Volume 282, number 1, 175–177 FEBS 09648 April 1991
© 1991 Federation of European Biochemical Societies 0014-5793/91/$3.50
ADONIS 001457939100356A

Sequence similarity between opioid peptide precursors and DNA-binding proteins

Georgy Ya. Bakalkin[1,2], Dmitry Ponomariev[2], Raphik A. Sarkisyan[2] and Lars Terenius[1]

[1] *Department of Drug Dependence Research, Karolinska Institute, S-104 01 Stockholm, Sweden* and [2] *National Cardiology Research Center, Moscow, 121552, USSR*

Received 4 February 1991

The opioid peptide precursors, preprodynorphin and preproenkephalin show structure similarity with a transcription factor, *hunchback* and the putative helix-loop-helix DNA-binding proteins, *lil-1*, *tal* and *twist*. Segments with similarity contain the three enkephalin sequences in preprodynorphin and one in preproenkephalin which are present within heptapeptide repeats characteristic of an α-helical coiled-coil structure distinctive of an amphipathic helix-loop-helix DNA-binding motif. *Hunchback* and the opioid prohormones also have cystein-rich regions characteristic of zinc-finger domains in common.

Opioid precursor; DNA-binding protein; Sequence homology

1. INTRODUCTION

Like other regulatory peptides, the opioid peptides which include enkephalins, endorphins and dynorphins are synthesized as high molecular weight precursors containing the bioactive peptides and 'spacer' segments [1–3]. We show here that amino acid and nucleotide sequences of preprodynorphin, preproenkephalin and several eukaryotic transcription factors display striking similarity. The similarities cover the enkephalin sequences and structural motifs which have been proposed for DNA-binding domains of transcription factors, the helix-loop-helix and zinc-finger [4–6].

2. RESULTS AND DISCUSSION

A comparison of the opioid precursors and transcription factors reveals sequence similarity between some segments of these proteins. Fig. 1 shows an alignment of the amino acid (A, C) and nucleotide (B, D) sequences of bovine preproenkephalin (A, B) and human preprodynorphin (C, D) segments with segments of the helix-loop-helix DNA-binding proteins *lyl-1* [7], *tal* [8], *twist* [9], and the transcription factor *hunchback* [10]. The preproenkephalin (70–103) segment is similar to *lyl-1* (159–191) and to a lesser extent to *tal* (54–86) or *twist* (397–415) (Fig. 1A, B), while human preprodynorphin (175–231) and (502–600) are related to *hunchback* (92–148) and (5032–5130), respectively (Fig. 1C, D). The observed relationship was found to be statistically significant (see legend Fig. 1). It should be pointed out that the segments of opioid peptide precursors showing similarity to DNA-binding proteins include three enkephalin sequences in preprodynorphin and one in preproenkephalin.

The opioid peptide precursors have characteristic heptapeptide repeats including all enkephalin sequences with alternating hydrophobic and hydrophilic amino acid residues (Fig. 2). If the amino acid residues in heptapeptide repeats are designated a, b, c, d, e, f and g, the residues a and d are hydrophobic and the residues b, c, e, f and g are predominantly hydrophilic. It is well known that a link of heptapeptide repeats is a distinctive feature of an α-helical coiled coil structure [18] in which the helices interact through their hydrophobic surfaces formed by a and d residues on one side of the helix. The characteristic heptapeptide pattern has been observed in the potential amphipathic helix-loop-helix DNA-binding motif of several proteins involved in the control of cellular proliferation and differentiation [5,7,8]. It has been suggested that the helix-loop-helix motif is crucial for both the dimerization of amphipathic helices and their binding to DNA. It should be mentioned that DNA-binding protein segments with similarity to preproenkephalin (region of helix II of *lyl-1*, *tal*, and *twist* [5,7–9]) and preprodynorphin (*hunchback* segment) are present within characteristic heptapeptide repeats ($P < 0.03$).

Further examination of preproenkephalin and preprodynorphin sequences reveals the presence of cysteine-rich regions in the N-terminus with amino acid sequences $Cys\text{-}X_3\text{-}Cys\text{-}X_2\text{-}Cys\text{-}X_{14}\text{-}Cys\text{-}X_3\text{-}Cys$ and

Correspondence address: G.Ya. Bakalkin, Department of Drug Dependence Research, Karolinska Institute, Box 60500, S-10401 Stockholm, Sweden. Fax: (46) (8) 341939.

Fig. 1. Sequence similarities between opioid peptide precursors and DNA-binding proteins. A segment of bovine preproenkephalin [1] is compared with segments of *lyl-1*, *tal* and *twist* (A, B) and human preprodynorphin [2] with *hunchback* (C, D) at the amino acid or nucleotide level, respectively. The enkephalin sequences are indicated by solid lines. The alignment was found using the Needleman-Wunsch algorithm [11]. Note that all possible preproenkephalin (70–103) segment alignments along the *lyl-1* protein, and all possible *lyl-1* (159–191) segment alignments along preproenkephalin, preprodynorphin (175–231) segment alignments along *hunchback* protein, as well as *hunchback* (92–148) segment alignments along preprodynorphin were tested. All four pairwise comparisons of proteins gave only one alignment with significant similarity (depicted in Fig. 1). The probability score was computed taking into account the number of identical pairs and the number of gaps (with the gap penalty −3 for the amino acid (for example see [12]) and −5 for the nucleotide sequence (for example see [13]). For comparison 1000 random sequences were generated by computer and z-values [z = (observed score − mean of shuffled scores)/standard deviation of shuffled scores] were calculated. The z-values were for preproenkephalin/*lyl-1* amino acid and nucleotide sequences −5.27 and 7.62, respectively; for preprodynorphin/*hunchback* −8.75 and 7.39. According to Doolittle [12] who employed a value of $z > 3$ as the cut-off for significance (z measure based solely on aligned identities), the observed z-values can be expected to represent authentic relationships. Assuming random z'-values [$z' = \ln(1 + z - z_{min})$, where z_{min} is the smallest z among all shuffled scores] to be normally distributed [14], the significance (P) of differences between the expected and observed z' was evaluated by Student's t-test; for preproenkephalin/*lyl-1*, $P < 0.001$ (amino acid sequences) and $P < 0.00002$ (nucleotide sequences); for preprodynorphin/*hunchback*, $P < 0.00002$ (amino acid sequences) and $P < 0.00005$ (nucleotide sequences) were recorded. Examination of the highest scores of the 1000 shuffled sequences showed that no random score approached the observed score in any of the 4 cases. According to Pearson and Lipman [15] it is unlikely that the observed similarity became significant only by chance. A quantitative estimate of a one-side confidence limit for a binomial distribution [16] which is independent of the z-distribution, it might be affirmed with a significance level of 0.99 that the probability of the observed z-values (viz. 5.27, 7.62, 7.39 and 8.75) is fairly small ($P < 0.005$). It was evaluated by formula and values of one-sided confidence limit for a binomially distributed population [16] and is independent of the z-distribution form.

Cys-X_3-Cys-X_2-Cys-X_{15}-Cys-X_3-Cys, respectively. The cysteines of these regions are perfectly conserved in opioid peptide precursors of different species [2]. It is noteworthy that the cysteine-rich regions of the precursors perfectly match a test filter (Cys-X_{2-4}-Cys-X_{2-15}-Cys-X_{2-4}-Cys) that was devised to locate potential metal-binding sites and to identify possible zinc-finger domains in proteins [19]. Cysteine-rich regions of proteins, particularly zinc-finger domains, are structural motifs proposed for binding to DNA and RNA [4,6,20,21].

Besides a DNA-binding domain, eukaryotic transcription factors contain a transcription activation region that is generally located in a separate part of the protein. The transcription activation functions are contained in short (30–100 residues) acidic fragments with minimum amino acid sequence similarity [4,6]. These fragments may adopt amphipathic α-helices with one face containing negatively charged residues and the other hydrophobic residues. Inspection of opioid peptide precursors reveals a 38-amino acid domain in preproenkephalin (preproenkephalin 140–177) containing 32% Glu and/or Asp and a 29-amino acid domain in preprodynorphin (preprodynorphin 139–167) containing 31% Glu and/or Asp. The acidic and hydrophobic amino acids in these domains are arrang-

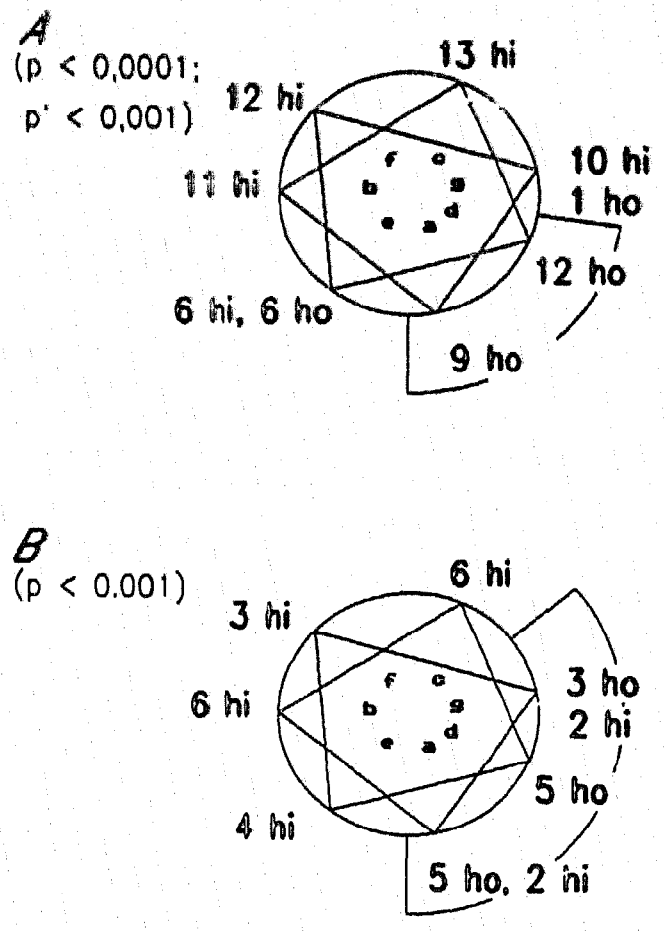

Fig. 2. Heptapeptide repeats in enkephalin containing fragments of preproenkephalin (A) and preprodynorphin (B). The amino acid sequence of bovine preproenkephalin (91–101, 102–113, 127–142, 176–191, 200–215, 220–235, 252–263 amino acid residues) and human preprodynorphin (171–182, 203–215, 222–233) fragments is displayed on a helical wheel. The face with hydrophobic amino acids is bracketed. The significance of differences (P, P') between the expected and observed frequencies of hydrophobic (Tyr, Phe, Leu, Trp, Met, Val, Ile) and hydrophilic amino acid residues at (a and d) and (b, c, e, f and g) positions was evaluated with the χ^2-test ($r = 1$; [17]) (P'-enkephalin sequences with adjacent basic residues were excluded from enkephalin fragments). A Tyr residue was placed at position a (preproenkephalin fragments) or d (preprodynorphin fragments) of the heptapeptide repeat. Abbreviations: ho, hydrophobic and hi, hydrophilic residues. Numbers indicate the number of residues at a position.

ed in such a way that they can form an amphipathic α-helix in which negative charges are predominantly localized along one surface of the helix while hydrophobic residues are localized along the other.

The amino acid and nucleotide sequence similarity of the opioid precursors with *lyl-1* and related proteins and *hunchback*, as well as the presence of heptapeptide repeats, zinc-finger-like segments and acidic regions in preproenkephalin and preprodynorphin point to a common evolutionary origin of these groups of proteins.

Acknowledgements: The authors thank Mr D. Sarkisyan for programming and Dr I. Novikov for consultations.

REFERENCES

[1] Noda, M., Furutani, Y., Takahashi, H., Toyosato, M., Hirose, T., Inayama, S., Nakanishi, S. and Numa, S. (1982) Nature 295, 202–206.
[2] Horikawa, S., Takai, T., Toyosato, M., Takahashi, H., Noda, M., Kakidani, H., Kubo, T., Hirose, T., Inayama, S., Hayashida, H., Miyata, T. and Numa, S. (1983) Nature 306, 611–614.
[3] Udenfriend, S. and Kilpatrick, D.L. (1983) Arch. Biochem. Biophys. 221, 309–323.
[4] Mitchell, P. and Tjian, R. (1989) Science 245, 371–378.
[5] Murre, C., McCaw, P.S. and Baltimore, D. (1989) Cell 56, 777–783.
[6] Struhl, K. (1989) Trends Biol. Sci. 14, 137–140.
[7] Mellentin, J.D., Smith, S. and Cleary, M. (1989) Cell 58, 77–83.
[8] Chen, Q., Cheng, J.-T., Tsai, L.-H., Schneider, N., Buchanan, G., Carrol, A., Crist, W., Ozanne, B., Siciliano, M.J. and Baer, R. (1990) EMBO J. 9, 415–424.
[9] Thisse, B., Stoetzel, C., Gorostiza-Thisse, C. and Perrin-Schmitt, F. (1988) EMBO J. 7, 2175–2183.
[10] Tautz, D., Lehmann, R., Schnurch, Schuh, R., Seifert, E., Kienlin, A., Jones, K. and Jackle, H. (1987) Nature 327, 383–389.
[11] Needleman, S.B. and Wunsch, C.D. (1970) J. Mol. Biol. 48, 443–453.
[12] Doolittle, R.F. (1981) Science 214, 149–159.
[13] Kumar, C.S., Muthukumaran, G., Frost, L.J., Noe, M., Ahn, Y.H., Mariano, T.M. and Pestra, S. (1989) J. Biol. Chem. 264, 17939–17946.
[14] Wilbur, W.J. and Lipman, D.J. (1983) Proc. Natl. Acad. Sci. USA 80, 726–730.
[15] Pearson, W.R. and Lipman, D.J. (1988) Proc. Natl. Acad. Sci. USA 85, 2444–2448.
[16] Sachs, L. (1984) Applied Statistics. A Handbook of Techniques. Springer Series in Statistics, p. 336 (formula 4.24).
[17] Cramer, H. (1946) Mathematical Methods of Statistics. Princeton Mathematical Series 9 (Morse, M., Robertson, H.P. and Tucker, A.W., eds), Almqvist and Wiksells Akademiska Handbocker, Gebers, p. 446 (formula 30.5.3).
[18] Cohen, C. and Parry, D.A.D. (1986) Trends Biol. Sci. 11, 245–248.
[19] Berg, J.M. (1986) Science 232, 485–487.
[20] Garcia, J.A., Harrich, D., Pearson, L., Mitsuyasu, R. and Gaynor, R.B. (1988) EMBO J. 7, 3143–3147.
[21] Gorelick, R.J., Henderson, L.E., Hanser, J.P. and Rein, A. (1988) Proc. Natl. Acad. Sci. USA 85, 8420–8424.