Volume 268, number 1, 231–234 FEBS 08705 July 1990

Related domains in yeast tRNA ligase, bacteriophage T4 polynucleotide kinase and RNA ligase, and mammalian myelin 2′,3′-cyclic nucleotide phosphohydrolase revealed by amino acid squence comparison

Eugene V. Koonin[1] and Alexander E. Gorbalenya[2]

[1]*Institute of Microbiology, The USSR Academy of Sciences, 7 Prospekt 60let Oktyabrya, 117811 Moscow* and [2]*Institute of Poliomyelitis and Viral Encephalitides, USSR Academy of Medical Sciences, 142782 Moscow Region, USSR*

Received 31 May 1990

Related domains containing the purine NTP-binding sequence pattern have been revealed in two enzymes involved in tRNA processing, yeast tRNA ligase and phage T4 polynucleotide kinase, and in one of the major proteins of mammalian nerve myelin sheath, 2′,3′-cyclic nucleotide 3′-phosphohydrolase (CNPase). It is suggested that, similarly to the tRNA processing enzymes, CNPase possesses polynucleotide kinase activity, in addition to the phosphohydrolase one. It is speculated that CNPase may be an authentic mammalian polynucleotide kinase recruited as a structural component of the myelin sheath, analogously to the eye lens crystallins. Significant sequence similarity was revealed also between the N-terminal regions of yeast tRNA ligase and phage T4 RNA ligase. A tentative scheme of the domainal organizations for the three complex enzymes is proposed. According to this model, tRNA ligase contains at least three functional domains, in the order: N-ligase-kinase-phosphohydrolase-C, whereas polynucleotide kinase and CNPase encompass only the two C-terminal domains in the same order.

CNPase; Myelin; Polynucleotide kinase; tRNA ligase; tRNA splicing

1. INTRODUCTION

Yeast tRNA ligase is a multifunctional enzyme involved in tRNA splicing. In addition to the ligase activity as such, it displays the activities of polynucleotide kinase (PNK) and 2′,3′-cyclic nucleotide 3′-phosphohydrolase (CNPase) required for the complex reaction of the joining of tRNA half-molecules [1]. The three enzymatic activities of tRNA ligase are equal to the combined activities of two enzymes of bacteriophage T4, PNK and tRNA ligase, involved in tRNA reprocessing in some strains of *E. coli* [2]. Surprisingly, one of these activities, that of the CNPase, has also been observed in an abundant structural protein of the mammalian nerve myelin sheath [3]. It was of interest to compare the amino acid sequences of all these proteins to identify putative domains responsible for distinct enzymatic activities, and to gain insight into possible function(s) of the myelin CNPase.

2. METHODS

2.1. *Amino acid sequences and their comparison*

The sequences were from references [4] (rat CNPase), [5] (bovine CNPase), [6] (human CNPase), [7] (T4 PNK), [8] (yeast tRNA ligase), and [9] (T4 RNA ligase). Initial search for similar sequence segments was by program DOTHELIX which is a version of the dot matrix method generating the full local similarity map and allowing automatic determination of the optimal length of similar segments [10,11]. Detailed alignments were done by program OPTAL [12]; some of the alignments were subsequently modified using the program MULTALIN [13]. Screening of the Swissprot database (release 12) for sequence patterns was by program SITE [14].

Correspondence address: E.V. Koonin, Institute of Microbiology, The USSR Academy of Sciences, 7 Prospekt 60let Oktyabrya, 117811 Moscow, USSR

Abbreviations: CNPase, 2′,3′-cyclic nucleotide 3′-phosphohydrolase; PNK, polynucleotide kinase

3. RESULTS AND DISCUSSION

3.1. *The relationships between the sequences of myelin CNPases and tRNA processing enzymes*

Inspection of the amino acid sequences of CNPases of different species revealed, in their amino terminal domains, the purine NTP-binding pattern [15,16] found previously in the PNK sequence too [7,16]. A somewhat deviant form of the same pattern was identified in the central part of the tRNA ligase sequence. This underscored the analogy between CNPase and the tRNA processing enzymes and prompted further detailed sequence comparisons.

Comparison of the sequences of putative NTP-binding domains of the CNPases with those of the tRNA processing enzymes showed statistical significance of over 5 SD above the mean of 25 random simulations. As shown in Fig. 1, impressive similarity was observed in the segments surrounding the 'A' and 'B' motifs of the NTP-binding pattern, especially between the CNPases and PNK around the 'A' motif (10 identical residues of 12), and between CNPases and

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

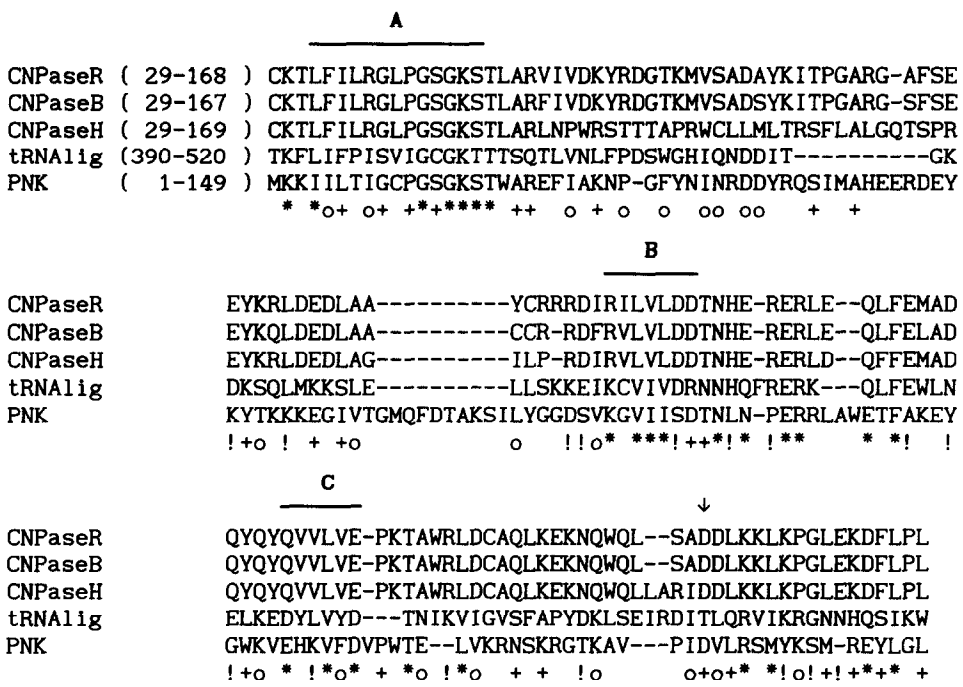

Fig. 1. Alignment of putative NTP-binding domains of CNPases, PNK and tRNA ligase. The alignment was generated by program OPTAL and subsequently modified by program MULTALIGN to maximize the number of conserved amino acid residues. Identical or similar residues in all aligned sequences (*), in the CNPases and PNK (!), in the CNPases and tRNA ligases (+), and in PNK and tRNA ligase (○) are highlighted. The grouping of similar residues was as follows: S,T; A,G; K,R; D,E,N,Q; I,L,V,M,F,Y,W. The numbers of the first and the last residues of the aligned segments in the protein sequences are indicated in parentheses. 'A' and 'B' motifs of the NTP-binding pattern and the new 'C' motif are designated. The arrowhead shows the N-terminal boundary of the CNPase fragment retaining the phosphohydrolase activity. CNPase R, B, H, rat, bovine, and human myelin CNPases, respectively.

tRNA ligase around the 'B' motif (15 identical or similar residues of 18, with a one residue insertion). The obtained alignment allowed definite identification of the 'B' motif in PNK instead of the alternative tentative identification suggested previously [16]. This motif in PNK is unusual in that the Asp residue adjacent to the hydrophobic stretch and conserved in a number of NTP-binding proteins [16] is substituted by Ser.

Despite modest sequence conservation outside the segments around the two motifs, a third conserved segment ('C') was highlighted in the distal part of the putative NTP-binding domains (Fig. 1). This motif was analogous to the 'A' and 'B' motifs, and to additional conserved motifs delineated in other NTP-binding proteins [16,17] in that it consisted of a potential hydrophobic β-strand terminated by a conserved residue (Asp/Glu, in this particular case). Additionally, a number of residues were identical or similar in the CNPases and PNK, or in the CNPases and tRNA ligase. Screening of the Swissprot database showed that the patterns of amino acid residues conserved in the segments of the CNPases, tRNA ligase and PNK surrounding 'A' and 'B' motifs (Fig. 1) are not encountered in any other protein sequence. These observations suggested a genuine functional and evolutionary kinship between the NTP-binding domains of the three enzymes. Among the other NTP-binding proteins, the closest resemblance of the patterns conserved in the new family was found in phage T4 dNMP kinase (data not shown). Perhaps this might be indicative of a distant evolutionary relationship between different types of kinases.

tRNA ligase has three distinct enzymatic activities (see above), two of which are present also in PNK, and presumably in the CNPases. Most likely, the NTP-binding pattern-containing domain encompasses the kinase activity. It was of interest to find out if the second common activity, the phosphodiesterase one, is also reflected in sequence similarity. It has been shown that the nucleotide phosphohydrolase activity resides in the C-terminal domain of the myelin CNPase polypeptide ([18]; Fig. 1). Comparison of the sequences of the C-terminal domains of PNK, tRNA ligase and the CNPases yielded an alignment significant at about 4 SD level (not shown), but in the absence of any clues as to the role of any specific segments or residues, the relevance of this marginal similarity was not obvious. A short region of similarity has been found between PNK and yeast inorganic pyrophosphatase [5]. However, upon detailed analysis, we were unable to confirm the statistical significance of this observation. Neither could we demonstrate a meaningful relationship between the sequences of the (putative) tRNA pro-

```
tRNAlig    (24-160)  LSGRGRAYRRVCDLSHSNKKVISWKFNEWDYGKNTITLPCNARGLFISDD
                     .:. :: : :. :           . .:         : :  ::.    :
RNAlig T4  (25-149)  VSASGRTY-RI--FSYN----YA-SYSDWLLPD---ALEC--RGIMFEMD

                                                          *
tRNAlig              TTNPV-IVARGYDKFFNVGEVNFTKWNWIEENCTGPYDVTIKANGCIIFI
                        :: :  :  .:::::. :  ::  : :. :    : .: : .: ..
RNAlig T4            GEKPVRIASRPMEKFFNLNENPFTM-N-IDLN-DVDYILT-KEDGSLV-S

tRNAlig              SGLEDGTLVVCSKHS-TGPRADVDRN-HAEAGEKQL---LRQL
                     . :.   ..  :: :     : .     .     :   :..:
RNAlig T4            TYLDGDEILFKSKGSIKSEQALMANGILMNINHHRLRDRLKEL
```

Fig. 2. Alignment of (putative) ligase domains of phage T4 RNA ligase and yeast tRNA ligase. (Colons) Identical residues; (dots) similar residues. Asterisk denotes the AMP-coupling Lys residue of the phage ligase.

cessing enzymes discussed here, and those of various phosphatases extracted from the Swissprot database.

Finally, we compared the sequences of yeast tRNA ligase and phage T4 RNA ligase. As could be expected, only the N-terminal domain of tRNA ligase displayed similarity to the phage enzyme. In particular, convincing similarity (5.4 SD, 30.1% identity) was observed between the very N-terminal regions of about 130 amino acid residues (Fig. 2). Indeed, it has been shown that it is the N-terminal domain of the phage RNA ligase that encompasses the activity [19]. Moreover, it has been demonstrated that Lys-99 of the T4 ligase is involved in the formation of the covalent enzyme-AMP intermediate during the RNA ligation [20]. This residue is conserved in tRNA ligase as well as the preceding run of hydrophobic residues (Fig. 2). On the other hand, another residue found to be crucial for the T4 RNA ligase activity, Asp-101 [21], is substituted by Asn in tRNA ligase, suggesting that differences in the catalytic mechanisms may exist. Previously, a degree of similarity has been reported between the regions around the (putative) catalytic Lys residues of T4 RNA ligase and DNA ligases from T4, T7 and yeast [22]. The relationship between the two RNA ligases noted here is, however, obviously closer (unpublished observations).

3.2. *Domainal organization of the enzymes of the new family*

The present observations allowed tentative delineation of the domains encompassing distinct enzymatic activities in yeast tRNA ligase, T4 PNK, and in the mammalian CNPases (Fig. 3). As described, the location of the kinase and ligase domains was confirmed by convincing sequence alignments. Though this was not the case for the phosphodiesterase domain, it has been identified experimentally in the CNPases, and in fact the other assignments left no room for this domain except the assignments shown in Fig. 3.

3.3. *Implications for the myelin CNPase functions*

The relationship between a myelin structural protein and components of tRNA splicing machinery is unexpected and intriguing. We believe that, most likely, CNPase is a bona fide mammalian polynucleotide kinase, related to the T4 enzyme not only structurally, but also functionally. This understanding relies on the two-domainal organization of this protein, the C-terminal domain encompassing the experimentally determined CNPase activity, and the N-terminal one bearing sequence similarity to the putative NTP-binding domains of PNK and tRNA ligase. Conservation of all the key residues of the NTP-binding pattern in the latter domain makes it an unlikely possibility that CNPase is an inactive homolog of polynucleotide kinase. The physiological function of this putative enzyme is not immediately obvious as the only well-characterized mammalian system of tRNA splicing, that from HeLa cell nuclei, apparently did not require PNK activity [23]. On the other hand, in the system from L cells internal inclusion of γ-phosphate from ATP into tRNA upon splicing has been observed [24]. Also, multiple PNK activities have been definitely identified in mammalian cells [23,25]. The finding of the putative mammalian PNK having also the CNPase activity suggests existence of multiple pathways of tRNA splicing (processing) in mammals. Compatibly with this view, CNPase is expressed in various mammalian tissues outside the nervous system ([4], and references therein). It seem likely that the putative polynucleotide kinase is simply recruited to fullfil the structural func-

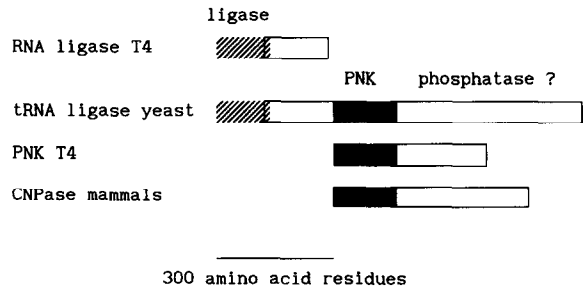

Fig. 3. A tentative scheme of the domainal organization of the CNPases and the tRNA processing enzymes. Regions of sequence similarity are highlighted by identical hatching.

tion in myelin, the enzymatic activity as such not being important for this function. This invokes the provocative analogy with crystallins, the major structural proteins of vertebrate eye lens, that are either bona fide enzymes, or are closely related to enzymes; these include alcohol dehydrogenase, lactate dehydrogenase, argininosuccinate liase, enolase, and some others (reviewed in [26]). It is of further interest that T4 RNA ligase also performs a second, structural function, as a component of the phage basal plate involved in tail fiber attachment [9]. Perhaps recruitment of enzymes for structural roles may be, after all, a more or less general principle employed in many different systems, rather than an exception. A more extravagant, and in our view less plausible possibility would be that the postulated PNK activity of CNPase is somehow involved in the myelin sheath formation.

Acknowledgements: The authors thank Dr F. Corpet for sending the program MULTALIN.

REFERENCES

[1] Phizicky, E.M., Schwartz, R.C. and Abelson, J. (1986) J. Biol. Chem. 261, 2978–2986.
[2] Amitsur, M., Levitz, R. and Kaufmann, G. (1987) EMBO J. 6, 2499–2503.
[3] Bradbury, J.M. and Thompson, R.J. (1984) Biochem. J. 33, 1143–1150.
[4] Bernier, L., Alvarez, F., Norgard, E.M., Raible, D.W., Mentaberry, A., Shembri, J.G., Sabatini, D.A. and Colman, D.R. (1987) J. Neurochem. 7, 2703–2710.
[5] Vogel, U.S. and Thompson, R.J. (1987) FEBS Lett. 218, 261–265.
[6] Kurihara, T., Takahashi, Y., Nishiyama, A. and Kumanishi, T. (1988) Biochem. Biophys. Res. Commun. 152, 837–842.
[7] Midgley, C.A. and Murray, N.E. (1985) EMBO J. 4, 2695–2703.
[8] Westaway, S.K., Phizicky, E.M. and Abelson, J. (1987) J. Biol. Chem. 263, 3171–3176.
[9] Rand, K.N. and Gait, M.J. (1984) EMBO J. 3, 397–402.
[10] Gorbalenya, A.E., Leontovich, A.M. and Brodsky, L.I. (1988) in: Theoretical Research and Data Banks on Molecular Biology and Genetics (Ratner, V.A. ed.) p. 64, Novosibirsk.
[11] Brodsky, L.I., Gorbalenya, A.E. and Leontovich, A.M. (1990) Biopolymery i Kletka (in press).
[12] Gorbalenya, A.E., Blinov, V.M., Donchenko, A.P. and Koonin, E.V. (1989) J. Mol. Evol. 28, 256–268.
[13] Koonin, E.V., Chumakov, K.M. and Gorbalenya, A.E. (1990) Biopolymery i Kletka (in press).
[14] Corpet, F. (1988) Nucleic Acids Res. 16, 10881–10890.
[15] Walker, J.E., Gay, N.J., Saraste, M. and Runswick, M.J. (1982) EMBO J. 1, 945–951.
[16] Gorbalenya, A.E. and Koonin, E.V. (1989) Nucleic Acids Res. 17, 8413–8440.
[17] Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P. and Blinov, V.M. (1989) Nucleic Acids Res. 17, 4813–4830.
[18] Kurihara, T., Fowler, A.V. and Takahashi, Y. (1987) J. Biol. Chem. 262, 3256–3261; 16754.
[19] Rand, K.N., Singh, M., Thogersen, H.C. and Gait, M.J. (1985) Proc. Int. Symp. Biomol. Struct. Interact. J. Biosci. 8 (Suppl)., 89–106.
[20] Thogersen, H.C., Morris, H.R., Rand, K.N. and Gait, M.J. (1985) Eur. J. Biochem. 147, 325–329.
[21] Heaphy, S., Singh, M. and Gait, M.J. (1987) Biochemistry 26, 1688–1696.
[22] Barker, D.G., White, J.H.M. and Johnston, L.H. (1985) Nucleic Acids Res. 13, 8323–8337.
[23] Filipowicz, W. and Shatkin, A.J. (1983) Cell 32, 547–557.
[24] Winicov, I. and Button, J.D. (1982) Mol. Cell. Biol. 2, 241–249.
[25] Shuman, S. and Hurwitz, J. (1979) J. Biol. Chem. 254, 10396–10404.
[26] Wistow, G.J. and Piatigorsky, J. (1988) Annu. Rev. Biochem. 57, 479–504.