Histone Sequence Database: sequences, structures, post-translational modifications and genetic loci

Izabela Makalowska, Erik S. Ferlanti, Andreas D. Baxevanis and David Landsman

Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Building 49, Room 2C-76, Bethesda, MD 20892, USA and Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Building 38A, Room 8N-805, Bethesda, MD 20894, USA

Received October 5, 1998; Accepted October 7, 1998

ABSTRACT

The Histone Sequence Database is an annotated and searchable collection of all available histone and histone fold sequences and structures. Particular emphasis has been placed on documenting conflicts between similar sequence entries from a number of source databases, conflicts that are not necessarily documented in the source databases themselves. New additions to the database include compilations of post-translational modifications for each of the core and linker histones, as well as genomic information in the form of map loci for the human histone gene complement, with the genetic loci linked to Online Mendelian Inheritance in Man (OMIM). The database is freely accessible through the World Wide Web at either http://genome.nhgri.nih.gov/histones/ or http://www.ncbi.nlm.nih.gov/Baxevani/HISTONES

INTRODUCTION

The histones are amongst the most highly conserved of all proteins throughout evolution, as these proteins are required for the proper compaction and organization of cellular DNA. These proteins fall into two general groups: core histones and linker histones. The core histones (H2A, H2B, H3 and H4) form an octameric assembly that comprises the protein core of nucleosomes (1) and facilitates the histones (H2A, H2B, H3 and H4) form an octameric assembly that into two general groups: core histones and linker histones. The core compaction and organization of cellular DNA. These proteins fall throughout evolution, as these proteins are required for the proper The histones are amongst the most highly conserved of all proteins INTRODUCTION

The histones are amongst the most highly conserved of all proteins throughout evolution, as these proteins are required for the proper compaction and organization of cellular DNA. These proteins fall into two general groups: core histones and linker histones. The core histones (H2A, H2B, H3 and H4) form an octameric assembly that comprises the protein core of nucleosomes (1) and facilitates the histones (H2A, H2B, H3 and H4) form an octameric assembly that into two general groups: core histones and linker histones. The core compaction and organization of cellular DNA. These proteins fall throughout evolution, as these proteins are required for the proper The histones are amongst the most highly conserved of all proteins throughout evolution, as these proteins are required for the proper compaction and organization of cellular DNA. These proteins fall into two general groups: core histones and linker histones. The core histones (H2A, H2B, H3 and H4) form an octameric assembly that comprises the protein core of nucleosomes (1) and facilitates the histones (H2A, H2B, H3 and H4) form an octameric assembly that into two general groups: core histones and linker histones. The core compaction and organization of cellular DNA. These proteins fall throughout evolution, as these proteins are required for the proper The histones are amongst the most highly conserved of all proteins throughout evolution, as these proteins are required for the proper compaction and organization of cellular DNA. These proteins fall into two general groups: core histones and linker histones. The core histones (H2A, H2B, H3 and H4) form an octameric assembly that comprises the protein core of nucleosomes (1) and facilitates the histones (H2A, H2B, H3 and H4) form an octameric assembly that into two general groups: core histones and linker histones. The core compaction and organization of cellular DNA. These proteins fall throughout evolution, as these proteins are required for the proper The histones are amongst the most highly conserved of all proteins throughout evolution, as these proteins are required for the proper compaction and organization of cellular DNA. These proteins fall into two general groups: core histones and linker histones. The core histones (H2A, H2B, H3 and H4) form an octameric assembly that comprises the protein core of nucleosomes (1) and facilitates the histones (H2A, H2B, H3 and H4) form an octameric assembly that into two general groups: core histones and linker histones. The core compaction and organization of cellular DNA. These proteins fall throughout evolution, as these proteins are required for the proper The histones are amongst the most highly conserved of all proteins throughout evolution, as these proteins are required for the proper compaction and organization of cellular DNA. These proteins fall into two general groups: core histones and linker histones. The core histones (H2A, H2B, H3 and H4) form an octameric assembly that comprises the protein core of nucleosomes (1) and facilitates the histones (H2A, H2B, H3 and H4) form an octameric assembly that into two general groups: core histones and linker histones. The core compaction and organization of cellular DNA. These proteins fall throughout evolution, as these proteins are required for the proper The histones are amongst the most highly conserved of all proteins throughout evolution, as these proteins are required for the proper compaction and organization of cellular DNA. These proteins fall into two general groups: core histones and linker histones. The core histones (H2A, H2B, H3 and H4) form an octameric assembly that comprises the protein core of nucleosomes (1) and facilitates the histones (H2A, H2B, H3 and H4) form an octameric assembly that into two general groups: core histones and linker histones. The core compaction and organization of cellular DNA. These proteins fall throughout evolution, as these proteins are required for the proper The histones are amongst the most highly conserved of all proteins throughout evolution, as these proteins are required for the proper compaction and organization of cellular DNA. These proteins fall into two general groups: core histones and linker histones. The core histones (H2A, H2B, H3 and H4) form an octameric assembly that comprises the protein core of nucleosomes (1) and facilitates the histones (H2A, H2B, H3 and H4) form an octameric assembly that into two general groups: core histones and linker histones. The core compaction and organization of cellular DNA. These proteins fall throughout evolution, as these proteins are required for the proper The histones are amongst the most highly conserved of all proteins throughout evolution, as these proteins are required for the proper compaction and organization of cellular DNA. These proteins fall into two general groups: core histones and linker histones. The core histones (H2A, H2B, H3 and H4) form an octameric assembly that comprises the protein core of nucleosomes (1) and facilitates the histones (H2A, H2B, H3 and H4) form an octameric assembly that into two general groups: core histones and linker histones. The core compaction and organization of cellular DNA. These proteins fall throughout evolution, as these proteins are required for the proper The histones are amongst the most highly conserved of all proteins throughout evolution, as these proteins are required for the proper compaction and organization of cellular DNA. These proteins fall into two general groups: core histones and linker histones. The core histones (H2A, H2B, H3 and H4) form an octameric assembly that comprises the protein core of nucleosomes (1) and facilitates the histones (H2A, H2B, H3 and H4) form an octameric assembly that into two general groups: core histones and linker histones. The core compaction and organization of cellular DNA. These proteins fall throughout evolution, as these proteins are required for the proper The histones are amongst the most highly conserved of all proteins throughout evolution, as these proteins are required for the proper compaction and organization of cellular DNA. These proteins fall into two general groups: core histones and linker histones. The core histones (H2A, H2B, H3 and H4) form an octameric assembly that comprises the protein core of nucleosomes (1) and facilitates the histones (H2A, H2B, H3 and H4) form an octameric assembly that into two general groups: core histones and linker histones. The core compaction and organization of cellular DNA. These proteins fall through...
Table 1. Histone Sequence Database statistics

<table>
<thead>
<tr>
<th>Histone</th>
<th>Total Sequence Set</th>
<th>Non-Redundant Sequence Set</th>
<th>Structures</th>
<th>Human Gene Loc</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>209</td>
<td>78</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>H2A</td>
<td>235</td>
<td>83</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>H2B</td>
<td>216</td>
<td>75</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>H3</td>
<td>233</td>
<td>84</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>H4</td>
<td>185</td>
<td>61</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>H5</td>
<td>13</td>
<td>5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Octamer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nucleosomal Core Particle</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total Histone Entries: 1091; 386; 4; 60

Histone Fold Proteins: 52; 4

Modification was collected from both SWISS-PROT and GenBank, and color-coded multiple sequence alignments showing the amino acid positions that can undergo modification are available on the Web site. As the information collected represents only what has been documented in the target databases, information which appears only in the literature may not be reflected in the multiple sequence alignment. Users are encouraged to inform the authors of any updates or corrections to this or any other information.

As a result of both the Human Genome Project and a number of directed gene-hunting efforts (10 and references cited therein), the chromosomal location of the human histone genes has been deduced. Most of the histone genes are located in two discrete clusters, at 1q21–23 and at 6p21.3–6p22 (Fig. 1). A listing of all of the human histone genes are available on the database Web site, giving the gene name, alternate names, and map locations, with each entry being linked to Online Mendelian Inheritance in Man (OMIM). The map locations used are those found within the individual OMIM entries. From OMIM, users can link to MEDLINE, sequence entries, UniGene (11), the OMIM gene map and other offsite resources as available.

In addition to the histone sequences, the database contains sequences of non-histone proteins that contain the conserved histone fold motif (12). For this group, both full-length and motif-length sequence sets in FASTA format are available. PostScript files containing multiple sequence alignments of these histone fold proteins are also available. Finally, links are provided to coordinate data for all histone and histone fold NMR and X-ray structures that have been determined to date.

DATABASE AVAILABILITY

The Histone Sequence Database is available through the World Wide Web at either http://genome.nhgri.nih.gov/histones/ or http://www.ncbi.nlm.nih.gov/Baxevani/HISTONES. A menu bar appears to the left of each page, allowing users to easily navigate the Web site without having to return to the home page to examine different parts of the site. Studies utilizing the data within this database should cite this paper as the primary reference.

REFERENCES