Volume 334, number 2, 149–152 FEBS 13236 November 1993

Hypothesis

A hybrid protein kinase-RNase in an interferon-induced pathway?

Peer Bork*, Chris Sander*

EMBL, Meyerhofstr. 1, D-69012 Heidelberg, Germany

Received 31 August 1993; revised version received 24 September 1993

The sequence of RNase L has been re-examined by computer analysis. We propose a molecular architecture of RNase L, with an unusual combination, in one protein chain, of 9 ankyrin-like repeats, a functional active protein kinase and a C-terminal catalytic RNase similar to the yeast protein, IRE1. The protein kinase may be involved in a new signal transduction pathway which remains to be discovered.

RNase; Protein kinase; Homology; Ankyrin-like repeat; Domain

1. INTRODUCTION

Sequence analysis using computers is a powerful tool for elucidating the molecular function of proteins, but it can also lead to pitfalls when interpreting similarities detected in database searches. An example of the latter problem is provided by the recently sequenced 2-5A-dependent RNase [1], an interferon-induced enzyme that is activated by 5′-phosphorylated, 2′-5′-linked oligoadenylates (2-5A). The protein is also called RNase L and may be involved in the inhibition of viral replication and/or in tumor suppression [1,2]. Zhou et al. [1] report several sequence motifs within the approximately 740 residue long sequences of human and murine 2-5A-dependent RNases. These are two nucleotide triphosphate binding sites (P-loops), a zinc finger, motif VI and VII of protein kinases, and a region with similarity to *Escherichia coli* RNase E. We have re-analyzed the sequences of both these 2-5A-dependent RNases by a variety of methods [3], combining the results of standard database searches with information from multiple sequence alignments and known 3D structures. Surprisingly, we have found a molecular architecture of the enzyme that is almost completely different from that proposed in [1]. The 2-5A-dependent RNases each consist of (i) 9 ankyrin-like repeats (ANK), (ii) a complete protein kinase-like domain and (iii) a C-terminal, 130 residue-long region which we presume to contain the RNase activity (Fig. 1).

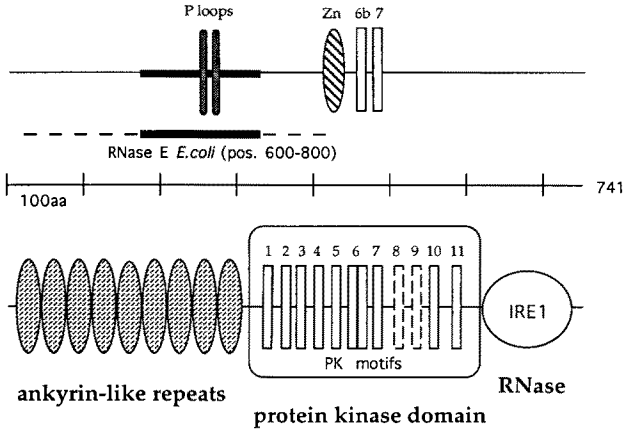

Fig. 1. Molecular architecture of 2-5A-dependent RNase. (Top) Original assignments [1]. (Bottom) Revised assignments, following extensive sequence analysis using a number of methods [3]. Notation for original assignments: P-loops (phosphate binding loops), Zn (zinc finger), 6b,7 (protein kinase motifs [13]); for revised assignments: IRE1 (region similar to yeast IRE1 protein). The highly significant similarity of 2-5A-dependent RNase to ANK-repeats (e.g. BLASTP P values $<10^{-20}$) contradicts the presence of P-loops which have been proposed on the basis of the very frequently occurring and therefore not significant tripeptide GKT [1]. The reported similarity to RNase E from *E. coli*, which also falls within the ANK repeat region, is contradicted as well by the fact that the region of RNase E proposed to match 2-5A-dependent RNase is (i) biased towards negatively charged residues (composition-biased regions can lead to spuriously high search scores) and (ii) likely to have a coiled coil region, as predicted by the method of Lupas et al. [17] and is therefore not homologous to the ANK region of 2-5A-dependent RNase. In positions 395–334 a zinc finger was proposed by Zhou et al. [1]. A detailed analysis of this region, however, did not reveal any significant similarity to known zinc fingers. Furthermore, human and mouse 2-5A-dependent RNase differ in the positions of the cysteines and the cysteine pattern in the human enzyme is atypical of that observed in classical zinc fingers.

*Corresponding authors. Fax: (49) (6221) 387–306. Email: bork@embl-heidelberg.de.

```
ANKS:  G TPLHhAht        thht LLt  GAt t
cons:  G T Lh Aht        thht LLtt GAt t
  24   EDNHLLIKAVQNEDV...DLVQQLLEG.GANVNFQEEE
  58   GGWTPLHNAVQMSRE...DIVELLLRH.GADPVLRKK
  91   NGATLFILAAIAGSV...KLLKLFLSK.GADVNECDF
 124   YGFTAFMEAAVYGKV...KALKFLYKR.GANVNLRRKTKEDQERLRK
 167   GGATALMDAAEKGHV...EVLKILLDEMGADVNACDNM
 202   GRNALIHALLSSDDSDVEAITHLLLDH.GADVNVRGE
 238   RGKTPLILAVEKKHL...GLVQRLLEQEHIEINDTDS
 272   DGKTALLLAVELKLK...KIAELLCKR.GASTD....
 301   CG.DLVMTARRNYDH...SLVKVLLSH.GAKEDFHPPAEDWKPQS..
       [          ANK repeat                 ]
```

Fig. 2. Alignment of the 9 internal repeats in the N-terminal region of human 2-5A-dependent RNase. In the top two lines the consensus derived from the alignment (cons) is compared with the consensus of about 650 ANK repeats (ANKS) [5]. Numbers on the left are residue positions of the beginning of the repeat in the 2-5A-dependent RNase sequence.

2. MOLECULAR ARCHITECTURE OF 2-5A-DEPENDENT RNase

The N-terminal region of 2-5A-dependent RNase has more than 35% amino acid identity over 320 residues to consecutive ANK repeats in human erythrocyte ankyrin [4]. More detailed analysis shows that the first 330 residues in 2-5A RNase indeed contain a 9-fold approximate internal repeat of average length 33 residues (Figs. 1 and 2). Moreover, the consensus sequence of the repeat is nearly identical to a consensus [5] derived from almost 650 ANK repeats (Fig. 2). ANK repeats occur as consecutive copies in functionally diverse proteins

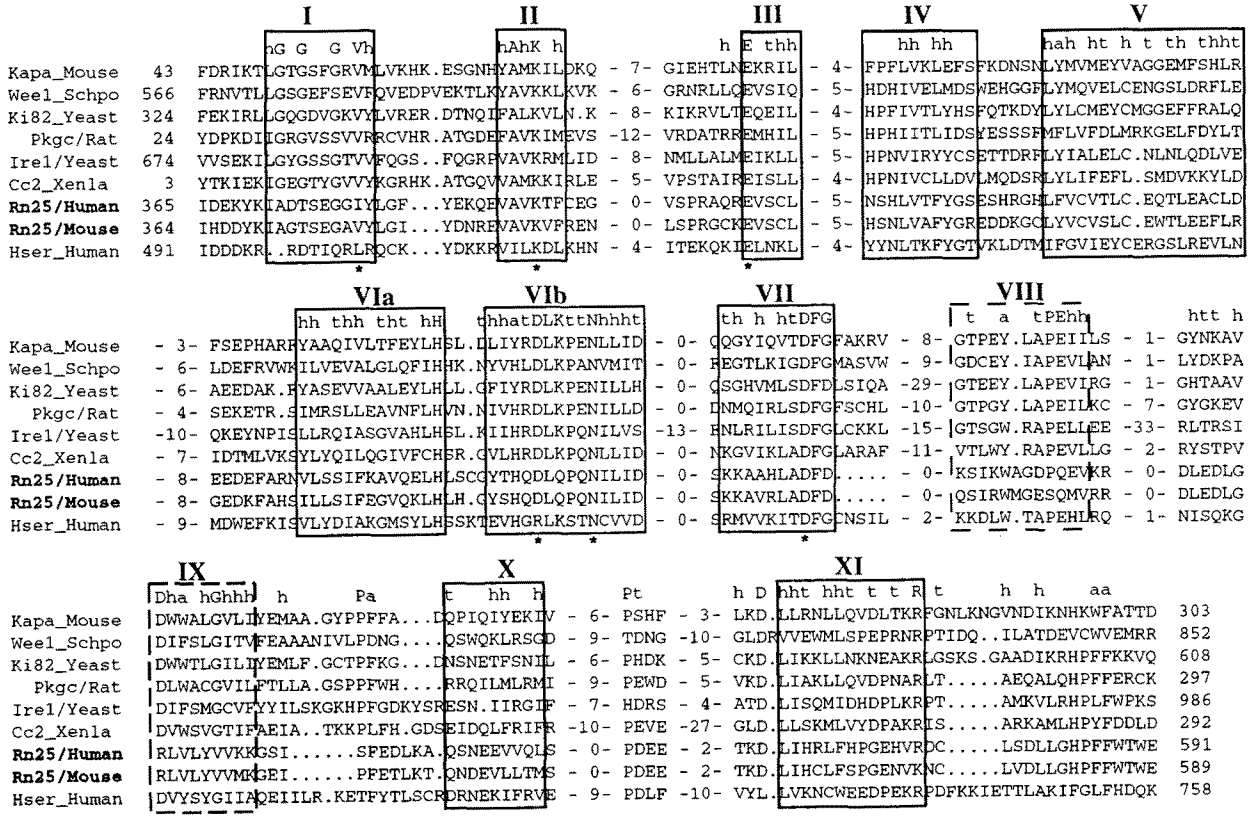

Fig. 3. Alignment of selected protein kinases with the predicted kinase domain in 2-5A-dependent RNase. The protein kinases shown are representatives of different subfamilies and are as similar to 2-5A-dependent protein kinase as they are to each other. Only the regions around the 11 conserved protein kinase boxes (Roman numerals) are shown. (Top line) Upper case letters denote conserved amino acids, lower case letters stand for conserved amino acid properties (a = aromatic, h = hydrophobic, t = turn-like or polar, +/− =th?positively/negatively charged). (Left) Protein names as in the SWISSPROT database [18], if available (e.g. Hser = guanylyl cyclase GC-C, Kapa = cAMP dependent protein kinase), followed by residue numbers at the beginning of the domain. (Between boxes) Number of intervening residues. Asterisks mark all strictly conserved residues for which the side chains are known to be directly involved in ligand binding/catalysis [14]. Agreement with protein kinase motifs is as follows. Motif I deviates slightly from the consensus (top line), but more strongly modified variants have already been observed in other protein kinases. Motif I appears to be completely absent in the kinase domain of the membrane-bound guanylyl cyclases GC-C [15,19], one of the kinase subfamilies most similar in sequence to 2-5A-dependent RNase. In motif VII, a conserved glycine (G) is replaced by an aspartate (D), as observed in several putative plant protein kinases and in KIN82. Since the loop following in the sequence [15] appears to be much shorter than in most of the other protein kinases, the D sidechain can probably be accommodated and is unlikely to have a drastic steric effect on substrate binding. Although the general features of motif VIII remain conserved, the otherwise invariant glutamate (E) is replaced by the physicochemically similar glutamine (Q). The absence of a conserved aspartate (D) in motif IX is no cause for concern. This residue neither participates directly in catalysis, nor does it appear to be crucial for stabilizing the fold [15]. Thus, the agreement with most of the boxes in the appropriate spacing scattered throughout this domain in the 2-5A-dependent RNase shows that this domain may have protein kinase function. In any event, it is very likely to have the same 3D fold as the catalytic subunit of other kinases.

```
          Wt tt+ht L tVt-t -It Rt  S hL  h- tt    tS D WT Kh-t hMtth-+ YtK     attth DLL+hhRN   Hh - ttt htL  t Pt    YFtKtFP-LhI VY
Ire1/Yeast WPKSKKLEFLLKVSDRLEIENRDPPSALLMKFDAGSDFVIPSGD.WTVKFDKTFMDNLER.YRK....YHSSKLMDLLRALRNKYHHFMDLPEDIAELMGPVPDGFYDYFTKRFPNLLIGVYM
Rn25/Human WTWESRYRTLRNVGNESDIKTRKSESEILRLLQPGPSEHSKSFDKWTTKINECVMKKMNKFYEKR.GNFYQNTVGDLLKFIRNLGEHIDEEKHKKMKLKIGDPSL...YFQKTFPDLVIYVYT
Rn25/Mouse WTWENRYRTLRNVGNESDIKVRKCKSDLLRLLQHQTLEPPRSFDQWTSKIDKNVMDEMNHFYEKRKKNPYQDTVGDLLKFIRNIGEHINEEKKRG>>>
```

Fig. 4. Alignment of the C-terminal, probably catalytic, domain of 2-5A-dependent RNases with the C-terminus of yeast IRE1 protein [10]. Bold residues are conserved between IRE1 and at least one of the two 2-5A RNases. Mouse 2-5A RNase has not entirely been sequenced yet. Invariant charged and polar residues in conserved regions are likely to participate in catalysis.

and are thought to mediate protein–protein interactions [4,6]. An involvement of ANK in DNA binding was shown for the GA-binding protein complex [7] and for transcription factor complex SWI4/SWI6 [8]. The ability of ANK repeats in 2-5A-dependent RNase to bind oligoadenylates would explain the abrupt loss of 2-5A binding affinity when truncating the 7th and 8th repeat, i.e. position 265–294 [1].

The putative protein kinase domain of 2-5A-dependent RNase is located immediately following the ANK repeats (Figs. 1 and 3). The closest protein kinase relative appears to be yeast KIN82 [9], with 30% amino acid sequence identity over the entire kinase domain. Other kinases, such as yeast IRE1 [10], are also significantly similar, with high FASTA opt scores >150 [11] and/or BLASTP P values $<10^{-6}$ [12]. These similarities are verified by a multiple sequence alignment which indeed reveals the presence of all 11 boxes conserved in the protein kinase family [13], with some modifications (Fig. 3). Further evidence (data not shown) comes from mapping conserved residues onto the known 3D structure of mouse cAMP-dependent kinase [14]. It confirms that the changes are neither in functionally nor in structurally essential positions. Not only are all essential hydrophobic 3D contacts conserved (in particular near the active site), but also all residues known to participate in ATP- and peptide binding in protein kinases (Fig. 3).

So, if most of the protein sequence consists of ANK repeats and a protein kinase domain, then the RNase activity is most probably located in the C-terminal 130 residues. Several facts support this hypothesis. (i) The C-terminal domain of 2-5A-dependent RNase is more conserved between mouse and human (73% identical residues) than the two proteins are on average (64%), indicating strong selective pressure on this domain, typical of catalytic function. (ii) Another protein family, the membrane-bound guanylyl cyclases [15], with similar modular construction (protein kinase plus C-terminal domain), has its catalytic domain also at the C-terminus. (iii) Interestingly, the C-terminal domain of 2-5A-dependent RNase has 29% sequence identity (FASTA opt. score of 155 [11]) to the C-terminal domain of yeast IRE1 protein (Fig. 4). (Yeast IRE1 protein is involved in inositol phototrophy and has been identified by genetic complementation of *myo*-inositol auxotrophic yeast mutants [10].) (iv) The IRE1 protein also contains a protein kinase domain (Fig. 1, bottom panel) and the two putative catalytic domains are in the same relative location in the two proteins. Taken together, these facts indicate that the C-terminal domain of 2-5A-dependent RNase is a catalytic domain, which very plausibly has RNase function.

As neither 2-5A-dependent RNase nor IRE1 (for which an RNase function can be predicted by analogy) have obvious sequence similarity to other RNase families, they might form a new structural class of RNases. Although the functionality of the kinase domain has not yet been proved, its presence could imply a regulatory function for 2-5A-dependent RNase in the interferon-dependent pathway.

3. RECENT EXPERIMENTAL EVIDENCE

As this manuscript was about to be submitted, a report by Hassel et al. [16] appeared which reports the presence of 9 ANK repeats and provides experimental evidence for the location of the RNase function within the C-terminus (a clone lacking the last 89 residues has no RNase activity) [16]. This confirms parts of our conclusions and focusses attention on experimental verification of the predicted protein kinase function. The C-terminal similarity to IRE1 is useful for identification of conserved polar residues that might be involved in catalysis (Fig. 4).

REFERENCES

[1] Zhou, A., Hassel, B.A. and Silverman, R.H. (1993) Cell 72, 753–765.
[2] Lengley, P. (1993) Proc. Natl. Acad. Sci. USA 90, 5893–5895.
[3] Bork, P., Ouzounis, C., Sander, C., Scharf, M., Schneider, R. and Sonnhammer, E. (1992) Protein Sci. 1, 1677–1690.
[4] Lux, S.E., John, K.M. and Bennett, V. (1990) Nature 344, 36–42.
[5] Bork, P. (1993) Proteins (in press).
[6] Bennett, V. (1992) J. Biol. Chem. 267, 8703–8706.
[7] Thompson, C.C., Brown, T.A. and Knight, S.L. (1991) Science 253, 762–768.
[8] Sidorova, J. and Breeden L. (1993) Mol. Cell. Biol. 13, 1069–1077.
[9] Wilson, C., Bergantino, E., Lanfranchi, G., Valle, G., Carigani, G. and Frontali, L. (1992) Yeast 8, 71–77.
[10] Nikawa, J.I. and Yamashita, S. (1992) Mol. Microbiol. 6, 1441–1446.
[11] Pearson, W. R, and Lipman D. (1988) Proc. Natl. Acad. Sci. USA 85, 2444–2448.
[12] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990 J. Mol. Biol. 215, 403–410.
[13] Hanks, S.K. and Quinn, A.M. (1991) Methods Enzymol. 200, 38–62.
[14] Knighton, D.R., Zheng, J.H., Ten Euck, L.F., Ashford, V.A.,

Yuong, N.H., Taylor, S.S. and Sowadski, J.M. (1991) Science 253, 407–414.
[15] Schulz, S., Green, C.K., Yuen, P.S.T. and Garbers, D.L. (1990) Cell 63, 941–948.
[16] Hassel, B.A., Zhou, A., Sotomayor, C., Avudaiappan, M. and Silverman, R.H. (1993) EMBO J. 13, 3297–3304.
[17] Lupas A., Van Dyke, M. and Stock, J. (1991) Science 252, 1162–1164.
[18] Bairoch, A. and Boeckmann, B. (1992) Nucleic Acids Res. 20 Suppl., 2019–2022.
[19] de Sauvage, F.J., Camerato, T.R. and Goeddel, D.V. (1991) J. Biol. Chem. 266, 17912–17918.