Vol. 162, No. 1, 1989 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 14, 1989 Pages 427-434

# HOMOLOGY OF NUCLEAR FACTOR I WITH THE PROTEIN KINASE FAMILY

Riitta-Maaria Mannermaa and Jouko Oikarinen

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

Received May 31, 1989

It is observed that a putative DNA binding domain in nuclear factor I (NF I) which is a eukaryotic sequence-specific DNA binding protein participating in both regulation of DNA replication and transcription displays sequence homology with catalytic domains in various protein kinases. In addition, hydropathy analysis reveals that the NF I polypeptide chain is likely to fold into similar secondary and tertiary structures to those of the protein kinases. Although it is not known whether NF I functions as a protein kinase in addition to recognizing a specific sequence on DNA, it is concluded that some of the eukaryotic DNA binding proteins and protein kinases may belong to a gene family, various members of which are evolutionarily related and responsible for the regulation of metabolism in diverse compartments of the cell. The present finding also suggests that a number of kinases may to varying extents be capable of direct interaction with DNA. 

Nuclear factor I (NF I) is a eukaryotic sequence-specific DNA binding protein that preferentially recognizes the 5'-TTGGCAnnnTGCCAA-3' motif displaying a perfect dyad of symmetry (1 - 3). The protein participates both in regulation of DNA replication and transcription of a number of host cell and viral genes (4, 5). Both synthetic oligonucleotides and point mutagenesis to a putative consensus sequence have been used to reveal the recognition sequence. Several reports have recently been published on the purification of proteins that bind to this motif (6, 7), and it is evident that, in addition to NF I, histone H1 binds to the same motif or a similar one (8).

Three reports have recently been published on the cloning and sequencing of cDNAs (9, 10) or the gene of NF I (11), and on amino acid sequences deduced from these experiments. The protein contains maximally about 500 amino acids, but shorter forms are encountered due to differ-

0006-291X/89 $1.50

ential splicing of primary transcripts. The areas conserved in the different subtypes comprise a basic aminoterminal domain of about 240 residues, thought to be responsible for binding to the specific sequence on DNA, an acidic domain in the middle and a proline-rich carboxy-terminal domain.

The primary structure of the putative DNA binding domain in NF I is here compared with sequences in a genetic data bank and homology is observed with catalytic domains in various protein kinases, thus suggesting a common evolutionary origin for the two entities.

## MATERIALS AND METHODS

The computer programs described by Pustell and Kafatos (12 - 14) were used for DNA and protein sequence analysis, and the GeneBank$^R$ Genetic Sequence Data Bank to search for various sequences.

## RESULTS AND DISCUSSION

The primary structures of rat, porcine and human NF I were compared with those of a number of protein kinases. Interestingly, dot matrix analysis demonstrated considerable homology between the putative DNA binding domain in NF I, covering 240 amino acid residues from the amino terminus, and catalytic domains for ATP hydrolysis in the catalytic subunit of protein kinase A (cPKA-α), protein kinase G (cPKG) and casein kinase (CKIIα), for example (Fig. 1). When the corresponding sequences were aligned it was observed that major variations between NF I and the protein kinases take place in areas which are variable even between different kinases. Another short area of homology was observed in the carboxyterminal part of NF I, but the two areas formed a continuous stretch in the kinases.

The protein kinases comprise a large family of eukaryotic regulatory proteins (15) catalyzing transfer of the terminal phosphate from ATP to an acceptor amino acid residue in a polypeptide chain. This covalent modification subsequently serves many regulatory functions. The cata-

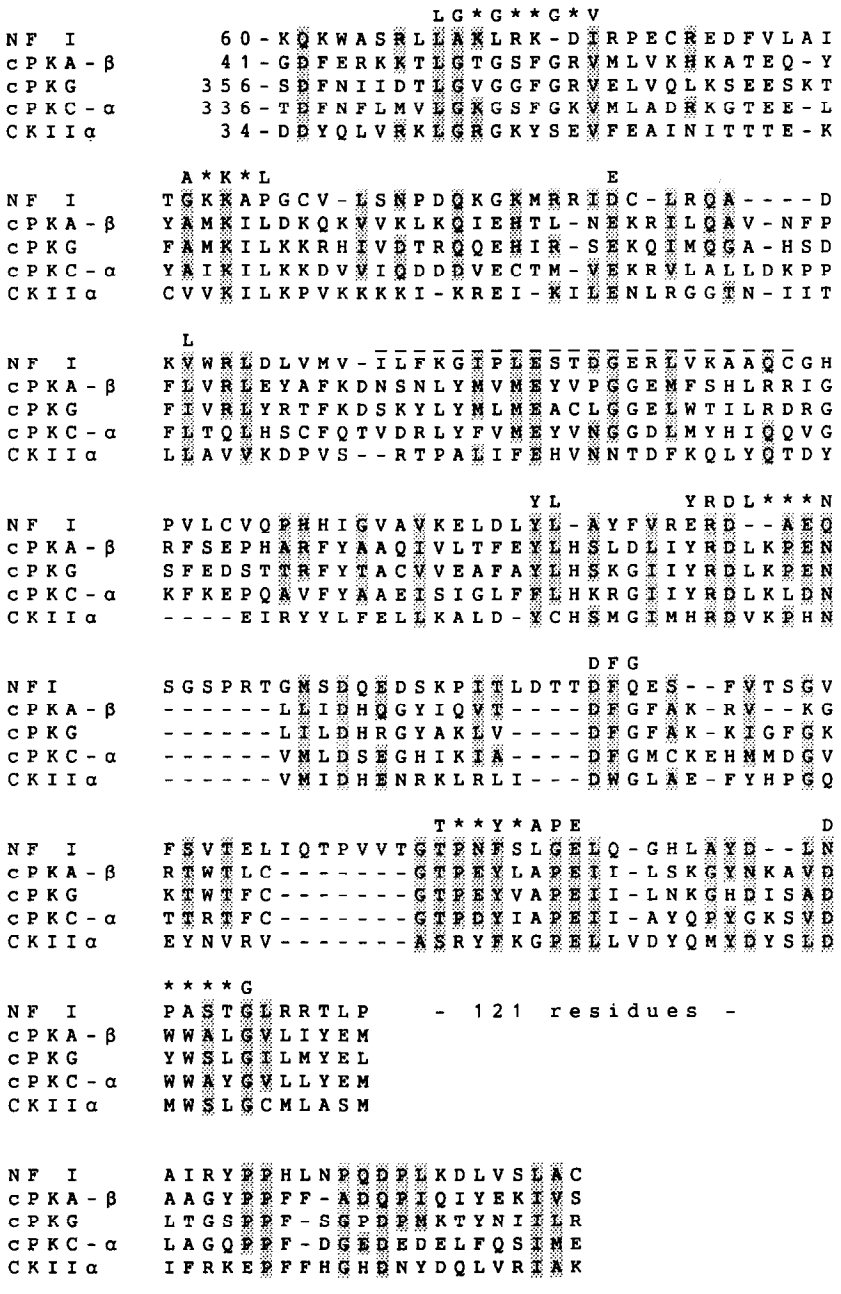


Fig.1 Alignment of the primary structures of human nuclear factor I (NF I), the catalytic subunit of human protein kinase A (cPKA-β), human protein kinase G (cPKG), human protein kinase C (cPKC-α) and Drosophila casein kinase II (CKIIα). The single letter codes for amino acids are used, and gaps are introduced for maximal alignment. The hatched residues display homology or conservative displacements in that position in at least three of the five polypeptide chains when compared with the NF I chain, the residues being grouped (V, I, M, L), (G, P, A, S, T), (E, D, N, Q), (F, Y, W) and (K, R, H). The residues above the five sequences compared denote those generally conserved in protein kinases, and the overlined area depicts a putative helix-turn-helix structure, as described in the legend to Fig. 3. A star denotes any amino acid.

lytic sites of all the protein kinases characterized so far from yeast to human display reasonable amino acid homology, and this 200 amino acid structure may therefore be considered indispensable for kinase activity. In addition, a number of the kinases contain regulatory ligand binding domains, such as those for cAMP, cGMP, diacylglycerol and $Ca^{2+}$, located in the aminoterminal parts of the molecules. In some cases the regulatory regions may be encoded by separate genes, including the regulatory subunits for PKA (rPKA) (16, 17). These are obviously evolutionarily related to the protein kinases but have diverged to perform only one of the several functions of the whole peptide.

Hydropathy analysis of the polypeptide chains of various kinases and NF I revealed considerable conservation of hydrophilic and hydrophobic areas, thus suggesting similar folding properties (Fig. 2). There are no conserved cysteines in the catalytic domains of the protein kinases, and therefore no conservation can be expected between the kinases and NF I. The positions of residues capable of forming reverse turns, prolines and glycines, displayed conservation, however, thus suggesting similar helical structures for NF I and the catalytic domains of the protein kinases.

Although several reports exist on the purification of NF I, it is not known whether it functions as a protein kinase. We have recently demonstrated that a protein preparation isolated from rat liver by DNA recognition site affinity chromatography using the NF I consensus sequence as the ligand contains a protein kinase, but it is the other protein, histone H1, recognizing the same motif as NF I, that seems to be responsible for the activity (18). At the moment, therefore, only the capability of NF I for sequence-specific binding to DNA is indisputable, having been demonstrated by virtue of a fusion protein produced in bacteria (10).

The finding that the DNA binding domain of a regulatory DNA binding protein displays homology with the catalytic domains in protein kinases

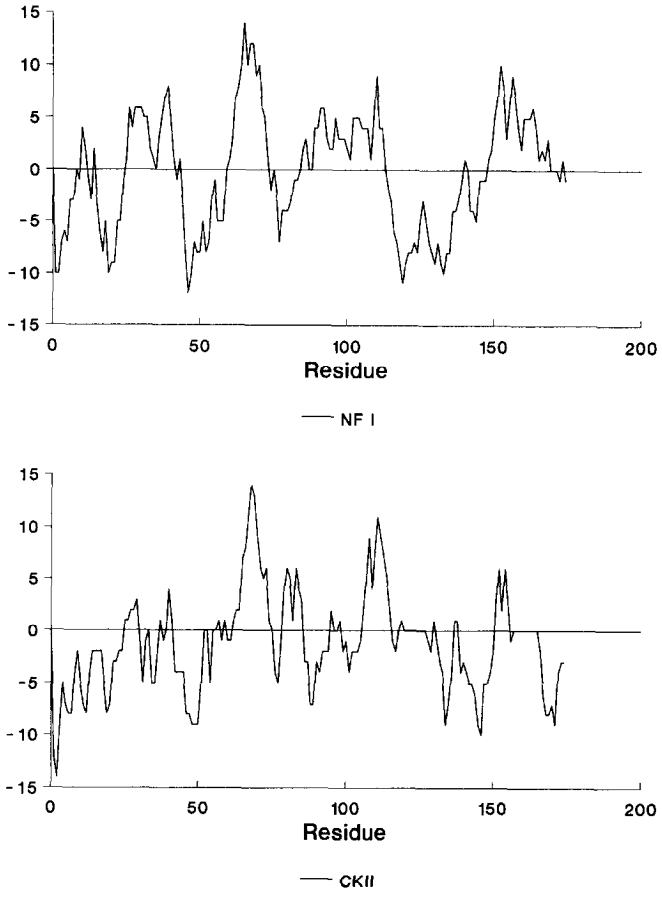


**Fig.2** Hydropathy analysis of the homologous region in human NF I and Drosophila CKIIα. Most hydrophobic areas rank positive.

raises interesting questions. Firstly, what is the evolutionary relationship between the DNA binding proteins and protein kinases, and secondly, are some of the protein kinases capable of sequence-specific binding to DNA? Several lines of evidence, covering organisms from bacteria to human beings, already exist suggesting that all these proteins may belong to a large gene family, members of which have evolved from a primordial DNA binding protein inducible by an exogenous ligand, and that they may variably either recognize a DNA sequence, function as a protein kinase or be capable of both functions.

The α-carbon backbone of H1 has previously been demonstrated crystallographically to be superimposable on the DNA binding domain in

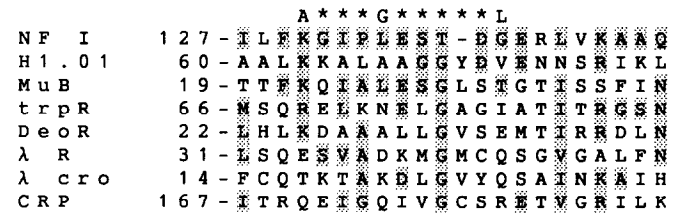


**Fig.3** Alignment of a region in NF I with helix-turn-helix domains in chicken histone H1.01 and a number of prokaryotic DNA binding proteins. The hatched residues display homology or conservative displacements. The deduced consensus amino acid sequence for the helix-turn-helix structures, as suggested previously (21), is shown above the eight sequences.

bacterial cAMP receptor protein (CRP) containing a helix-turn-helix structure (20) The primary structures of the putative DNA binding domain in NF I and the globular domain in H1 (19), were here compared with the helix-turn-helix structures in a number of bacterial DNA binding proteins (21) (Fig. 3). Consequently, putative domains were identified in NF I and H1 that may be responsible for the sequence-specific interaction in the major groove of the DNA secondary structure. The corresponding region displays conservation even in the protein kinases, thus confirming its functional importance (Fig. 1).

If the DNA binding domains and the catalytic domains of the kinases are related, the former must have appeared prior to the latter during evolution, since no kinase so far has been characterized in prokaryotes. However, in the light of the observation that CRP and cPKA display amino acid homology (22), together with the observations that both NF I and H1 bind to the 5'-TTGGCAnnnTGCCAA-3' motif on DNA and that the three-dimensional structures of the globular domain in H1 and the DNA binding domain in CRP are similar, it can be deduced indirectly that CRP may belong to the same gene family while being capable only of sequence-specific binding to DNA. The fact that CRP contains a cAMP binding domain in the aminoterminal part of the molecule in the same position where a number of the eukaryotic kinases contain a ligand binding domain (23, 24) (Fig. 4), lends further support to the hypothesis of a common origin for NF I, H1, CRP and the protein kinases.

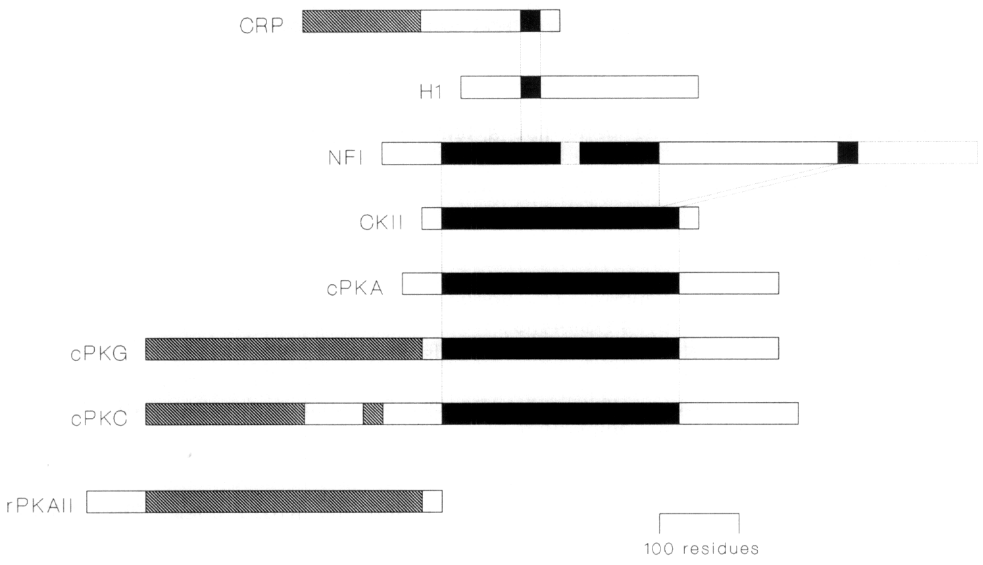


**Fig.4** A schematic representation of the homologous areas in NF I, H1, CRP and various protein kinases. The black areas in the kinases denote the catalytic domains and that in NF I the homologous region, as depicted in Fig. 1. The black areas in H1 and CRP denote the helix-turn-helix structure responsible for DNA recognition and its relationship to the catalytic domain in the kinases. The striated areas in the various proteins denote ligand binding domains. rPKAII, the regulatory subunit of protein kinase A.

The question of the extent to which various protein kinases are capable of direct interaction with DNA remains to be answered, but several hints of such a capability already exist. A cAMP responsive element binding protein (CREB) has been purified (25, 26), but it is different from cPKA (27). However, the present sequence comparisons and some pieces of experimental evidence suggest that cPKA may bind to a specific sequence on DNA (28) and thereby at least partly elicit the effects of cAMP on responsive genes.

Protein kinases are ubiquitous in all cell types from yeasts to higher organisms, participating in regulatory functions in diverse compartments of the cell. Several kinases have been characterized which, after completion of their synthesis, will be primarily located in the nucleus and function there, while others will be directed by hydrophobic signals to the cell membrane or stay in the cytoplasm. On the basis of the present data we suggest that a number of prokaryotic and eukaryotic

DNA binding regulatory proteins and these eukaryotic protein kinases belong to the same large gene family of regulatory proteins, which have evolved from a primordial gene for a DNA binding protein, and that their two functions may in many cases be inseparable.

REFERENCES


1. Meisterernst, M., Gander, I., Rogge, L., and Winnacker, E.-L. (1988) Nucl. Acids Res. 16, 4419-4435.
2. Gronostajski, R.M. (1986) Nucl. Acids Res. 14, 9117-9132.
3. Gronostajski, R.M. (1987) Nucl. Acids Res. 15, 5545-5560.
4. Gronostajski, R.M., Knox, J., Berry, D., and Miyamoto, N.G. (1988) Nucl. Acids Res. 16, 2087-2098.
5. Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J., and Tjian, R. (1987) Cell 48, 79-89.
6. Rosenfeld, P.J., and Kelly, T.J. (1986) J. Biol. Chem. 261, 1398-1408.
7. Rupp, R.A.W., and Sippel, A.E. (1987) Nucl. Acids Res. 15, 9707-9726.
8. Ristiniemi, J., and Oikarinen, J. (1989) J. Biol. Chem. 264, 2164-2174.
9. Paonessa, G., Gounari, F., Frank, R., and Cortese, R. (1988) EMBO J. 7, 3115-3123.
10. Santoro, C., Mermod, N., Andrews, P.C., and Tjian, R. (1988) Nature 334, 218-224.
11. 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.
12. Pustell, J., and Kafatos, F.C. (1982) Nucl. Acids Res. 10, 51-59.
13. Pustell, J., and Kafatos, F.C. (1984) Nucl. Acids Res. 12, 643-655.
14. Pustell, J., and Kafatos, F.C. (1986) Nucl. Acids Res. 14, 479-488.
15. Hanks S.K., Quinn, A.M., and Hunter, T. (1988) Science 241, 42-52.
16. Titani, K., Sasagawa, T., Ericsson, L.H., Kumar, S., Smith, S.B., Krebs, E.G., and Walsh, K.A. (1984) Biochemistry 23, 4193-4199.
17. Takio, K., Smith, S.B., Krebs, E.G., Walsh, K.A., and Titani, K. (1984) Biochemistry 23, 4200-4206.
18. Mannermaa, R.-M., and Oikarinen, J. Submitted for publication.
19. Clore, G.M., Gronenborn, A.M., Nilges, M., Sukumaran, D.K., and Zarbock, J. (1987) EMBO J. 6, 1833-1842.
20. de Crombrugghe, B., Busby, S., and Buc, H. (1984) Science 224, 831-838.
21. Brennan, R.G., and Matthews, B.W. (1989) J. Biol. Chem. 264, 1903-1906.
22. Weber, I., Takio. K., Titani, K., and Steitz, J.A. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 7679-7683.
23. Takio, K., Wade, R.D., Smith, S.B., Krebs, E.G., Walsh, K.A., and Titani, K. (1984) Biochemistry 23, 4207-4218.
24. Parker, P.J., Coussens, L., Totty, N., Rhee, L., Young, S., Chen, E., Stabel, S., Waterfield, M.D., and Ullrich, A. (1986) Science 233, 853-866.
25. Roesler, W.J., Vandenbark, G.R., and Hanson, R.W. (1988) J. Biol. Chem. 263, 9063-9066.
26. Yamamoto, K.K., Gonzalez, G.A., Biggs III, W.H., and Montminy, M.R. (1988) Nature 334, 494-498.
27. Gonzalez, G.A., Yamamoto, K.K., Fischer, W.H., Karr, D., Menzel, P., Biggs III, W., Vale, W.W., and Montminy, M.R. (1989) Nature 337, 749-752.
28. Riabowol, K.T., Fink, S., Gilman, M.Z., Walsh, D.A., Goodman, R.H., Feramisco, J.R. (1988) Nature 336, 83-86.