120                 140
cos   GCCGCCTGGCCCGCCGAGGTGGTGTCAAGCGCATCTCTGGCCTCATCTAC
 87   ..........G..G..C..C..C..............G..G.........
 88   ....TT.......T.....C.....G........A.....A.....A...
 89   .............G..C.....G..............C............
 90   --..-.....G....:T..A..A..............C............
 91   ....A..C----------------------.A........T..T......
 92   .T..A..-..TA.AA.G..A.........A.G........T.........
 93   C...T.........A...C...A..T........................
 94   ..........A...A....G..A..............C............
 95   .......A..A..GA....G..A.......................T..T
 96   .A.....T.:......C..C..C..G...........C............
 97   .T..T..T........T..............T...AG...T.........
 98   .T..T..T........T..............T...AGC..T.........
 99   .G..G.....G..G..G..C..C..G...........G..G.........
100   ....T..C..T..T..T..............T.......C.A.G......
101   .AA.AT.A..TA.AA................T..T.....TT.G......
102   .G..T:.A..T..G..T..C..C..T.....G..............T...
103   .AA.AT.A..TA.AA................T..T.....TT.G......
104   ..........A..GA....G..A.......................T..T
105   .AA.AT.A..TA.AA..........T...A.A..T..CTCTT....T...
106   .......A......A....G...............---------------
107   ....T.....T.....T..C..C..G.....T.....C..G..G......

      160                  180             200
cos   GAGGAGACCCGCGGGGTGCTGAAGGTCTTCCTGGAGAATGTCATCCGGGA
 87   ........G.....C.....C...........T.....C........C..
 88   .....A..G..T..C..T........T...T.......C..A..T..T..
 89   ...........T..............G....................C..
 90   ...........:.:T.....C.....T::............G...A.::.
 91   ........A..A.............-------------------------
 92   ..A.....A.....T..A.............................T..
 93   .....A..T..............A..T..............T.....C..
 94   ........T..............A..G...........C..T........
 95   .....A..T..T.....C..C.....T.....A.................
 96   ........T.................G...T.......C..A........
 97   ..A........T..T..C..C...........T........T.....T..
 98   ..A........T..T..C..C.....T.....T..............C..
 99   ..............C.....C...A.......C.....C........C..
100   ...........T..T..C..C...AC......C...GG.........T..
101   ..A..AGT.A.A.CC..CT....ATC....T....ATCC......A....
102   ..............T........A..G...T..........G..T.....
103   ..A..AGT.A.A.CC..CT....ATC....T....ATCC......A.A..
104   .....A..T..T.....C..C.....T.......................
105   ..C..CT..A.ACAA..CT.....TCT...T.A..A..C..TG.TA.A..
107   ..............T...........G.....T.....C..G.....C..
```

Figure 9. (cont)

```
                 220                 240
cos   CGCCGTCACCTACACCGAGCACGCCAAGAGGAAGACCGTCACCGCCATGG
 87   ....................................G.....G.......
 88   T.....G...----------------------------------------
 89   ...A........................C.C..............T.....
 90   ...G----------------------------------------------
 92   T..A.........TG.........T...C.A.....T.....A.......
 93   .......................................G.....T....
 94   ...G.........................................T....
 95   ...T...................................T..........
 96   ......A.....T..A............C.C.....G.............
 97   T............TG.........T.........................
 98   T...........TTG.............C.A...................
 99   T........................CGCC.C...................
100   T...........................C.C............T..C.C.
101   .T.T..T..T.....T..A...........A.....T..T..TT.TT...
102   ...A........................C.C...........A.......
103   .T.T..T..T.....C..A...........A.....T..T..TT.TT...
104   ...T...................................T..........
105   ...T.....T.....T..A.....T.GA..A..A........T..T....
106   ----------------------------------.....T..........
107   .......................................T..G.......

       260                 280                 300
cos   ATGTGGTCTACGCTCTCAAGCGCCAGGGCCGCACCCTCTACGGCTTCGGAGGTTAA
 87   .C...........G..............A....................C......
 89   ........G...........................................C..G
 91   --------------------------.....A..A..G...........C..C...
 92   .C.....G..T..A..A...A.G.....T..T..AT.G...........C..C...
 93   ..........T........A..T...........T........T............
 94   .......G..T.....G........A..A.....T..G.....A............
 95   .......G..........................T..............C..A...
 96   .............G.............................T.....T....G.
 97   .C..C........C..G...A.G.....A..T...........A............
 98   .C..C........C..G...A.G.....A..T...........A............
 99   .C..C........G...................................C..C...
100   .C..T........C.................T...........T.....T......
101   ....T..T..T...T.G...A.A..A..TA.A...T.A..T..T.....G......
102   .......G.....G........:.....GA.:....................C..G
103   ....T..T..T...T.G...A.A..A..TA.A...T.A..C..T.....T......
104   .......G..........................T..............C..A...
105   .C..T........C......A.A..A...A.A..T.....T..T.....T....G.
106   .......G..T.....G...........A.....T..G.....A............
107   .C..............G..A........A........G.....T.....C......
```

Figure 10A. Histone H1 5' Upstream sequences

```
          -200                -180                -160                -140
1 TCGTGTTGGCGGAATTGTAGAAAAAACGCGCTTTTTCGCCTGTTAAGTTAAGAAACACAAAATAGCGGGAG
2    GGGGCGATTTGGTGGCAGAAATTCCGAGGAAAATACACTTTTGTTAGTCCAAAGAACACAAATCGAGCA
4 AACCCCCCATTTTCCAGACTCTAAAAACACAGACTCCCACCTGCCCATAGTTTCCACCAGCGAGAATTAATG
5 AACCCCCCATTTTCCAGACTCTAAAAACACAGACTCCCACCTGCCCATAGTTTCCACCAGCGAGAATTAATG
7         CAAAGTTTTCTGTTATAATCATCTGCTTGTCAGCAATTTCACTCAAGTTTCGACCTGCAAACAC

                  -120                -100                -80
1 CGAAGGGAGCTCTGCGCCGTGCGGCGGGGCGGGCTCTGCAGCGCACCAATCACCGCGCGGCTCCGCTCTATA
2 CACCGAAGGGCTCCCCGGCCGTGCAGCGGGGCGGGTTAGCAACGCACCAATCACCGCGCGGCTCCTCTTCTA
4 GGCGGGGTTTGCTTCTCAACCAATGAAGGTTTAACTATAAAAGCTGGCCAATCAGGGACCCAGAAAAGATGA
5 GGCGGGGTTTGCATCTCAACCAATAAAGGTTTAACTATAAAAGCTGGCCAATCAGGGACCCAGAAAAGCTGA
6                              CGGACGACCCGGGACTGTCTCCTCCCACGTACGCAACAATGCC
7 ACGCTGATCGGCAGTGAGGCAAAACAGACAAATCGGACAAATTGTTTCCCACCACGTACGCAACCGCGCGGG
8                                                            CCCGGGCCCGAGC

       -60                 -40                 -20
1 AATACGAGGCCGCCGACTTGCTCCGGGCCCAGTGGTTCCCCCCGATCTGTGGAACGACGTCCGTCAC ATG
2 AAAATACGAGCATCTGACCCGCGCCAGCCCAATTGTGTTCGCCTGCTCCGCAGAGGACTGCGCCGCG ATG
4 ATATAAGGAGGTTCGATATAAACTGAAAGTTTAGATTTTAGTTCGCGTGAGTGAATTTACTTCAAAG ATG
5 ATATAAGGAGGTTCGAGACAAACAAGAACTTTAGATTTTAGTTTGCGTTAGTGAATTTACTTCGAAA ATG
6 TTATATTGAGCGTTGCCGAGCCGATGGTTATTCGTTTTGTTAACCTCCCGACGCACCGTATATCAAG ATG
7 ATATAGGTGAGGTTGCCGTGAGGGCCGTCACTTGTTTTGTTAACTCCGCTACGCAACGTTTACCAAG ATG
8 ATAGCAGCAACGCAAAACCTGCTCTTTAGATTTCGAGCTTATTCTCTTCTAGCAGTTTCTTGCCACC ATG
```

Figure 10B. Histone H1 3' Downstream Alignments

```
                         20                  40                  60
1  TAA GATGACAGAAGAAATTCGAGTCTGCTCATTTAAAAACCCCAAAGGCTCTTTTAAGAGCCACCCAT
2  TAA GTTATCCCAGAAGAGTCCTGCTCTACCTATTTTGATATCCAACGGCTCTTTTAAGAGCCACCCAC
4  TAA CTTGGTTCCGGTGTCCCTGTCCATCCCCCCCCAACCTCAAAGGCTCTTTTCAGAGCCACCACCAA
5  TAA CTTGGTTCCGGTGTCCCTGACTATCCCCCCCCAACCTCAAAGGCTCTTTTCAGAGCCACCACCAA
6  TGA TTTGCACGCCAACTTTCCCATCCTACCAAAACGGCTCTTTTCAGAGCCACCACATACCC
7  TGA TGTTGCACGTCCTACTCGTGTCACCACAACACAAACGGCTCTTTTCAGAGCCACCACATTTCCAC
9  TAA ATTGTGAAAAAGTGCAGTATTTGGTACATGTTCGCAATTAAAATTTTAGATTTATGATTTAGTAG

                80                 100                 120
1  TTATTCTCAGAAAGAGCTGGAATGCTGCGGGAACCGCGGCAGCACAACTAATTATCTCAGTTGCAGAGAT
2  ACTTTCCCTAAAGGAGCTGAGGCACCGAGGTCGTCAGAAACTTCCAGCACGGAGGCAGCAATTCGTAAGT
4  CTCAGTGAGAAGAGCCGATACTGTGACTGCTCATTGGCTCCATTATGTCAGGGAGGCTTTCTGGTATCTG
5  CTCCGTCAGAAGAGCCGATACTGTGACTGCTCATTATGTCAGGGAGGCTTTCTGGTATCTGCTGAATTGT
7  GTAAGACC
9  ATCTGTAAATTTGTTTAAACAAGTCCTTTTCAGGGCTACAACGTTCCGTTGCAAGAGAAAAAAACTTTTA
```

Figure 11A. H2A 5' Upstream Sequences

```
        200                 180                 160                 140
18   CTAAAGGAATGAGAAAGCTGTACTCCACTACATACTCTGGTGTACTCTGGCTCAGTTCTTGGACTCCTCT
19   GGCCTCGAGAACACGCCTTTCCTCCCTTATAACTGCTTTTCATTGGTTCAAATTCGATTCGCTTTCTCAT
25   GATGGCGGTATTCATGAAATCGAGGTGCGGGCAGCGTCCCGCTGATTGGACAATTGTCACAATGCCCTCG
26                                                     CCGTTGGTGCCGAATACACT
27   ATACTAGCAATAAGTCTTGACCTAAAAAATATATAAATAAGACTCCTAATCAGCTTGTAGATTTTCTGGT
28   TCAAACATAACTCTATATATAAGGGATGAAGATGTATGCTTTCTTAGAATTTCAAACATATTCCGTTAAA
29   ATTTTCACAAACACAATTCACTTATCGAATGGGCCGAACGCGTTCACGTTTATACTTTTTTCGAGCAGT
30   CTCGTCACTCTGCTGCCTGCCTGCCTGCGGGTGGGTGGGGGCGGGCTTGGCTGACTGCCTGTCTGTCTGT

                  120                 100                  80
18   TTTCTTGGCGAACTCAACTGGTATGAATTCCTCACAGCCTACCTCCAGTCAGTATAAATACTTCTCTGCC
21   TGGCTGCCCAGAGCGGACCCAGACGTCAGCCCATCAGCGCAGAGGCGAGGAGCCAAAGCGAGACGTAGCG
22                                                                    ATTGT
24                                                                 TCCATTCA
25   CTGACCGGTCTCTCCGATCCCGACGTTTGGTATAAATAGCCAGCAAAAAAGATAGGTGGTCAACCATTCA
26   GTGTTGGCTGGTCGAACTCATCCAATTAAAGAAGAGAGGTGTGTCCTGCGCTGCCTATAAATATCAGTAA
27   CTTGTTGAACCATCATCTATTTACTTCCAATCTGTACTTCTCTTCTTGATACTACATCATCATACGGATT
28   GTTTTACTTTTCGATTTCAATTTCGACTGCATGATGCTTTTCTTAGGTAGTTTTTTGTTATTAAATAGTA
29   CAATTCAGGTCTAAGTCACCCACCCCTAACTGAATGCGCGGGCAAACGGAAAAGTATAAATATTTCGCTG
30   CCCCTCCCTCACTCCAATGGATAGGCCACACCGGTCCGGTGGCCAATCGATGGCTTTTGTGGCGAGGTAT

         60                  40                  20
18   TTGCGTTCTAATGTAGTTTCATTACATTTTCTTGTGGCGATTTTCCCTTCTTATCAGAAGTAGTT ATG
21   GAGCCCGTAGGTTGCGCGCTGCGTTCTTGGCTTGTTTGCTCTCCTGAGTGTGTTCAGTCGCTGCG ATG
22   CTGGCGGCTCCTGAGGCGGTTCGAGCGGCGGATTCGGGGACCGGGGCTCGGGCGGCGGCGGCACC ATG
24   AGTCATCGAACATTGTTACGTTCTGAACTTCGTCTTCCGATTTATTCTAACTCATCAACAACATC ATG
25   AGCCAGCGCACATCGCTTCGTTCACAACCTCGCTTCGCTCTCAGCTCGTTAACCAACCAACCATC ATG
26   GTAGGGGAGTGCAGCTTCAGTCTACAACATCTTCTTGATTGTGGTTGATTTGTAGCACAGTAATC ATG
27   TGGTTATTTCTCAGTGAATAAACAACTTCAAAACAAAACAAAATTTCATACATATAAAATATAAA ATG
28   TCATAAATTCTTGCCTTTTACATAAGAATTAGGAAAGTACAGAACAAGAGCAAATTTAATATATA ATG
29   TCGGGGTTAGGCGAGCATTCGTGTTCGTGTGTAAAGTGAACTAAGTGAAATAAACGCTGAGCAAA ATG
30   AAGTAAGGCTCTCGAGGTGCCCAGCGGCTCATTCAGACTTTCTGTCACATACTGAAGCTACCAAT ATG
```

Figure 11B. Histone H2A 3' Downstream Sequences

```
                           20                  40                   60
18   TGA AGAGTTAACGCTTCATGCACTGCTGTTTTTCTGTCAGCAGACAAAATCAGCCTAACAGCAAGGCT
19   TAA GCCAGTGAGCTAAGTTTTTTTTTTTTTTTTTTTTTTTTTTTAAACAAAACCCAAGGCTCTTTTCA
21   TGA GCACCGGCGAGGCAGCGCTGTCTGAGAGAACAGTCCAAGCTCTTTTCAGAGCCACCCACAGCATC
22   TAG AGGACGGGGGGTCCCACCCGGCGCCCGTCGGCGGCCCAACGCGGCCCAACGCGGCCGGGCTGCGC
24   TGA AACCTCAACGGCCCTTATCAGGGGCCACCAATTACTCACGAAAGAATTGTTTCATTTATGAATTC
25   TAA ATTTGTTTGCTACCTCTTGCAACCTCAACAACGGCCCTTATCAGGGCCACCAAATATTCAAGAAA
26   TGA AATGCCCAGCTTCCCAGCGCCCCCATCAGGGACAACACAAGGGCTCTTTTCAGAGCCGCCACACC
27   TAA GATCGGTTCTGGTATTTTAAAGAAGGCGGAAGGAATTAAATTGCTTCTCCATTGT
28   TAG TATTACTATTAATAAAGAAAATAAAAAACGTATCAAGTTTACGG
30   TAA AATCGCTGGTGCGGCTGCAACCTGACTACTCAACCCCCAAAGGCTCTTTTAAGAGCCAACCACCT

        80                 100                 120                 140
18   CTTTTCAGAGCCACCTACGACCTTCCATTAAATGAGCTGTTGTGCTTTGGATTATGCCGCCCATAAAG
19   GAGCCACCACTTCTTCATATAAGAG
21   GCAGGAGAGCTCAGAAATCCGCAATACAGTCGTGCAGGTCTATGAATTACTCGA
22   GGCGCCGCTCCGCTCCGCGGAGGGAGGAAACGTTTCTATGGCTTTTGTGTTCCCCGCCGTGGGCGGCGG
24   CCTTCCCCCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCC
25   GATAAAAGTCTCTG
26   CGCAAATCAGAGCTCACGTGATCACATGGGATTACGAGGAGAGATTTGTAATAAGAGAATGAATAGCGC
30   AGCTCAACAAAAGCGCAAAGTGTCC
```

Figure 12A. Histone H2B 5' Upstream Sequences

```
               -200                -180                -160
42  AGTTTAATTTCAACCAATAGTAGTGCGTCTTCTGGATTTGCGAATCCTGATTGGGCAGACCTGACCTCTGACG
43                                             TGATACTTACGCAACTAACCACTGAGCGAA
44  CAGCCAATGAGAAAGCGAATCGAATTTGAACCAATGAAAAGCAGTTATAAGGGAGGAAAGGCGTGTTCTCGAG
48  CTCTAGGCAGGGGATGGAACAGGCACTAANCTGCGACGCCTAAGACCAATGAAAGGATCGAGACCGAGGCTCA
51                 ACCTTGGTTGGTTATCTTGAACGATTGGTAAGAAAGGGGCATCTCTGTTTTCTTGATG
52                                        GTAACTGTTATTTGAGCATAACACAGGTTTTTAAA

   -140               -120                -100                 -80
42  TTACCCTGAATAACTACCAATCAGACACAAGACTTCAACTCTTCACCTTATTTGCATAAGCGATTCTATATAA
43  TATGCTTCCTTGATGGACAGTTAGTGCTTGACGTTTGCAGACTCTCTGACAAGGACAGCCACCGCTTTATTTA
44  TCCGACCAATGAAAGAGTGCGAAAGGAATGCTTCTCATTTGCATAGAGGGGCTATAAATAAATGCCTACGACC
45                      CTGTTATCCAATCAGAGAGCAGATACAGAAGGCACTCGATTTGCATACTGCCC
46     GCGACACGTCACGGACTCTGTTATCCAATCAGAGAGCAGATACAGAAGGCACTCGATTTGCATACTGCCC
48  TTTGCATACGGACCGCAGCATACGGATCCGGCCCCGTGTATAAAAAGGAAAGGTTCTCGCTGGCCATTCACAG
50        TTATACCATGTGACAAAACCTACCAGTAATATTACAAGATATCGGACTGCCTTATTTGCATGGGAAG
51  TATATAAACAACATGATTTGATCATCTCAAGATGGTCAGATTTATTAAAGACGTTTCTCTTTCCGCATTTTTG
52  ATTATTATATATCATGGTATATGTGTAAAATTTTTTTACTGACTAATTTTGTTTATTTATTTAGCTTTTAAAA
53  ACTTTTCCGTTTGCCCGCGCATTCAGTTAGGGGTGGGTGACTTAGACCTGAATTGACTGCTCGAAAAAAAGTA

          -60                 -40               -20
42  AAGCGCCTTGTCATACCCTGCTCACGCTGTTTTTCCTTTTCGTTGGCGCTTTATAGCTACACAGTGCT ATG
43  AAAGAGCAGGAAAGGAACGGAACAGTTCAATATCTCTTTTCTTGGCCTACCTTCATTCTCTGTTCACT ATG
44  CCTTCGTTTCCATTCAGCGTCTCCTGGTCGTTTTTGTTCGCCTCGCTTCGTGAGCGCGTTGTGCCACT ATG
45  CTATAAATAGGCGAGCAGTGCTCGCAGCCGGCACTCCGCTGCGCCGAAGGGATCGTGGAGAGTTCGAC ATG
46  CTATAAATAGGCGAGCAGTGCTCGCAGCCGGCACTCCGCTGCGCCGAAGGGATCGTGGAGAGTTCGAC ATG
48  TATCCAAAGAATATTGCTTGACATACTCGTTTCGCTGCATCTTTACAGACCAGAAAACCTCAATTCAT ATG
50  GCTATAAAAGCAGGAGCCCGGGAGGCGAAGGAAACAGTTTTGTAGGCTGAGAGAGAAGCAGCACAATT ATG
51  CATTATTGTTATATTAAATTTATCCTATATAGACAAGTCAAACCACAAATAAACCATACACACATACA ATG
52  ATTTTACTTTCTTGTTAATTTTTCTGATTGCTCTATACTCAAACCAACAACAACTTACTCTACAAATA ATG
53  TAAACGTGAACGCGTTCGGCCCATTCGATAAGTGAATTGTGTTTGTGAAAATATAAGTAACGTGAACA ATG
```

Figure 12B. H2B 3' Downstream Sequences

```
                         20                  40                  60
42  TAA ACAGTGAGTTGGTTGCAAACTCTCAACCCTAACGGCTCTTTTAAGAGCCACCCAATGTTCTCAAAGA
43  TGA GTGCTCAAGACTCAGCTCTTAACCCAAAGGCTCTTTTCAGAGCCACTCAAGACTTCAAAATTGGAGC
44  TAA GCTGTTTTCATCCCCCGTGCCAGAGTCGCTCGACCCAAAGGCTTCAGAGCCACCCACCTTGCCAGAG
45  TAG AGCGGTGCGGATTACTCGATTTTAACCCAAAGGCTCTTTTCAGAGCCACCATTTGTTCTAATAAAAG
46  TAG AGCGGTGCGGATTACTGATTTTAACCCAAGGCTCTTTTCGGAGCCACCATTTGTTCTAATAAAAGGG
48  TAG ACAGGTCATATCCTGCTCTAATTGGACAATAATACAACGGCCNNACAAATAATCAAGAAAGAATGAT
49  TAA ACGGTTACACNNNNNNNNNNNNNNNNNNNNGGCCACCAAACATCCAAGAAAGAATTGTGTCATAAAT
50  TAA TTGCTGCTGCCCGACCCCTGTCCGACTCCAACACAAAGGCTCTTTTCAGACCACCCATCTTCTCCCG
51  TAA TGAAATCACTTC
52  TAA GTCACTCACTAG

               80                 100                  120
42  AAGAGCTGGTGCTTGTATTTCCTCCTC
43  TTTAATGCTACCAAGCGACTTAGTGACTACCGGGAAAATAACCGACTTCATGCAGGATGTGTGACAACAC
44  AAAGAGCTGTAGTCATATCAGAAGAAAATAGCATAAGTTAATCCGGCTTTCTACTTCACATCTTTAAATAC
45  GGCTGTATTACTTTTTTTCTTTTTTTCCTGAGGGGTATAGCGTGGGTTAACTGAGTGAATGGAAAGCGAGT
46  CTGTATTACTTTTTTTCTTTTTTTCCTGAGGGGTATAGCGTGGGTTAACTGAGTGAATGGAAAGCGAGTGC
48  ATCCGTAGTAATGTAGCGTAGTTGTTTATTAAATAAATAATGTTGTATAACGAATATTAAGATATAATCAA
49  CAGATGAAGAGAGTAGTCAGCTATTACATAACACATAAGGGGTGCAAATTACATATAAGCACCACCATGAA
50  AAAAGATCT
```

Figure 13A. Histone H3 5' Upstream Sequences

```
                                    -200                -180
60                               ACGGTAATGACAGGAATCTCTCTTAATCTGCAACTAGGCACAGA
61                                                 TCTTTGCAACCTGGGACAGGACAGGA
63                                        CTCAGCCAATAGGACTACTGCGCGGGACACTTGAAAA
64                                                                   CACAAATTTG
69  AAGCGCGCCTCGGCGAGGGTGGGGGTTGCATTTTGGGGCGGCACGGGAGGCCGAGCCTCCCGTCCAATGGGGC
70  TCGACACATTTTATTGACTTTGCACATACGGCATCGCCAAGCCCCCTTCCCGTCACGCGCTAAACAAAAGAGC
74  AGCAAAATAAGATTCAATTTTTCGTGTTGGAAGCGAACAGATGGACCGCATGACTAATGCGTAATGCATGAGG
75  CTGAAAATGGTAAATTAATTCCTAGATTCATGTCAGACGACACGCAGACGATGGTTGCCAAGATCGTGAGTGG
76  TATTTATAGGTAAAACGACAAAAACCCGAGAGTACGAACGATATGTTCGTTCGCTTTTCGCTCGTCAAATGAA
77  GCATTCCTACCCGGATCGCCATCTCGACCGTCCACTCCATCCGCATCCAACGGCAAGTACATGCCTCTCGTTC

   -140                 -120               -100                 -80
60  GATGGGCCAATCCAAGAAGGGCGCGGGGATTTTTGAATTTCTTGGGTCCAATAGTTGGTGGTCTGACTCTAT
61  GGACAGGCAGAAGGCTTAGAGTTAGCCGGTTAAATTCATTGATTTATTGACCAATGAGAGGCGAATGGGCGG
62                                   TGTTCGCAGCCGCCGCCGCGCCGCCGTCGCTCTCCAACG
63  GCAGACACGCCTATCAGGATGCTTTCTCGGTGGGAAGGAGGGGTACGAGCGCGGGTACGTGTGTTGCGCGTGT
64  AAGTTGAGACCTGTTATCCAATTACCAAGTACTTCCGCATACATGATCATAGGCATTTGAAGATTTCAACCAA
69  GGAGGAGGCCTCCGCAACGGGCCAATCA GGGCGGTGCGGGGATGGTGACCAATGAGCAGACGCCGCTGCCGG
70  AACCCGGTTGACCAATCAAGAGAGCTTTACAAACGGCACCAGGATCCCGCAGCACATATAAATAGCTGAAAAT
71                                                                  TTGCGCCG
74  GGCAGCGATCCGTAGAGGAAGCGACCAATGAGGAGTGATCTGAACCTCTATTCACTGGTATAAGCCAAAAGCG
75  AACATCATGAATGGTCAACTCGAGTGAACCAATGGGACTGGACAGTTTCACTATCCAATCAACGCGCACGATA
76  ATGGCCTCTGTTTTTCTCTCTCTCTCTCTCCTCTTTCACCGTCCACGATTGCTATATAAGTAGGTAGCAAATG
77  CGAAAAACCAAAACCCAAACCCGCCCGCTCTCTCGGTGTCCCTCCTATTTAACTCCGCCCCGTCCCCTTCTTC
78                                                                TCGACCACCA
79                                           ACTGTGTATTCTTCGGGATACATCTCTTT
80                                           TTCTTGTTCTTTTATATAGGACCACTGTT
81                                    GATCTCTCGGAGCTCTTATAAATAGCGGTCATATTT

            -60                -40                 -20
60  AAAAGAAGAGTAGCTCTTTCCTTTCCTCCACAGAACGTCTCTGGAGGGAAGCTTTTCTGTGGTTTTGCC ATG
61  GTTTCATCTACTATAAATAAGAGCCGTGCAACGAGACCGCCTACTTTCGGTTGCAGAGCAGTTCTGCGA ATG
62  CCAGCGCCGCCTCTCGCTCGCCGAGCTCCAGCCGAAGAGAAGGGGGGTAAGTAAGGAGGTCTCTGTACC ATG
63  GCGACGCAAGCGTACTAAAGGCCAAAGTGCGCTACTTAGGTATCTCACTTTTCCCTACGGTTACTTGCC ATG
64  TCAGGAGCATGTTCCTTCTTATAAAGGAACCCAGAACCTAACCTCTGCATTTCCTATTTCTTTGTAGAA ATG
69  CCTTATAAACTTCACATAGGCATTTTGAGGCTATACTCCGACTGTGAAAGAAGGAAGCTAGCTAGCGCC ATG
70  TGCCAGTGGTTCTCATTCATCCCGTCACTCGTATTTGAAGTACTGAATTACCTACTGTCCCAAGCAACT ATG
71  TGCGACACCTCTTCATCTCGTACACCCTACGTTTGAAACACACTGANTCCAACTGTCTCCCAATCAACC ATG
74  CAGTCTGGAGGTATAAATACGTCGCGGTTACTTTGAAAATTTATCAGTTGACTATCTGAAAAATCAACA ATG
75  CCATAGCTATAAAGATCAGCCGGGAATTTCGAAAATCAGTTTTCGTCTCATCAACGATCAATCAACAAA ATG
76  CTCTGATCGTTTATTGTGTTTTACAACGTGAAGTAGTGAACGTGAACTTTAGTGAAACCCAAATCGGAG ATG
77  CCTCCTCACCCCAATCTCCCCAGCCCCCAACCACCAGAGCCTCCTCCTCCCCTCCGGCGCATCGACGAC ATG
78  GCGCAGGTCATCACTTAATCATCGCTCATCCATCAAAAAACACGTTATCAACACACATAAACCATCACA ATG
79  CCTCAACCTTTATATTCTTTCTTTCTAGTTAATAAGAAAAACATCTAACATAAATATATAAACGCAAAC ATG
80  TTGTGACTTCACTTTGGCCCTTCCAACTGTTCTTCCCCTTTTACTAAAGGATCCAAGCAAACACTCCAC ATG
81  TCAGCGGCAGTCACTCATCTCGTCTTCTACGTTCGATTCGCAGACTTACCATCAGACAAGTAACCAACC ATG
```

Figure 13B. Histone H3 3' Downstream Sequences

```
                           20                  40                  60
60  TAA ATGTAAAGTTACTTTTTCATCAGTCTTAAAACCCAAAGGCTCTTTTCAGAGCCACCCACTTATTCCAA
62  TAA GAATCCACTATGATGGGAAACATTTCATTCTCAAAAAAAAAAAAAAAATTTCTCTTCTTCCTGTTATT
63  TAA GGGTTTCTGTTAATCCACACAACCACTTTAAAGGCTCTTCTTAGAGCCACCCATCTTCCAAAAAAAGA
64  TAA TAGGTACGCTTTCTACACTGGCACGTAAACCAAAACGGCTCTTTTAAGAGCCACCTCCATTATCCACC
66  TAA ACTTCACCTTGGTGGGTTTGTCATTCTCGAGCAAACTTCTTCCGTTATTGGTAGTAATGAACGTTAGA
69  TAA ACGATGACCTGATCTCCAAAATCCCCCAAAGGCTCTTTTAAGAGCCACCTCCATATTTCAGTCAAAAA
70  TAG AACCATCGGTACAGCATGTAGCCCATGCACCGCATACACAAACGGCTCTTTTCAGAGCCACCACAACC
71  TAG ATC----------------------------AATCTCGTTTAAATAGTTATATATTACATTATATGGT
74  TAA ATTGTCTTCTTTGACTGACACAATATAAACGGCTCTTTCGAGACCACCAAAAGCTCAAGAAAGAATCA
75  TAA GTCGTTTAACCTTGTTTGAACGCAAAACCGGCTCTTTTCAGAGCCACTAAATTTACAAGAAAGACACA
76  TAA GCTGACACGGCATTAACTTCGAGATAAAGCGCTAGCGTACTCTATAATCGGTCCTTTTCAGGACCACA
77  TAG GCTGCTGCACCTGCAATCCGTGTTAGGCTGAGTTCATCGGTGGAAAATAGTGGTGTTTCAGAATGTGC
78  TAA GCGACTCTTCGATATGGAGTAGTTTGCTTTGGGTTTTCGGGGTAGTCTAGTCAGATTCTGGGGTTATA
79  TAG TTTGTTGATTGTCATCAGTTTTAGTAAAAAACGAACAAAAACACAATAAAATATAAATCAATATATTT
80  TGA ATATAAAGCGGGAATTTTTTTTTTTCTATGCATTTAGACTGGGGGACATCATACAAACATCCTTTTT
81  TAG AGGCAATGCACTACAGCCTGTAGGGCATATAACCAAAACGGCTCTTTTCAGAGCCACCATAACATCAC

                80                  100                 120                 140
60   CGAAAGTAGCTGTGATAATTTTTTGTTGTCTCAA
62   GGTAGTTCTGAACGTTAGATATTTTTTTTTCCATGGGGTCAAAGGTACCTAAGTATATGATTGCGAGTGGAAAA
63   ACTGTGCGCTTTTTCCAAACTTGTGGGTATTAATCAGTTTCATTTGTCAAAAGTGCTAGGTCTCCTAGGGGAC
64   AAAGATGCTTGAAGTACAAGTTGTGAGAGTTTTCTAGGGTTTCCTATTATAGCCTTTCTTGACAATGTGAGC
66   TATTTTTTTCCAATGGGGTTGAAAGGTACCTAAGTATATGGTTGCAAATGGAAAARAAGGGGTCAGAATTGGGT
69  GACACAATTGTTCCATTTGTACACGCCCCTTTCCCACCGTGTATGTTCCCTGTTCTCGAGTCACTACAGACCG
70  CCAAGAAAGAXTCATTGATTTCTAAAATGAACACGATGATGCATGACAATCGACGCTATTATAAGAAGAAAGA
71  ACTATCCTTAAACATACTAGTAATAAATTTATAAATTACAGTCATTCACGACCACGATAAGTACATCCGTACG
74  AATGATCGATCTTCTTTGTGTTGTCTTTGTGTTTTTATATCCTTTTTCGCACTGTAATTACCAGTTCATTTTA
75  ATCTGATTTTTATATAGTGTTCATAAATAACTTATTCTTGAAATGATGGGCATCATAATAAATGTCAATCAGT
76  AAACCAGATTCAATGAGATAAATTTTGCGTTGCCGACTATTTATAACTTAAAAAAAATTAGAACAAAATTCCA
77  GTCGTGTTAGATCCATGTTAGATCCCTGTTCCTCTCATGGTGGGTGGATGTGTTATGCCTAATCTGATGGTAC
78  TCACGGGGCGGTCATGACAACTTTTCTCGCGTCACACTTCGCACTACATTTGGG
81  GAAAGAATCACTGATATTATATGCAGTATATATGAATAATTGGTAAAAAATAATGAATCCCTGGGGTAATTGA
```

Figure 14A. Histone H4 5' Upstream Sequences

```
               -200               -180               -160               -140
 87  AAACAGACTGTAATAGGAAATGAACGCATTGCAGCCAGGAGAAAAAAAACAATTAGCTAGGAGGAGGCCCCG
 89  AAAAAAATAAAGGCTTGAGACTGTAAGGAACCGGTAGAGGGCAGAGAAGAGAAAAGAAAAACAGGAAGATGA
 91  GTCAGTTGTGTGCGCGGCCTCCAGTGAGCTACCACCGGGCCGTCGCGGAGGGGCGCACCTGTGCGGGAGGGG
 92  GCTAGCGAATACTCGCCACAAGGGGGCGCAACTCGAATGGGAGTCTCCGCACTCCAGTCCCGCATACCGTA
 99  TCTCGGCCACGTCACCGATCCGCGGCATCTCTCCCCCGGATCGCCGTCTCGACCGTCCACTCCATCCGCATC
101                                GCTGTTGTTATTCGGCTAGATACATACGTGTTTGTGCGTATG
102  TTAAAAAAAAAAAAGAGTGAGAGAGGGACTGAGCAGAGTGGAGGAGGAGGGAGAGGAAAACAGAAAAGAAAT
104  CTCCGCCTCTATGGAGTGAGTTCTCTGTCAGGTCCTCTCCAGCTGCATATAAAGAGGTGGAGCGGTCCTGAA
107  GTGTTTCGCCGCGCATGCAAGTAGCTTTGATTTCTTCTTCCTCCATAGTTCCCTAGAACGAACGCCCTCCCC

                     -120                -100                 -80
 87  CCCTCCAGGAGGAGGAGGCAAGTGGCGCTCCCGCCCCCGCCTGGGGCCCCGCCCCTGGTTTCAATCGGTCCG
 88  GCGAAAAGCGAACGAACATATCGTTCGTACTCTCTCGGGTTTTTGTCGTTTTACCTATAAATAGGGGCACAG
 89  TGCAACATCCAGAGCCCGGATAATTTAGAAAGGTTCCCGCCCGCGCGCTTTCAGTTTTCAATCTGGTCCGAT
 91  TCATCGGAGGGCGATCGAGCCTCGTCATCCAAGTCCGCATACGGGTGACAATACCCCCGCTCACCGGGAGGG
 92  ACGATGCCGCAATCTCGGTCACCCAAGTCCGCAATGGTGTAACAATACTCGGTGCAATCCGGTTGAGGCATC
 96                 GCGGGACTCGGGACTTCCCGCCGACTTCTTTCAGGTTCTCGTTCGGTCCGCCAACTGT
 97  CCCTCTACGTTTGATGAACAGAGAGAGCTCATTGATCCAGAGAGCATTACTATTCCTTCGACGTCCGGAATA
 98  TTGGTTTTGTGACGGGTTGCTGTATTCGCGGAACAAAGGAATTAGTCGAGTGGTCACTCACGTTCGCATAAA
 99  CAACGGCAGCCACACGCCTCCTCCAACCTCTCGACCCCTTTAAGACGCCCTTCGCCCCACCCAGCAAATCAC
100  AACCGCGTTCCCGTACTTAAAGTCCGCTCTCTTCCCCGCCTTCTTCCCACCTTCCACGAACGACTTCCAATC
101  TAGTTATATCATATATAAGTATATTAGGATGAGGCGGTGAAAGAGATTTTTTTTTTTCGCTTAATTTATTCT
102  GACGAAATGTCGAGAGGGCGGGGACAATTGAGAACGCTTCCCGCCGGCGGCTTTCGGTTTTCAATCTGGTCC
103            AAAAGGTCTAATTCTTTTTCCTATAAATACCGAGATATTTTTTCTATATGATGGTTTCCGTC
104  CAATCAGAAATTACATGACGAACACACTGATATACTCCGCCTCTATGGAGTGAGTTCTCTGTCAGGTCCTCT
105  AAATTAGATTAATTATTAATCATGATTTGAATAGGATAGCAAGAATATTTGTTGGTTTAAACGGGTAATTAT
107  TTTGCAGCAGAGGCCAGGCAAAAGCGCGCTTTCTGTGCCACTGGGCTCAATCAGGTCCACCAAACGCAGA

          -60                 -40                 -20
 87  ACCATACGCATAACACCCGGCGCGCGCCCCGCCACATCCTCACTGGGTGTCGACTCAGGCTCTCGGC ATG
 88  AAACGTTGAAATTAGTTCTTTAGTGACTTTCGTGCTGTGCGTGTATAATAGTATAGAACAGTGAAAA ATG
 89  CCTCTCATATATTAGTGGCACTCCACCTCCAATGCCTCACCAGCTGGTGTTTCAGATTACATTAGCT ATG
 91  TTGGTCAATCGCTCAGCGAAACGTCCAGTCGTCAGCATCGCACTAAGACTCTCTCTCAATCTCCATA ATG
 92  ATTCGCTTAGCGTAATATCCAGTCTACAGGATCACACAGAACTCGCTCTCAACTATCAATCATCATC ATG
 93                                                            GACTATGTC ATG
 96  CGTATTAAAGGCCTGCCTCAGGTCAGAGGCCACAAAGCGTTGGGTGAGACGCCTCTTGCTCGTCGTC ATG
 97  GGGGTATATATACCGGTCGAAATCCTATCACGTCAATTCGACTCGTCGATACCAAATAAATCTAATC ATG
 98  GGCTATATATACCGCACGAACAGCAGAATTGAGTATCAGTTTGAATCTCAAACAGGAAACTATCAAC ATG
 99  AGCACCAGACGCCACCCACCACCGTTCCTCCCATCCCACACTCGCTCGCAGCTCGAGATCGTCGGCC ATG
100  TCCCAAACGACCGCTTCAGCCTTACCACCTACACCACTTTCACTTTCACCCCCCAACACATATCAAA ATG
101  TTTCTCTATCTTTTTTCTACATCTTGTTCAAAAGAGTAGCAAAAACAACAATCAATACAATAAAATA ATG
102  GATATCTCTGTATATTACGGGGAAGACGGTGACGCTCCGATCGANCNNCTATCGGGCTCCTGCGGTC ATG
103  GCATTATTGTACTCTATAGTACTAAAGCAACAAACAAAAACAAGCAACAAATATAATATAGTAAAAT ATG
104  CCAGCTGCATATAAAGAGGAGGAGAGGCCCTGATACGTTATATTGTGTTTCAAGAGCTCAAGAAAGA ATG
105  CAAAAATTTATAAATAATTTAAAACAATAAATAGAAAAACAAATAAGATTATAAAAACTTACAAAA ATG
107  TAATAAGAGCACCGAGCACCGAAAATTAACGTCATTAATTCACTTGAACTCAGCTAGTGAACACACG ATG
```

Figure 14B. Histone H4 3' Downstream sequences

```
                           20                  40                   60
 87 TAA ACTCGTCTCCGATTCCGGCCACCCGAACTCGTTTTTAGCAACCCAAAGGCTCTTTTCAGAGCCGCCC
 89 TAG ACGCCGCCGCTTCAATCCCCCCCCCCCCCCCATCCCTAACGGCCCTTTTTAGGGCCAACCACAGTCTC
 91 TAA GTGAAGCAGACTTGGCTAGAATAACG
 92 TAA GTGTAGCAGACCTGCTAGAATAACAAACGGCTCTTTTCAGAGCCACCAAATAATCAAGAAAGAATAC
 93 TAA GCCATCGCTCCTCTTTTCCTATCACAACGGCCCTTTTAAGGGCCACA
 94 TAA AGGCTCGCTTCTGTTTCCTCATCAAAACGGCCCTTTTAAGGGCCACCA
 95 TAA ATCTCTTCCTTCCCACTAATCTTATACCCAAGGCTCTTTTCAGAGCCCCCCAGCTCTCACTCCTAGA
 96 TGA GCGTCCTCTCTACCAATAAGAGGCCTTTTCAGGGCCCCTACTTCCTCAGCTGAAGGTGGTAACACTG
 97 TAA GAAGTTCATCTCTTCAAACCCATCAACAAACCGGCTCTTTTCAGAGCCACCAAATATTCAAGAAAGA
 98 TAA GAAGTTACTTCTAACCAACGGCTAAACGGCTCTTTTTAGAACCACCAAATACTCAAAGAAAGAATCA
 99 TAA GGGCCGGCCGGCCGACGGGAGTCACTCTTTGTCGCCGCCTGCAGATTCCAGAAGCCTGATGAAGCCC
100 TAA ATGTCTCGCCGTCATTAGCAGCAGCGTGTCTTTTTTCTTTTTTCTTCACCATAACGACGACGAATAA
101 TAA ACAATCGGTGTTTGAAATTATTTTCATGCCTTTCAAAAAATAAAATAACGATCATTCTATTGGGAAT
102 TAG GCCGCCGCTCCAGCTTTGCACGTTTCGATCCCAAAGGCCCTTTTTGGGCCGACCACTTGCTCATCCT
103 TAA TTTAGCTAATTCTAAGAAAACGGTTTCAACAAGCAAATATTTGGGATTCGTATTCAACTGCCCGGTT
104 TAA ATCTCTTCCTTCCCACTAATCTTATACCCAAAGGCTCTTTTCAGAGCCCCCCACACTCTCACTCCTA
105 TGA ACAAAATATTTATCTTAAAAAATTAAAAAGTAAAAAGCTGCATGCTTACTCAAAGGTAATAGTGTAA
106 TAA AGGCTCGCTTCGGTTTTTTCTGTTTTTTGCCCCCCACAATCAAGATAAGCCCTTTTAAGGGCCACAA
107 TAA ACGCACTCTTCTCGGAACGTCAACATCCCGACTTGAACCCAAAGGCTCTTTTAAGAGCCACCCACAT

                80                    100                 120               140
 87 ACTTGGTCCAACAAAGAGCATGAATT
 88 TTCCTCAGCTGAAGGTGGTAACACTGACGAGGTGTTTTGGTAGGTACGGAATTTTGCTTGGTTCTGAGTCAG
 89 TTCAGGAGAGCTGACACTGACTTGGGTCGTACAGGTAATAACCGCGGGTTTAGGACTCACGCTACTAGGTGT
 92 TGTTGTATGTTATGTTACTACCGTAAAGAAAGTAAAGAAAGAAGAAGAAGAA
 95 GCTGTTACTGCTTTTTTTACATATTTAT
 96 GGAGGTGTTTTGGTAGGTACGGAATTTTGCTTGGTTCTGAGTCAGTTCTGGGGGGAACAGTTTTTTGAACAC
 97 AACAATTTGAATGTCTCTTATTTATGTGTTATAATTTATGTAC
 98 ATGGTTATTTCTCTGTCATCAATGATAATTTTTTTTGTGTTATTCAAGTCAAATCCCCTTTCCAAAAGAAAA
 99 CGACTTGTTTAGTTCGCTATTTCCTCTGTAGTTTGAACTCAATCCGTGGAACAAAGTTATTCGCAATATATT
100 TCAGTTTC
101 ATAATAGAACCTTTATGTATGCATTTTTCGGTTTTTTATTATTATTTCATGTACTATATATATATTAAA
102 GAGGAGTTGGACACTTGACTGCGTAAAGTGCAACAGTAACGATGTTGGAAGGTAACTTTGG
103 TTTCTTCTACAAATATTGAGTGTATTATAGCGGGTTTAGGAGAAATATGTACAATTATAAACATATAAAGGG
104 GAGCTGTTACTGCTTTTTTTACATATTTA
105 TTATCTAGTTCTTTTATCTGAGAGAGTATGCTTTTTTCTATCGAGTGTTAGTGTAGCAATTTTCTAAAGTGC
107 CCGCTTCAAAAGGGTCAAATCCATTATCGTAGGAAACCATGAGCCACTCTTGAGTGGACAGGGAGGGAGTAT
```

Figure 15. Conserved Upstream Sequences

```
  1 AAACACA..5.GC.42.ACCAATCA..7.GCTCC..3.TATAAATACGAG.20.CCAGTGGT.28.CACATG
  2 GAACACA..5.GC.42.ACCAATCA..7.GCTCC..3.TAAAAATACGAG.18.CCAATTGT.28.GCGATG
  4 AAACACA.40.GC.34.GCCAATCA..2.GACCC..9.GAATATAAGGAG.27.TTAGATTT.20.AAGATG
  5 AAACACA.40.GC.38.GCCAATCA..2.GACCC..9.GAATATAAGGAG.27.TTAGATTT.20.AAAATG
  6                              GTACG..8.CCTTATATTGAG.17.TTATTCGT.29.AAGATG
  7 AAACACA....35....ACAAATTG.11.GTACG.11.GATATAGGTGAG.16.TCACTTGT.29.AAGATG
H1  AAACACA    GC    RCCAATCA    GNNCC     ATATAAANGAG    YYANTTGT    AAGATG

 18                  CTCAACTG.21.CCTCC.4..AGTATAAATAC..6.TTCATTAC..37.GTTATG
 21                              TTTCC.3..CTTATAACTGC.21.TCGATTCG.141.GCGATG
 22                                                         ATTGT..62.ACCATG
 24                                                      TCCATTCA..62.ATCATG
 25                  GACAATTG.29.GATCC.8..GGTATAAATAG.23.ACCATTCA..62.ATCATG
 26                  TCCAATTA.13.TGTCC.7..CCTATAAATAT.19.TTCAGTCT..39.ATCATG
 29                  GTCAATTC.....47......AGTATAAATAT.20.AGCATTCG..41.AAAATG
 30                  GCCAATCG.....16......GGTATAAGTAA.21.CTCATTCA..27.AATATG
H2A                  YCAATT      NNTCC      TATAAATA      YCATTC      ANNATG

 42    TACCAATCA.22.CCTTATTTGCATNN...5..TCTATATAAAA..19..TCACGCT..48..GCTATG
 43    AACCACTGA.33.TGACGTTTGCAGAC..25..TATTTAAAAGA..21..TCAATAT..32..ACTATG
 44    GACCAATGA.21.TCTCATTTGCATAG...4..GCTATAAATAA..19..CCATTCA..49..ACTATG
 45    ATCCAATCA.20.CTCGATTTGCATAC...4..CCTATAAATAG..18..GCACTCC..28..GACATG
 48    ATCCAATCA.20.CTCGATTTGCATAC...4..CCTATAAATAG..18..GCACTCC..28..GACATG
 49    GACCAATGA.17.GCTCATTTGCATAC..37..TGTATAAAAAG..14..CCATTCA..68..CATATG
 50    TTACAAGAT.10.CCTTATTTGCATGG...4..GCTATAAAAGC..20..ACAGTTT..25..ATTATG
 51     ............CATGATTTGATCAT..16..TTATTAAAGAC..12..GCATTTT..59..ACAATG
 52     ............CATGGTATATGTGT..27..TATTTATTTAG..10..TTACTTT..55..ATAATG
 53    CCGCATTCA.21.CCTGAATTGACTGC...9..AGTATAAACGT..12..CCATTCG..39..ACAATG
H2B      CCAATNA    CYTNATTTGCATAC        TATAAATA        CAYTCY      RNNATG

 60            GGCCAATCC.. 51.GGTCT..4..CTATAAAAGA..20..TCCACAGA..32..GCCATG
 61            GACCAATGA.. 16.GTTTC..5..CTATAAATAA..19..CCTACTTT..19..CGAATG
 63            CGCCTATCA.. 59.GACGC..4..GTACTTAAAG..18..GGTATCTC..20..GCCATG
 64            ATCCAATTA.. 59.GTTCC..2..CTATAAAGGA..17..TGCATTTC..13..GAAATG
 69            GGCCAATCA.. 30.GACGC.10..CTTATAAACT..18..GCTATACT..29..GCCATG
 70            AGCCAATCA.  24.GATCC..8..ATATAAATAG..19..CTCATTCA..47..ACTATG
 71                                                    . TTCATCTC..47..ACCATG
 74            GACCAATGA..43..AGTCT..4..GTATAAATAC..22..ATCAGTTG..16..ACAATG
 75            AACCAATGG..34..GATAC..5..CTATAAAGAT..18..ATCAGTTT..24..AAAATG
 76            GTCAAATGA..40 .GCTCC..7..CTATATAAGT..21..TTTATTGT..49..GAGATG
 77            GTCCACTCC..72 .GTCCC..2..CTATTTAACT..23..CTCACCCC..54..GACATG
 81                           GATCT.10..TTATAAATAG..20..GTCACTCA..49..ACCATG
H3              CCAATCA       GATCC      TATAAA          TCAYTT       RNNATG

 87  Pu-Rich..AGGCAAGTG..42...GGTCC..8..GCATAACAC...16..CACATCCT..24..GGCATG
 88           AAACATATC..25...GTTTT..3..CTATAAATA...19..ATTAGTTC..33..AAAATG
 89  Pu-Rich..TTTCAATCT...5...GATCC..3..TCATATATT...21..CTCACCAG..21..GCTATG
 91           GGTCATCGG..24...AGTCC..8..GTGACAATA...22..GTCAATCG..54..ATAATG
 92           GTGCAATCC...8...GTTTC.10..CGTAATATC...17..ATCACACA..21..ATCATG
 96                           GGTCC.10..GTATTAAAG...18..GCCACAAA..25..GTCATG
 97           GAGCATTAC..26...CGTCC..9..GTATATATA...19..GTCAATTC..25..ATCATG
 98           GAACAAAGG..21...CGTTC.10..CTATATATA...23..ATCAGTTT..22..AACATG
 99           GTCCACTCC..32...CAACC..9..CTTTAAGAC...21..ATCACAGC..61..GCCATG
102  Pu-Rich..GGACAATTG..37...GGTCC..9..GTATATTAC...18..CCGATCGA..20..GTCATG
104  Pu-Rich..GAACAATCA..58...GGTCC.11..ATATAAAGA...19..GTTATATT..20..AGAATG
105           TAGCAAGAA..24...TATCA..6..TTATAAATA...22..AACAATAA..18..AAAATG
107           TGCCACTGG...9...GGTCC..9..AGATAATAA...18..ATTAACGT..32..ACGATG
H4               CAAT         NRTCC      TATAAATA        RYCA         ANCATG
```

Figure 16. Conserved Downstream Sequences

```
  1 TAA..15...TCGA..20..CAAAGGCTCTTTTAAGAGCCACCCA...8..CAGAAAGAGCTGGAA
  2 TAA..15...TCCT..20..CAACGGCTCTTTTAAGAGCCACCCA..10..CTAAAGGAGCTGAGG
  4 TAA..13...TCCC..20..CAAAGGCTCTTTTCAGAGCCACCAC...8..TGAGAAGAGCCGATA
  5 TAA..13...TCCC..20..CAAAGGCTCTTTTCAGAGCCACCAC...8..TCAGAAGAGCCGATA
  6 TGA..15...TCCC...9..AAACGGCTCTTTTCAGAGCCACCAC
  7 TGA...9...TCCT..18..AAACGGCTCTTTTCAGAGCCACCAC...6..ACGTAAGACC
  9 TAA..14...CCCA---------------------------------------------------
 18 TGA........55......GCAAGGCTCTTTTCAGAGCCACCTA...7..CCATTAAATGAGCTG
 19 TAA........48......CCAAGGCTCTTTTCAGAGCCACCAC...8..TCATATAAGAG
 21 TGA........30......TCCAAGCTCTTTTCAGAGCCACCCA...7..GCAGGAGAGCTCAGA
 24 TGA.........3......CAACGGCCCTTATCAGGGGCCACCA...5..TCACGAAAGAATTGT
 25 TAA........25......CAACGGCCCTTATCAGGGCCACCAA...6..TCAAGAAAGATAAAA
 26 TGA........33......CAAGGGCTCTTTTCAGAGCCGCCAC...4..GCAAATCAGAGCTCA
 30 TAA........33......CAAAGGCTCTTTTAAGAGCCAACCA...6..TCAACAAAAGCGCAA
 42 TAA........26......TAACGGCTCTTTTAAGAGCCACCCA...6..TCAAAGAAAGAGCTG
 43 TGA........21......CAAAGGCTCTTTTCAGAGCCACTCA...5..TCAAAATTGGAGCTT
 44 TAA........31......CAAAGG    CTTCAGAGCCACCCA...5..CCAGAGAAAGAGCTG
 45 TAG........24......CAAAGGCTCTTTTCAGAGCCACCAT...5..CTAATAAAAGGGCTG
 46 TAG........22......CCAAGGCTCTTTTCGGAGCCACCAT...5..CTAATAAAAGGGCTG
 48 TAG........30......CAACGGCCN          NACAAA...3..TCAAGAAAGAATGAT
 49 TAA........30......................GGCCACCAA...4..CCAAGAAAGAATTGT
 50 TAA........19......CAAAGGCTCTTTTCAGACCACCCAT...5..TCCCGAAAAGATCT
 60 TAA........29......CAAAGGCTCTTTTCAGAGCCACCCA...6..CCAACGAAAGTAGCTG
 63 TAA........25......TAAAGGCTCTTCTTAGAGCCACCCA...4..CCAAAAAAAGAACTG
 64 TAA........28......AAACGGCTCTTTTAAGAGCCACCTC...9..CCAAAGATGCTTGAAG
 69 TAA........22......CAAAGGCTCTTTTAAGAGCCACCTC..10..TCAAAAAGACACAATT
 70 TAG........25......AAACGGCTCTTTTCAGAGCCACCAC...4..CCAAGAAAGAXTCATT
 74 TAA........19......AAACGGCTCTT TCGAGACCACCAA...4..TCAAGAAAGAATCAAA
 75 TAA........21......AACCGGCTCTTTTCAGAGCCACTAA...4..ACAAGAAAGACACAAA
 76 TAA........31......AATCGGTCCTTTTCAGGACCACAAA...7..TCAATGAGATAAATTT
 81 TAG........31......AAACGGCTCTTTTCAGAGCCACCAT...4..TCACGAAAGAATCACT
 87 TAA...Py-RICH(36)...CAAAGGCTCTTTTCAGAGCCGCCCA...6..CCAACAAAGAGCATGAA
 89 TAG...Py-RICH(35)...TAACGGCCCTTTTTAGGGCCAACCA...6..CTTCAGGAGAGCTGACA
 92 TAA.........(24)...AAACGGCTCTTTTCAGAGCCACCAA...4..TCAAGAAAGAATACTGT
 93 TAA...Py-RICH(24)...CAACGGCCCTTTTAAGGGCCACA
 94 TAA...Py-RICH(24)...AAACGGCCCTTTTAAGGGCCACCA
 95 TAA...Py-RICH(27)...CCAAGGCTCTTTTCAGAGCCCCCCA...9..CCTAGAGCTGTTACTGC
 96 TGA...Py-RICH(16)...TAAGAGGCCTTTTCAGGGCCCCTAC...3..CTCAGCTGAAGGTGGTA
 97 TAA...Py-RICH(27)...AACCGGCTCTTTTCAGAGCCACCAA...3..TTCAAGAAAGAAACAAT
 98 TAA...Py-RICH(23)...AAACGGCTCTTTTTAGAACCACCAA...3..CTCAAAGAAAGAATCAA
102 TAG...Py-RICH(31)...CAAAGGCCCTTTTTGGGCCGACCAC...8..CCTGAGGAGTTGGACAC
104 TAA...Py-RICH(29)...CAAAGGCTCTTTTCAGAGCCCCCCA..10..CCTAGAGCTGTTACTGC
106 TAA...Py-RICH(43)...GATAAGCCCTTTTAAGGGCCACAA
107 TAA...Py-RICH(39)...CAAAGGCTCTTTTAAGAGCCACCCA...7..TTCAAAAGGGTCAAATC

COS                     C  C   T        A      A
                         AA GGC CTTTTCAG GCCACC
                        A  A   C        G      C
```

## REFERENCES

1. Bains, W. (1986) Nucl. Acids Res. (in press).
2. Bannon,G.A., Bowen,J.K., Yao,M.C. and Gorovsky, M.A. (1984) Nucl. Acids Res. 12, 1961-1975.
3. Birnstiel, M.L., Portmann, R., Busslinger, M., Schaffner, W., Probst, E., and Kressman, A. (1978). In "Specific Eukaryotic Genes", Proc. Alfred Benzon Symposium (Engberg, J., Klenow, H. and Lieck, V., eds.). Munksgaard, Copenhagen, 13, 117-129.
4. Brandt, W.F., Strickland, W.N., and von Holt, C. (1974) FEBS Lett. 40, 349-352.
5. Busslinger, M., Portmann, R. and Birnstiel, M.L. (1979). Nuc. Acids Res. 6, 2997-3008.
6. Busslinger, M., Portmann, R., Irminger, J.C., and Birnstiel, M.L. (1980) Nucl. Acids Res. 8, 957-977.
7. Carozzi, N., Marashi, F., Plumb, M., Zimmerman, S., Zimmerman, A., Coles, L.S., Wells, J.R.E., Stein, G., and Stein, J. (1984) Science 224, 1115-1117.
8. Childs, G., Nocente-McGrath, C., Lieber, T., Holt, C., and Knowles, J.A. (1982) Cell 31, 383-393.
9. Choe, J., Kolodrubetz, D., and Grunstein, M. (1982) Proc. Nat. Acad. Sci. USA 79, 1484-1487.
10. Clark, S.J., Krieg, P.A., and Wells, J.R.E. (1981) Nucl. Acid Res.9, 1583-1590.
11. Clerc, R.G., Bucher, P., Strub, K., and Birnstiel, M.L.(1983) Nucl. Acid Res. 11,8641-8657.
12. Coles, L.S., and Wells, J.R.E. (1985) Nucl. Acids Res. 13, 585-594.
13. Conner, W., Mezquita, J., Winkfein, R.J., States, J.C., Dixon, G.H. (1984) J. Mol. Evol. 20, 227-235.
14. D'Andrea, R., Harvey, R., and Wells, J.R.E. (1981) Nucl. Acids Res. 9, 3119-3128.
15. DeLange, R.J., Fambrough, D.M., Smith, E.L., and Bonner, J. (1969) J. Biol. Chem. 244, 5669-5679.
16. DeLange, R.J., Hooper, J.A., and Smith, E.L. (1973) J. Biol. Chem. 248, 3261-3274.
17. Engel, J.D., Sugarman, B.J., and Dodgson, J.B. (1982) Nature 297, 434-436,
18. Fusauchi, Y., and Iwai, K. (1983) J. Biochem. 93, 1487-1497,
19. Goldberg, M.L. (1979) Ph.D. Thesis, Stanford Univ.
20. Grandy, D.K., Engel, J.D., and Dodgson, J.B. (1982) J. Biol. Chem. 257, 8577-8580.
21. Grunstein, M., Diamond, K.E., Knoppel, E., and Grunstein, J.E. (1981) Biochem. 20, 1216-1223.
22. Harvey, R.P., Robins, A.J., and Wells, J.R.E. (1982) Nucl. Acid Res. 10, 7851-7863.
23. Harvey, R.P., Whiting, J.A., Coles, L.S., Krieg, P.A., and Wells, J.R.E. (1983) Proc. Nat. Acad. Sci. USA 80, 2819-2823.
24. Heintz, N., Zernik, M., and Roeder, R.G. (1981) Cell 24, 661-668.
25. Hentschel, C.C., and Birstiel, M.L. (1981) Cell 25, 301-313.
26. Hooper, J.A., Smith, E.L., Sommer, K.R., and Chalkley, R. (1973) J. Biol. Chem. 248, 3275-3279.
27. Iwai, K., Hayashi, H., and Ishikawa, K. (1972) J. Biochem. 72, 357-367.
28. Jones, G.M.T., Rall, S.C., and Cole, R.D., J. Biol. Chem. 249, 2548-2553.
29. Kmiecik, D., Couppez, M., Belaiche, D., and Sautiere, P. (1983) Eur. J. Biochem. 135, 113-121.
30. Kootstra, A., and Bailey, G.S. (1978) Biochemistry 17, 2504-2510.
31. Krieg, P.A., Robins, A.J., D'Andrea, R., and Wells, J.R.E., (1983) Nucl. Acids Res. 11, 619-627.

32. Laine, B., Sautiere, P., and Biserte, G.(1976) Biochemistry 15, 1640-1645.
33. Laine, B., Kmiecik, D., Sautiere, P., and Biserte, G.(1978) Biochimie 60, 147-150.
34. Liao, L.W., and Cole, R.D.(1981) J. Biol. Chem. 256, 3024-3029.
35. MacLeod, A.R., Wong, N.C.W., and Dixon, G.H. (1978) Eur. J. Biochem. 78, 281-291.
36. Martinage, A., Belaiche, D., Dupressoir, T., and Sautiere, P.(1983) Eur. J.Biochem. 130, 465-472.
37. Moorman, A.F.M., DeLaaf, R.T.M., Destree, O.H.J., Telford, J., and Birnstiel, M.L. (1980) Gene 10, 185-193.
38. Moorman, A.F.M., DeBoer, P.A.J., DeLaaf, R.T.M., Van Dongen, W.M.A.M. and Destree, O.H.J. (1981) FEBS Lett.136, 45-52.
39. Moorman, A.F.M., DeBoer, P.A.J., DeLaaf, R.T.M., and Destree, O.H.J. (1982) FEBS Lett. 144, 235-241.
40. Ogawa, Y., Quagliarotti, G., Jordan, J., Taylor, C.W., Starbuck, W.C., and Busch, H. (1969) J. Biol. Chem. 244, 4387-4392.
41. Patthy, L., Smith, E.L., and Johnson, J. (1973) J. Biol. Chem. 248, 6834-6840.
42. Patthy, L., and Smith, E.L. (1975) J. Biol. Chem. 250, 1919-1920.
43. Rall, S.C., and Cole, R.D. (1971) J. Biol. Chem. 246, 7175-7190.
44. Roberts, S.B., Weisser, K.E., and Childs, G. (1984) J. Mol. Biol. 174, 647-662.
45. Rodrigues, J. de A., Brandt, W.F., and von Holt, C. (1979) Biochim. Biophys. Acta 578, 196-206.
46. Ruberti, I., Fragapane, P., Pierandrei-Amaldi, P., Beccari, E., Amaldi, F., and Bozzoni, I. (1982) Nucl. Acid Res. 10, 7543-7559.
47. Ruiz-Carrillo, A., Affolter, M., and Renaud, J. (1983) J. Mol. Biol. 170, 843-859.
48. Sautiere, P., Lambelin-Breynaert, M.-D., Moschetto, Y., and Biserte, G. (1971a) Biochimie 53, 711-715.
49. Sautiere, P., Tyrou, D., Moschetto, Y., and Biserte, G. (1971b) Biochimie 53, 479-483.
50. Sautiere, P., Tyrou, D., Laine, B., Mizon, J., Ruffin, P., and Biserte, G. (1974) Eur. J. Biochem. 41, 563-576,
51. Sautiere, P., Kmiecik, D., Loy, O., Briand, G., Biserte, G., Garel, A., and Champagne, M. (1975) FEBS Lett. 50, 200-203.
52. Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H.O., and Birnstiel, M.L.(1978) Cell, 655-671.
53. Seiler-Tuyns, A., and Birnstiel, M.L. (1981) J. Mol. Biol. 151, 607-625.
54. Sierra, F., Stein, G., and Stein, J. (1983) Nucl. Acids Res. 11, 7069-7086.
55. Sittman, D.B., Chiu, I.M., Pan, C.J., Cohn, R.H., Kedes, L.H., and Marzluff, W.F. (1981) Proc. Nat. Acad. Sci. USA 78, 4078-4082.
56. Smith, M.M., and Andresson, O.S. (1983) J.Mol. Biol. 169, 633-690.
57. Stephenson, E.C., Erba, H.P., and Gall, J.G. (1981) Nucl. Acid Res. 9, 2281-2295.
58. Strickland, M., Strickland, W.N., Brandt, W.F., and von Holt, C. (1977) Eur. J. Biochem. 77, 263-275.
59. Strickland, M., Strickland, W.N., Brandt, W.F., von Holt, C., Wittmann-Liebold, B., and Lehmann, A. (1978) Eur. J. Biochem. 89, 443-452.
60. Strickland, W.N., Strickland, M., de Groot, P.C., von Holt, C., and Wittmann-Liebold, B. (1980) Eur. J. Biochem. 104, 559-566.
61. Strickland, W.N., Strickland, M.S., de Groot, P.C., and von Holt, C.(1980) Eur. J. Biochem. 109, 151-158.
62. Strickland, W.N., Strickland, M., and von Holt, C. (1982) Biochim. Biophys. Acta 700, 127-129.

63. Sugarman, B.J., Dodgson, J.B., and Engel, J.D. (1983) J. Biol. Chem. 258, 9005-9016.
64. Sures, I., Lowry, J., and Kedes, L.H. (1978) Cell 15, 1033-1044,
65. Tabata, T., Sasaki, K., and Iwabuchi, M. (1983) Nucl. Acids Res. 11, 5865-5875.
66. Tabata, T., Fukasawa, M. and Iwabuchi, M. (1984) Mol. Gen. Genet. 196, 397-400.
67. Turner, P.C., and Woodland, H.R. (1982) Nucl. Acids Res. 10, 3769-3780.
68. Turner, P.C., Aldridge, T.C., Woodland, H.R., and Old, R.W.(1983) Nucl. Acids Res. 11, 4093-4107.
69. Van Helden, P., Strickland, W.N., Brandt, W.F., and von Holt, C. (1978) Biochim. Biophys. Acta 533, 278-281.
70. Van Helden, P.D., Strickland, W.N., Brandt, W.F., and von Holt, C.(1979) Eur. J. Biochem. 93, 71-78.
71. Wallis, J.W., Hereford, L., and Grunstein, M. (1980) Cell 22, 799-805.
72. Wang, S-W., Robins, A.J., D'Andrea, R., and Wells, J.R.E. (1985) Nucl. Acids Res. 13, 1369-1397.
73. Wells, D., and Kedes, L. (1985) Proc. Nat. Acad. Sci. USA 82, 2834-2838.
74. Wells, D., Bains, W., and Kedes, L. (1986) J. Mol. Evol. (in press).
75. Wilhelm, M.L., and Wilhelm, F.X., FEBS Lett. 168, 249-254, 1984
76. Wilson, R.K., Starbuck, W.C., Taylor, C.W., Jordan, J., and Busch, H. (1970) Cancer Res. 30, 2942-2951.
77. Winkfein, R.J., Conner, W., Mezquita, J., and Dixon, G.H. (1985) J. Mol. Evol. 22, 1-19.
78. Woudt, L.P., Pastink, A., Kempers-Veenstra, A.E., Jansen, A.E.M., Mager, W.H., and Planta, R.J.(1983) Nucl. Acids Res. 11, 5347-5360.
79. Wouters-Tyrou, D., Martin-Ponthieu, A., Briand, G., Sautiere, P., and Biserte, G. (1982) Eur. J. Biochem. 124, 489-498.
80. Wouters, D., Sautiere, P., and Biserte, G. (1978) Eur. J. Biochem. 90, 231-239.
81. Yaguchi, M., Roy, C., and Seligy, V.L. (1979) Biochem. Biophys. Res. Commun.90, 1400-1406.
82. Zernik, M., Heintz, N., Boime, I., and Roeder, R.G.(1984) Cell 22,807-815.
83. Zhong, R., Roeder, R.G., and Heintz, N. (1983) Nucl. Acids Res. 11, 7409-7425.