Volume 273, number 1,2, 11-14 FEBS 08981 October 1990

The predicted molecular structure suggests that CTF/NF-I may function as a histone acetylase

Jouko Oikarinen and Riitta-Maaria Mannermaa

Collagen Research Unit, Biocenter and Department of Medical Biochemistry, University of Oulu, Kajaanintie 52A, SF-90220 Oulu, Finland

Received 14 May 1990; revised version received 31 July 1990

The primary structure of nuclear factor-I (CTF/NF-I), a eukaryotic regulatory DNA-binding protein involved in both DNA replication and gene transcription, and the secondary structure predictable from it, are compared here with those of a number of prokaryotic acetylases. Hydropathy and Chou-Fasman analyses reveal that the polypeptide chain of CTF/NF-I is likely to fold to higher order structures similar to those of the acetylases, and significant conservation of functionally important regions of the acetylases is observed in CTF/NF-I. It is therefore suggested that CTF/NF-I may function as a histone acetylase.

Nuclear factor I; Histone H1; CCAAT binding protein; Acetylase; Transcriptional regulation

1. INTRODUCTION

Eukaryotic chromatin involves two turns of DNA, approximately 140 bp [1], twined around a core nucleosome [2] consisting of pairs of histones 2A, 2B, 3 and 4 while histone H1 binds to the linker region, a site between the entering and leaving DNA strands [3]. It has been suggested that H1 functions as a general repressor [4], and that this structure may therefore be analogous to prokaryotic nucleoids, in which repressors bind to loops of DNA bent by HU proteins [5]. Nucleosomes aggregate upon inactivation of the eukaryotic chromatin, forming continuous solenoid structures, and this process probably restricts the access of RNA polymerase [6].

A novel DNA binding unit has very recently been characterized in histones, recognition of the especially narrow DNA minor groove of A + T-rich regions taking place by means of an SPKK motif in the protein [7–9]. This interaction may be of particular importance in the case of H2B, H3, H4 and H1, and may serve to direct the formation of nucleosomes into A + T-rich regions [9,10]. On the other hand, the homeodomain-like globular domain of H1 is thought to be capable of sequence-specific interaction with DNA through an α-turn-α motif [6], and it may appear, therefore, that the interaction of the H1 C-terminus with A + T-rich regions serves only to seal the nucleosome and gather the entering and leaving DNA strands.

Correspondence address: J. Oikarinen, Department of Medical Biochemistry, University of Oulu, Kajaanintie 52A, SF-90220 Oulu, Finland

Eukaryotic RNA polymerases are generally thought to be capable of negotiating reconstituted eukaryotic nucleosomes in vitro [11], although at a reduced rate when compared with a nucleosome-free template, in spite of the fact that the formation of nucleosomes evidently competes for the binding of transcriptional initiation complexes to the start site [12,13]. Nucleosomes are present in vivo in the active areas of chromatin in most cases [14], although it has been suggested that H2A and H2B may become dissociated in some situations [15] and one or two nucleosomes disappear from the transcriptional start site [16]. It is generally thought that H1 becomes dissociated in the active areas, or is more loosely bound to the nucleosome although still present [17]. As activation of chromatin is accompanied by modifications to the histones, including reversible acetylation of the SPKK motifs in the core histones [18–20], phosphorylation of a number of serines and threonines in H1 [22,23] and ubiquitination of H2A [24], and as introduction of these modifications has been demonstrated to impair interaction of the histones with DNA [25], the enzymes catalyzing their syntheses may be of crucial regulatory importance.

The primary structure of nuclear factor I (CTF/NF-I), a sequence-specific DNA binding protein that belongs to the family of CCAAT-binding proteins [26–29] and participates in the regulation of both DNA replication and transcription [30], has previously been demonstrated to display homology with the protein kinase (PK) family [31]. We concluded in a previous report that CTF/NF-I and the eukaryotic PKs may belong to a large gene family of regulatory proteins, the members of which are capable, to variable extents, of

Published by Elsevier Science Publishers B.V. (Biomedical Division)

recognizing a specific sequence on DNA and/or functioning as a PK. The primary structure of CTF/NF-I is compared here with sequences of prokaryotic origin in a genetic data bank and is observed to display homology with bacterial chloramphenicol acetyltransferases (CAT) [32–34]. The observation thus suggests that CTF/NF-I may function as an acetylase, being evolutionarily related to the PKs. The acetylation of histones by CCAAT-binding proteins may cause the loss of the nucleosome during gene activation observable at the start of transcription.

2. MATERIALS AND METHODS

The computer programs described by Pustell and Kafatos [35–37] were used for DNA and protein sequence analysis, and the GeneBank Genetic Sequence Data Bank to search for the various sequences. The Chou-Fasman analysis was performed as described previously [38].

3. RESULTS AND DISCUSSION

Cloning of cDNAs for rat [26] and human CTF/NF-I [27] and that of the corresponding porcine gene [28] has recently been reported. Several forms of the protein, consisting of a maximum of 499 amino acid residues, have been synthesized as a result of differential splicing of primary transcripts from a single gene [27] or multiple genes [29], and the N-terminal part, containing a region subject to differential splicing in the middle of the final protein product and comprising about 200 residues, displays homology with catalytic domains in the PK family [31]. This region is thought to comprise the domain responsible for the sequence-specific recognition of DNA [39,40].

The primary structure of CTF/NF-I was compared with sequences in a genetic data bank, and was observed by dot matrix analysis to display significant homology with CAT from a number of bacteria [32–34]. The homologous area covered the entire length of the acetylases and 200 residues in the middle of CTF/NF-I, those containing the region subject to differential splicing (Fig. 1). The extreme C-terminal parts of the acetylases displayed homology with a C-terminal region in CTF/NF-I, thus leaving a 120-residue stretch of non-homology between the two homologous areas, similar to the comparison between CTF/NF-I and the PKs [31]. The 120-residue area in CTF/NF-I is acidic, *pI* 6.3, in contrast to the N-terminal part of the protein, which is basic, *pI* 11.8, and may therefore perform a specialized function.

When the sequences of CAT and CTF/NF-I were aligned, major variations were observed in CTF/NF-I only in positions which varied even between the acetylases (Fig. 2). There are only two regions that are completely conserved in the various CAT proteins compared here, one with the sequence FTSFDLN from 154 to 160 in the *E. coli* variant and the other with the sequence HHAVCDG from 193 to 199. These sequences

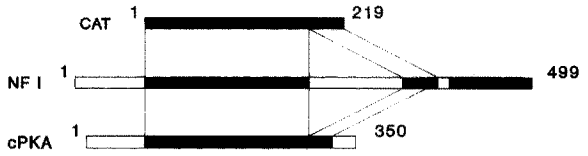

Fig. 1. Schematic representation of homologous regions in CTF/NF-I, CAT and the catalytic subunit of PKA. The black areas display homology and the hatched ones differential splicing from primary transcripts in the case of CTF/NF-I.

correspond to FSLGELQ and HTAIRYP in an analogous position in CTF/NF-I. It is also significant that the regions from 217 to 241 and 376 to 410 in CTF/NF-I, containing these conserved sequences, display respectively 32% and 27% identity when compared with the three CAT proteins.

Hydropathy (Fig. 3) and Chou-Fasman analyses (Fig. 2) revealed that CTF/NF-I may fold to higher order structures similar to those of the various CAT proteins, the plots for CTF/NF-I and *E. coli* CAT being remarkably similar. The primary structures of the 4 predicted α-helices in CTF/NF-I turned out to be without exception much less well conserved between the CAT variants, and CAT and CTF/NF-I, than those of the predicted turns and β-strands, although the locations of the helices approximately corresponded to those determined for CAT by crystallographic analysis of the type III variant of *E. coli* CAT [41].

Tertiary structures have been determined to date for a number of nucleotide-binding proteins including *E. coli* CAT, their characteristics being a core of several β-strands and a shield consisting of α-helices towards the aqueous phase [41]. It is significant that the two most well-conserved sequences between CAT and CTF/NF-I, FTSFDLN and HHAVCDG, define two conserved loops demonstrated to be involved in recognition and hydrolysis of acetyl-CoA by CAT [41]. The corresponding loops in CTF/NF-I may therefore take part in processing of acetyl-CoA by that protein. His-195 of the type III variant of *E. coli* CAT, this corresponding to the latter His residue in the sequence HHAVCDG, has been suggested to function as a general base catalyst in the acetyl-transfer reaction and is well-conserved in the CTF/NF-I variants.

The area undergoing differential splicing in CTF/NF-I may correspond to a loop in CAT which stabilizes the acetyl-CoA binding pouch, and may therefore confer a capability for processing acetyl-CoA on a portion of the CTF/NF-I molecules. The significance of the deletion of this region from some of the CTF/NF-I molecules remains to be elucidated, but it may be related to the dual role of CTF/NF-I, which is involved in the regulation of both DNA replication and the expression of different genes [30]. In the latter case a certain positioning of the CCAAT boxes is required for adequate transcription efficiency, and a pro-

Fig. 2. Alignment of the primary structures of human CTF/NF-I and CAT proteins from various bacteria. Single letter codes for amino acids are used and gaps are introduced for maximal alignment. The shaded residues denote homology or a conservative replacement, the amino acids being grouped (P, A, G, S, T), (V, I, M, L), (D, N, E, Q), (K, R, H) and (F, Y, W). The predicted helices in CTF/NF-I are depicted by horizontal bars above the sequences and the helices determined by crystallographic methods in the type III variant of *E. coli* CAT below the sequences. Asterisks denote the residues responsible for recognition and processing of acetyl-CoA by CAT [41]. The region in brackets is the area subjected to alternative splicing in the case of CTF/NF-I.

tein recognizing this motif may be of crucial importance for regulation purposes [42]. Differential splicing of certain regions of CTF/NF-I may most likely reflect variability in functionally unimportant areas of the protein, however.

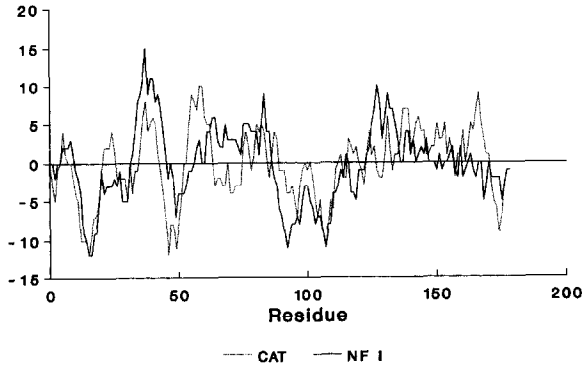

Fig. 3. Hydropathy plots of the homologous regions in human CTF/NF-I (continuous line) and *E. coli* CAT (broken line). The most hydrophobic areas rank positive.

It is therefore suggested here that at least a fraction of the CTF/NF-I molecules may function as histone acetylases in addition to recognizing a specific DNA sequence. Consistent with this rationale, a histone acetylase isolated previously displays several characteristics similar to CTF/NF-I [43]. This function of CTF/NF-I remains to be proved experimentally, but theoretically it is of the utmost interest, as it suggests that one mechanism for induction of the nucleosome loss during gene activation around the start of transcription [44,45] may be selective, directed protein lysine side-chain acetylation by sequence-specific DNA-binding proteins and consequent impairment of the DNA binding properties of the histones [18–20]. This nucleosome loss may in this way enable the access of other transcription factors and RNA polymerase to the start site [46].

The present observation thus lends further support to the hypothesis which posits the existence of a large family of eukaryotic regulatory proteins of common evolutionary origin [31], a number of which are capable

of sequence-specific binding to DNA through an α-turn-α motif, or specific recognition of exogenous proteins, and which display some nucleotide processing activity. This protein family may contain PKs with DNA-binding capability such as CKIIα and cdc2 or ones lacking in it, such as PKA, PKC and PKG, and also some acetylases with DNA-binding properties such as CTF/NF-I. It may also include various G proteins, both those anchored to the plasma membrane and histone H1, a nuclear G_α (Mannermaa and Oikarinen, unpublished results).

We have recently cloned the CTF/NF-I cDNA and are currently setting up a eukaryotic expression system for large-scale production of the corresponding protein in order to demonstrate the above enzymatic activity and to study the molecular structure of CTF/NF-I.

Acknowledgements: This work was supported in part by grants from the Medical Research Council of the Academy of Finland and the Finnish Cancer Research Foundation.

REFERENCES

[1] Morse, R.H. and Simpson, R.T. (1988) Cell 54, 285-287.
[2] Klug, A., Rhodes, D., Smith, J., Finch, J.T. and Thomas, J.O. (1980) Nature 287, 509-516.
[3] Staynov, D.Z. and Crane-Robinson, C. (1988) EMBO J. 7, 3685-3691.
[4] Weintraub, H. (1985) Cell 42, 705-711.
[5] Pettijohn, D.E. (1988) J. Biol. Chem. 263, 12793-12796.
[6] Mannermaa, R.-M. and Oikarinen, J. (1990) Biochem. Biophys. Res. Commun. 168, 254-260.
[7] Suzuki, M. (1989) EMBO J. 8, 797-804.
[8] Churchill, M.E.A. and Suzuki, M. (1989) EMBO J. 8, 4189-4195.
[9] Izaurralde, E., Käs, E. and Laemmli, U.K. (1989) J. Mol. Biol. 210, 573-585.
[10] Lukiw, W.J., Kruck, T.P.A. and McLachlan, D.R. (1989) FEBS Lett. 253, 59-62.
[11] Morse, R.H. (1989) EMBO J. 8, 2343-2351.
[12] Simpson, R.T. (1990) Nature 343, 387-389.
[13] Almouzni, G., Mechali, M. and Wolffe, A.P. (1990) EMBO J. 9, 573-582.
[14] Cordingley, M.G., Riegel, A.T. and Hager, G.L. (1987) Cell 48, 261-270.
[15] Gonzalez, P.J. and Palacian, E. (1989) J. Biol. Chem. 264, 18457-18462.
[16] Nacheva, G.A., Guschin, D.Y., Preobrazhenskaya, O.V., Karpov, V.L., Ebralidse, K.K. and Mirzabekov, A.D. (1989) Cell 58, 27-36.
[17] Ericsson, C., Grossbach, U., Björkroth, B. and Daneholt, B. (1990) Cell 60, 73-83.
[18] Profumo, A., Querzola, F. and Vidali, G. (1989) FEBS Lett. 250, 297-300.
[19] Hebbes, T.R., Thorne, A.W. and Crane-Robinson, C. (1988) EMBO J. 7, 1395-1402.
[20] Megee, P.C., Morgan, B.A., Mittman, B.A. and Smith, M.M. (1990) Science 247, 841-845.
[21] Csordas, A. (1990) Biochem. J. 265, 23-38.
[22] Hill, C.S., Parkman, L.C. and Thomas, J.O. (1990) EMBO J. 9, 805-813.
[23] Higurashi, M., Adachi, H. and Ohba, Y. (1987) J. Biol. Chem. 262, 13075-13080.
[24] Busch, H. (1984) Methods Enzymol. 106, 238-262.
[25] Norton, V.G., Imai, B.S., Yau, P. and Bradbury, E.M. (1989) Cell 57, 449-457.
[26] Paonessa, G., Gounari, F., Frank, R. and Cortese, R. (1988) EMBO J. 7, 3115-3123.
[27] Santoro, C., Mermod, N., Andrews, P.C. and Tjian, R. (1988) Nature 334, 218-224.
[28] Meisterernst, M., Rogge, L., Donath, C., Gander, I., Lottspeich, F., Mertz, R., Dobner, T., Föckler, R., Stelzer, G. and Winnacker, E.-L. (1988) FEBS Lett. 236, 27-32.
[29] Rupp, R.A.W., Kruse, U., Multhaup, G., Göbel, U., Beyreuther, K. and Sippel, A.E. (1990) Nucleic Acids Res. 18, 2607-2616.
[30] Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J. and Tjian, R. (1987) Cell 48, 79-89.
[31] Mannermaa, R.-M. and Oikarinen, J. (1989) Biochem. Biophys. Res. Commun. 162, 427-434.
[32] Cohen, J.D., Abrams, E., Eccleshall, T.R., Buchferer, B. and Marmur, J. (1983) Mol. Gen. Genet. 191, 451-459.
[33] Mongkolsuk, S., Duvall, E.J. and Lovett, P.S. (1985) Gene 37, 83-90.
[34] Bruckner, R. and Matzura, H. (1985) EMBO J. 4, 2295-2300.
[35] Pustell, J. and Kafatos, F.C. (1982) Nucleic Acids Res. 10, 51-59.
[36] Pustell, J. and Kafatos, F.C. (1984) Nucleic Acids Res. 12, 643-655.
[37] Pustell, J. and Kafatos, F.C. (1986) Nucleic Acids Res. 14, 479-488.
[38] Chou, P.Y. and Fasman, G.D. (1978) Annu. Rev. Biochem. 47, 251-276.
[39] Mermod, N., O'Neill, E.A., Kelly, T.J. and Tjian, R. (1989) Cell 58, 741-753.
[40] Gounari, F., De Francesco, R., Schmitt, J., van der Vliet, P.C., Cortese, R. and Stunnenberg, H. (1990) EMBO J. 9, 559-566.
[41] Leslie, A.G.W., Moody, P.C.E. and Shaw, W.V. (1988) Proc. Natl. Acad. Sci. USA 85, 4133-4137.
[42] McKnight, S.L., Lane, M.D. and Gluecksohn-Waelsch, S. (1989) Gen. Dev. 3, 2021-2024.
[43] Kelner, D.N. and McCarty, K.S. (1984) J. Biol. Chem. 259, 3413-3419.
[44] Han, M. and Grunstein, M. (1988) Cell 55, 1137-1145.
[45] Wolffe, A.P. (1989) EMBO J. 8, 527-538.
[46] Cheng, L. and Kelly, T.J. (1989) Cell 59, 541-551.