NOTES

Histone-Like Proteins of the Dinoflagellate Cryptecodinium cohnii
Have Homologies to Bacterial DNA-Binding Proteins

J. T. Y. Wong,* D. C. New,† J. C. W. Wong, and V. K. L. Hung

Department of Biology, Hong Kong University of Science and Technology, Kowloon, Hong Kong, People’s Republic of China.

Received 22 October 2002/Accepted 3 January 2003

The dinoflagellates have very large genomes encoded in permanently condensed and histoneless chromosomes. Sequence alignment identified significant similarity between the dinoflagellate chromosomal histone-like proteins of Cryptecodinium cohnii (HCCs) and the bacterial DNA-binding and the eukaryotic histone H1 proteins. Phylogenetic analysis also supports the origin of the HCCs from histone-like proteins of bacteria.

The evolutionary development of histones and nucleosomes led to more efficient packaging of DNA into structured chromosomes, which was probably a key step in the evolution of eukaryotes from prokaryotes (13). The dinoflagellates represent an exception to this histone-based increase of genome size among eukaryotes (2). The dinoflagellate genome ranges from 2 to 3 pg (similar to the haploid human genome) to 200 pg (18), but ultrastructural and biochemical studies of the dinoflagellate chromatin identified an absence of nucleosome structures (3, 9) and histone proteins (17). Even more surprisingly, despite the lack of histones the dinoflagellate chromosomes are permanently condensed throughout interphase and have been proposed to exist in a liquid crystal state (4). Proteins associated with DNA have been isolated and characterized as predominantly low-molecular-mass proteins (14 to 17 kDa) rich in basic amino acids (17). Six chromosomal proteins were identified that represented only 10% of the DNA mass (25), a ratio even smaller than that of the prokaryotes. Immunoscreening of a cDNA expression library constructed from mRNA isolated from Cryptecodinium cohnii (HCCs) and the bacterial DNA-binding and the eukaryotic histone H1 proteins. Phylogenetic analysis also supports the origin of the HCCs from histone-like proteins of bacteria.

and other prokaryotic and eukaryotic chromosomal binding proteins is necessary to facilitate a more complete understanding of the evolution of histones.

Accordingly, we designed two PCR primers against the 5’ and 3’ ends of the previously identified clones HCC1 and HCC2: pHCC1 sense, 5’-TATGGATCATGCCCCCAAGGAGGGC-3’; and antisense, 5’-TATAAGCTTGAAGTCTGTCTTATG-3’. Reverse transcription-PCR with mRNA isolated from C. cohnii amplified a number of DNA fragments of the expected size. Subsequent sequencing of these fragments allowed identification of two clones, HCC3 and HCC4, with homology to HCC1-HCC2 (Fig. 1A). The alignment of all four HCC variants enabled the proteins to be separated into two regions (Fig. 1A). The first 40 to 45 amino acids of each protein are highly variable apart from a region of 10 amino acids that is 100% conserved among HCC1, -2, and -4. Conversely, the final 70 to 75 amino acids of all four variants are highly homologous, with 64% absolute identity and 77% similarity when conservative substitutions are considered. HLP, an HCC-like gene product isolated by PCR from the dinoflagellate Lingulodinium polyedrum (6), has a high degree of homology to the HCCs (Fig. 1B). Our database searches with BLASTX revealed similarity to a number of bacterial DNA-binding proteins, especially the bacterial HUs (Fig. 1A). HUs are the most abundant small molecular multifunctional proteins in the Escherichia coli nucleoid (10). With the help of the software program ClustalX, we aligned the putative amino acid sequences of the HCCs with five randomly selected HUs (Fig. 1A). HUs are polypeptides of about 100 amino acids, while the HCCs are about 110 amino acids in length. However, since our PCR strategy was based on the amino and carboxyl termini of HCC1 and HCC2, we cannot completely rule out the possibility of extra sequences at the C terminus. Similarly, since the amino-terminal region of the protein contains a number of methionine residues, further library construction and screening as well as mRNA extension experiments will be required to determine whether the isolated genes encode full-length proteins. Given this, compared with the HUs, the HCCs have an extra amino-
FIG. 1. Alignment of HCCs with HUs, histone-like proteins, and histone H1. (A) Multiple alignments of HCCs and selective bacterial HUs: HCC1 (AY5681), HCC2 (B5681), HCC3 (AY12851), and HCC4 (AY12851); HU-Pa (Pseudomonas aeruginosa, JC4061); HU-Dr (Deinococcus radiodurans, Q9RZ89); HU-St (Salmonella enterica serovar Typhimurium, S01732); HU-Hi (Haemophilus influenzae, P43722); and HU-Sp (Streptococcus pneumoniae, AAK75224). (B) Multiple alignments of dinoflagellate and bacterial histone-like proteins: HLP (L. polyedrum, AF482694), HAF (A. fundyense, AAD20317), BpH2 (B. pertussis, AAB40156), and XHLP (X. fastidiosa 9a5b, AAF84745). (C) Multiple alignments of HCCs and selective eukaryotic histone H1 proteins: H1-Ce (Caenorhabditis elegans, NP_510410), H1-Gg (Gallus gallus, P09987), H1-Hs (Homo sapiens, XP_009973), and H1-Mm (Mus musculus, AAA37814). Asterisks indicate positions which have a single, fully conserved residue; colons indicate that one of the “strong” groups in parentheses is fully conserved (STA/NEQK/NHQK/NDEQ/QHRK/MILV/MILF/HY/FYW); periods indicate that one of the “weaker” groups in parentheses is fully conserved (CSA/ATV/SAG/STNK/STPA/SGND/SNDEQK/NDEQHK/NEQHRK/FVILIM/HFY).
terminal sequence of about 20 to 30 amino acids in which the peptide sequence MAPK appears frequently. The highly variable nature of the amino termini suggests that these proteins are not redundant in their functions. A recent study (11) identified regions of similarity between the basic regions of the sea urchin H1b protein and the HCC2 protein. We have also been able to identify similar regions of homology between histone H1 proteins and HCCs (Fig. 1C). There is an approximately 10% identity between the two groups of proteins, predominantly in C-terminal positions occupied by basic residues. Some of these residues are also conserved between HCCs and HUs (Fig. 1A). The repeated occurrence of these residues has been postulated previously to result in a proline-kinked AK helix organization (7). Prototypical examples of this structure in histones of animals confer the ability to bind to DNA and to facilitate chromatin folding and condensation (22). One

FIG. 2. A phylogenetic tree (UPGMA) inferred from the multiple alignments of amino acid sequences of histone-like proteins from dinoflagellate HCCs (a), bacterial histone-like protein family (b), and bacterial DNA-binding HU proteins (c) was constructed with PHYLIP version 3.5. The numbers at the nodes represent the bootstrap percentages by creation of 500 replicated data sets with SEQBOOT. The accession number for each protein sequence is as follows: HCC1 (C. cohnii), A56581; HCC2 (C. cohnii), B56581; BpH2 (B. pertussis), AAB40156; XHLP (X. fastidiosa), AAFF4745; HU (Salmonella enterica serovar Typhimurium), S01732; HU (Streptococcus pneumoniae), AAK75224; HU (Haemophilus influenzae), P43722; HU (Pseudomonas aeruginosa), JC4061; HU (Deinococcus radiodurans), Q9RZ89.
of the basic residues, K-107 of HCC1, is conserved in all eukaryotic H1 proteins and bacterial HUs analyzed.

The extent of similarity (23%) between HCCs and HUs is not restricted to the basic residues at the C terminus, as is found in the histones. Three glycine residues are identical in all HUs and HCCs but not in the H1 histones. One of these glycine residues (Gly-66 in HUs) lies within the DNA-binding β-ribbon arms of the dimeric HUs (23) and is located at the arm’s flexible tip, which has been postulated to encircle the DNA (26).

Interestingly, histone H1-like sequences from Bordetella pertussis (BpH2) and Xylella fastidiosa (XHLP) were identified during the search for HCC homologues. Sequence comparison reveals that BpH2-XHLP has more than 35% sequence homology with all HCCs. Similar to sequence identities among the HCCs themselves, the C-terminal 65 amino acids show the highest identity with BpH2-XHLP (over 50%), including the three glycines. Based on sequence similarities, BpH2 and XHLP are more closely related to the HCCs, and together they form a group that is intermediate between the histone H1 proteins and the bacterial HUs. HCCs have higher homology to BpH2-XHLP than to HAF, the other histone-like protein from the dinoflagellate Alexandrium fundyense (Fig. 1C). The genomes of B. pertussis and the intercellular bacterium Chlamydia trachomatis contain other histone H1-like gene products, BpH1 and Hc1, respectively (8, 21), which are more closely affiliated with the histone H1 proteins and do not contain the glycine residues (sequences not shown). It is possible that the Hc1-BpH1 group was derived from an ancestral form between the ancestral BpH2-HCC and the eukaryotic histone H1.

Using the HCC sequences and randomly chosen published sequences of HUs and HLPs, we constructed an unrooted UPGMA (unweighted pair group method with averages) tree (Fig. 2). As expected, the three glycines. Based on sequence similarities, BpH2 and XHLP are more closely related to the HCCs, and together they form a group that is intermediate between the histone H1 proteins and the bacterial HUs. HCCs have higher homology to BpH2-XHLP than to HAF, the other histone-like protein from the dinoflagellate Alexandrium fundyense (Fig. 1C). The genomes of B. pertussis and the intercellular bacterium Chlamydia trachomatis contain other histone H1-like gene products, BpH1 and Hc1, respectively (8, 21), which are more closely affiliated with the histone H1 proteins and do not contain the glycine residues (sequences not shown). It is possible that the Hc1-BpH1 group was derived from an ancestral form between the ancestral BpH2-HCC and the eukaryotic histone H1.

The present study was partly supported by a CERG grant (HKUST6096/02 M) from the Research Grant Council of Hong Kong to J.T.Y.W.

REFERENCES


