Histones bind and compact nuclear DNA in virtually all eukaryotes (17). Histones were therefore presumably a feature of the first nucleus and also of its immediate ancestor, but the origin of histones has been an unresolved issue. Bacteria do not have histones but Archaea do, a fact which argues that histones evolved in the archaeal lineage after the Bacteria-Archaebacteria divergence (Fig. 1A) (10, 21). Histones have, however, been documented only in Euryarchaea, and their conspicuous absence in Crenarchaea is often cited and noted as a difference between the two major archaeal lineages (1, 14). If valid, this would argue that histones evolved relatively late, after the Euryarchaea-Crenarchaea divergence (Fig. 1A), consistent with proposals, such as the hydrogen hypothesis (8), that posit a Crenarchaea origin for the eukaryotic nucleus. Alternatively, histones could have been lost in the lineages leading to all the other archaea. The Crenarchaea would argue that histones evolved relatively late, after the Crenarchaea-Euryarchaea divergence (Fig. 1A) (20), and the presence of two histone-encoding genes in the Crenarchaea genome (Fig. 1B) (11, 12). Only thermophilic and hyperthermophilic Crenarchaea (Fig. 1B) (14) have histones but very large numbers of as-yet-uncultured Crenarchaea are known to be present in cold marine environments (2, 5, 6, 9). We therefore screened the DNA sequences obtained from microorganisms filtered from the Sargasso Sea (18) for histone genes.

Genes that encode three closely related proteins (EAG39378 [hereafter designated EAG3], EAJ42364, and EAK74163) that have sequences, key residues, and basic pIs consistent with being archaeal histones are present in this database (Fig. 1B). These proteins do not, however, have two prolines near the N terminus as found in almost all other archaeal histones (e.g., P4 and P7 in HMfB) (Fig. 1B) (14). In an archaeal histone dimer, these residues form a proline tetrad motif that has been proposed to position surface-located residues appropriately for DNA binding (4).

Microorganisms were not cultured from the Sargasso Sea, and therefore, to pursue this observation, a gene encoding the EAG3 sequence was synthesized with codon usage optimized for expression in Escherichia coli (http://www.kazusa.or.jp/codon) and NdeI, SacI, SpII, and BamHI sites incorporated to facilitate gene assembly and cloning. Oligonucleotides (Operon Biotechnologies Inc., Huntsville, AL) were mixed in complementary pairs (Table 1), phosphorylated using T4 kinase (Invitrogen, Carlsbad, CA), purified through Sephadex G-50 spin columns (Amersham Biosciences, Piscataway, NJ), heated to 70°C, and allowed to cool to 25°C. The resulting double-stranded DNA molecules had single-strand extensions that facilitated their cloning into pUC19. The DNA molecules generated in oligonucleotide mixtures 1, 2, and 3 (Table 1) were ligated together and cloned into NdeI- and SacI-digested pUC19. The DNA molecules generated in mixtures 4 and 5 and in mixtures 6 and 7 were similarly ligated together and then cloned into SacI- and SpII-digested pUC19 and into SpII- and BamHI-digested pUC19, respectively. Preparations of the resulting plasmids, pLC1A, pLC2A, and pLC3A, were isolated after transformation and amplification in E. coli, and the EAG3-encoding DNA sequences were released by digestion with NdeI plus SacI, SacI plus SpII, and SpII plus BamHI, respectively. These DNAs were ligated together and cloned into SpII- and BamHI-digested pT7-7, generating pLC16B that was transformed into E. coli BL21 (DE3). Cultures of E. coli BL21 (DE3; pLC16B) were grown in LB medium to an optical density at 600 nm of 0.7. IPTG (isopropyl-β-D-thiogalactopyranoside) was added (1 mM final concentration), and incubation continued for 3 h at 37°C. The cells were harvested by centrifugation, resuspended in 1 M NaCl and 50 mM Tris-HCl (pH 8), and lysed and a clarified extract of soluble protein prepared as previously described (15). The polypeptides present in aliquots of each fraction generated from the extract by gel filtration chromatography (HiPrep16/60 Sephacryl S-100 HR column; Amersham Bio-

NOTES

Histones in Crenarchaea

L’ubomíra Čuboňová,*† Kathleen Sandman,1 Steven J. Hallam,2 Edward F. DeLong,2 and John N. Reeve1*

Department of Microbiology, Ohio State University, Columbus, Ohio 43210,1 and Department of Civil and Environmental Engineering and Division of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 021392

Received 10 April 2005/Accepted 14 May 2005

Archaeal histone-encoding genes have been identified in marine Crenarchaea. The protein encoded by a representative of these genes, synthesized in vitro and expressed in Escherichia coli, binds DNA and forms complexes with properties typical of an archaeal histone. The discovery of histones in Crenarchaea supports the argument that histones evolved before the divergence of Archaea and Eukarya.
FIG. 1. (A) Three-domain phylogenetic tree (10, 21). (B) Alignment of the sequences of crenarchaeal (c), euryarchaeal (e), nanoarchaeal (n), and *Xenopus laevis* eukaryotic histones (x). Residues in EAK74163 that differ from those in EAG39378 are shown above the EAG3 sequence, and residues identical in EAG3 and HCs are indicated by asterisks between the sequences. The histone fold, formed by three alpha helices separated by two loops (14, 17), must dimerize, as illustrated, for stability. Conserved residues that contact DNA (●), hydrophobic residues that form the dimer core (○), and an arginine-aspartate (R—D) salt bridge required for histone fold formation are identified. Most archaeal histones are just histone folds, but some have two histone folds (HMk), a C-terminal extension (MJ1647), or an insertion in loop 1 (NEQ288) (14). A similar insertion is also present in loop 1 of eukaryotic histone H3, but the eukaryotic histones also have N-terminal extensions (17). (C) Recombinant HMfB (3 g) and EAG3 (0.75 g) were incubated overnight at 20°C in 100 mM NaCl and 50 mM HEPES (pH 7.5) with (+) and without (−) 100 mM formaldehyde. The products generated were separated by SDS-PAGE and stained with GelCode blue. Lane S contained the size standards indicated. (D) Agarose gel shift assay of the complexes formed by incubation of increasing amounts of HMfB (50, 75, and 100 ng) or EAG3 (25, 75, 100, 125, and 150 ng) for 15 min at 20°C with 50-ng aliquots of EcoRI-linearized pBR322 DNA. Control lanes (−) had no protein added. (E) Topoisomers generated by histone binding to circular pUC18 DNA separated by linking number (upper gel) and by direction of supercoiling (lower gel) as described in detail previously (12, 15). The arrows connect aliquots of the same topoisomers separated in both the upper and lower gels. Complexes were assembled at EAG3-to-pUC18 DNA mass ratios of 0.5, 1, 1.5, 2, and 2.5. The mobility differences of negative (−ve) and positive (+ve) supercoiled plasmid DNAs are illustrated by the pUC18 topoisomers resulting from HMIB (4) binding at an HMIB to DNA mass ratio of 1. At this ratio, HMIB binding introduces negative supercoils but introduces positive supercoils at higher histone-to-DNA ratios (12, 15). EAG3 binding introduced positive supercoils at all histone-to-DNA ratios assayed. The control reaction (−) had no histone.
pooled and stored at II instrument, and fractions that contained only EAG were EAG3 by electrospray mass analysis using a Micromass QTOF identity of the material in a stained band was confirmed as visualized by GelCode blue staining after SDS-PAGE. The subjected to cation exchange chromatography (HiTrap CM Fractions that contained large amounts of a polypeptide with GelCode blue staining (Pierce Biotechnology, Rockford, IL). polyacrylamide gels (Bio-Rad, Hercules, CA) and visualized by amide gel electrophoresis (SDS-PAGE) through 15% (wt/vol) sciences) were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) through 15% (wt/vol) polyacrylamide gels (Bio-Rad, Hercules, CA) and visualized by GelCode blue staining (Pierce Biotechnology, Rockford, IL). Fractions that contained large amounts of a polypeptide with the electrophoretic mobility predicted for EAG3 were pooled, dialyzed against 0.1 M NaCl and 50 mM Tris-HCl (pH 8), and subjected to cation exchange chromatography (HiTrap CM Sepharose FF column; Amersham Biosciences). Polypeptides in fractions that eluted between 0.1 and 0.5 M NaCl were visualized by GelCode blue staining after SDS-PAGE. The identity of the material in a stained band was confirmed as EAG3 by electrospray mass analysis using a Micromass QTOF II instrument, and fractions that contained only EAG were pooled and stored at −70°C.

Dimerization is required for native histone fold formation (17), and all histones are dimers or higher-order oligomers in solution. Consistent with this, formaldialdehyde cross-linking of EAG3 resulted predominantly in molecules with the mobility of a dimer and, to a lesser extent, the mobility of tetramers (Fig. 1C). EAG3 bound spontaneously to DNA and formed complexes with linear DNA molecules that migrated faster during agarose gel electrophoresis than the original histone-complexes with linear DNA molecules that migrated faster during agarose gel electrophoresis than the original histone-free DNA (Fig. 1D). Binding to closed circular DNAs introduced positive supercoils (Fig. 1E). These are both unique and characteristic features of complexes formed by archaeal histones (12, 14, 15) and argue convincingly that EAG3 is an archaeal histone.

Most of the archaeal genes in the Sargasso Sea database appear to be derived from Archaea closely related to previously documented and abundant planktonic crenarchaeal groups (2, 5, 6, 9, 18). The origin of EAG3 was nevertheless uncertain until a gene encoding a close homologue (designated HCs [Fig. 5, 6, 9, 18]). The origin of EAG3 was nevertheless uncertain until a gene encoding a close homologue (designated HCs [Fig. 5, 6, 9, 18]). The origin of EAG3 was nevertheless uncertain until a gene encoding a close homologue (designated HCs [Fig. 5, 6, 9, 18]).

<table>
<thead>
<tr>
<th>Oligonucleotide mixture</th>
<th>Oligonucleotide sequences (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TATGAAAACCCAGCGATTTAGGTGTTT + TAGCAGAAACACCTTTAAATTCGCTGTTTTCA</td>
</tr>
<tr>
<td>2</td>
<td>CTGCTATGATCAGATTGAAAGAACAGCAGCGGT + CAGCTGCTTTCTTCAAATTACATCA</td>
</tr>
<tr>
<td>3</td>
<td>CGCGTGAAGCGATGAGCCCGGAAGAGCT + CTTCCGGGTTCCTATCGCTACGGCGATCG</td>
</tr>
<tr>
<td>4</td>
<td>CGGAGCGGAGGTTGCGAAGATTGCGATATCGAATGCGG + TTGGCAATACCAATATTCTTCAGACCCCTGTCGGAGCT</td>
</tr>
<tr>
<td>5</td>
<td>AAAACGCCGTATTGGCAGAATCTG + CTGGCCCATATCAGCAAGGCGGTTCCTTCGCAA</td>
</tr>
<tr>
<td>6</td>
<td>CGGGCCGAAAACCCAGCGAAGGCGAAGAGA + TTACATTTCCGGTCCTAAGCTATTTGCGCGCCATG</td>
</tr>
<tr>
<td>7</td>
<td>TGTAATGTGGCGCGCAAAACACCTAAATCGCTGGTTTTCA + GATCTCTTGAATTTGCCGCGTGTTTTCGCGCCAAT</td>
</tr>
</tbody>
</table>

This research was supported by grants from the Department of Energy, National Science Foundation, Gordon and Betty Moore Foundation, and National Institutes of Health.

REFERENCES


