J. Mol. Evol. 15, 355–358 (1980)

Journal of
Molecular Evolution
© by Springer-Verlag · 1980

Homology between Prokaryotic and Eukaryotic Ribonucleases

R.W. Hartley

Laboratory of Nutrition and Endocrinology, National Institute of Arthritis, Metabolism and Digestive Diseases, Bethesda, Maryland 20205, USA

Summary. There is homology between the amino acid sequences of the extracellular ribonucleases T1 and St, from the eukaryote *Aspergillus oryzae* and the prokaryote *Streptomyces erythreus*, respectively. Together with other extracellular ribonucleases homologous to each, these enzymes make up a family of interest to evolutionary biology and useful in studies of protein structure and function.

Key words: Homology – Amino acid sequence – Ribonuclease – Protein evolution

While homology between prokaryotic and eukaryotic proteins is not unprecedented, the number of known cases is still small and each case may provide a different clue to the difficult problem of how the eukaryotes originated. It is reported here that there is substantial sequence homology between several of the extracellular ribonucleases of bacteria and of fungi.

An alignment of several homologous segments of ribonuclease St (Yoshida et al., 1976) of the bacterium *Streptomyces erythreus* and ribonuclease T1 (Takahashi, 1965) of the mold *Aspergillus oryzae* is shown in Table 1. As reported by others (Aphanasenko et al., 1979), there is also homology between ribonuclease St and barnase (Hartley and Barker, 1972), the extracellular ribonuclease of *Bacillus amyloliquefaciens* and the close homolog of barnase from *Bacillus intermedius.* Also, ribonuclease T1 is homologous to two different extracellular ribonucleases, U1 (Hashimoto and Takahashi, 1974), and U2 (Sato and Uchida, 1975), from another mold, *Ustilago sphaerogena.* U1 has only been partly sequenced but it is clearly homologous to, and about equidistant from, both U2 and T1.

Figure 1 shows an overall alignment of the five available complete sequences. The alignment was obtained by a combination of computer and manual analysis and is not

0022–2844/80/0015/0355/$ 01.00

Table 1. Alignment of segments of amino acid sequences of ribonuclease T1 of *Aspergillus oryzae* and ribonuclease St of *Streptomyces erythreus.* Position in whole protein is indicated by the number above the first residue of each segment. Introduction of a one residue gap is represented by (-*-). Matching residues are shown in bold face

Enzyme	Sequences		
	23	44	56
T1	-**Gly-Tyr**-Gln-**Leu**-His-**Glu**-Asp-**Gly**-	-**Asn**-Tyr-**Glu-Gly**-	-**Tyr-Tyr**-*-**Glu**-
	26	47	59
St	-**Gly-Tyr**-Glu-**Leu**-Ile-**Glu**-Lys-**Gly**-	-**Asn**-Arg-**Glu-Gly**-	-**Tyr-Tyr**-His-**Glu**
	69		92
T1	-**Ser-Gly**-Pro-Gly-Ser-**Gly-Ala**-Asp-**Arg**-Val-**Val**-		-**His**-
	69		92
St	-**Ser-Gly**-Asp-Asp-Arg-**Gly-Ala**-Arg-**Arg**-Phe-**Val**-		-**His**-

represented as being an absolutely optimum match. The alignments were analysed, for each pair, by a procedure similar to that of Barker and Dayhoff (1972 and Dayhoff, 1976) and using their mutation data matrix. This procedure takes into account the probability, based on known related sequences, of particular amino acid substitutions. A raw score, S, is obtained for each aligned pair, with bias and gap penalties (each equal to two) as described by Barker and Dayhoff. Similar scores, plus a mean, S_r, and standard deviation, SD_r, are computed 50 times for the same pair but with one sequence randomized for each score. The alignment score, A, is then the number of standard deviations by which the raw score for the real sequences exceeds the mean of the random scores:

$$A = (S - S_r) / SD_r$$

```
                         10                  20                  30                  40
B.a.  A Q V I N T F D G V A D Y L Q T Y H K L P N D Y I T K S E A Q A L G W V A S K G N L
B.i.  A - V I N T F D G V A D Y L I R Y K R L P N D Y I T K S Q A S A L G W V A S K G D L
S.e.  - - E A P C G D T S G F E Q V R L A D L P P - - - - - - - E A T D T G Y - - - - - E L
A.o.  - - - - A C D Y T C G S N C Y S S S D V S T - - - - - - - - A Q A A G Y - - - - - Q L
U.s.  - - - - - C N I P E S T N C G G N V Y S N D - - - - - - - D I N T A I Q - - - - G A L
                    50                  60                  70                  80
B.a.  D V A - P - G K S I G G D I F S N R E G K L P G K S G R - - T W R E A D I N Y T - -
B.i.  E V A - P - G K S I G G D V F S N R E G R L P S A G S R - - T W R E A D I N Y V - -
S.e.  E K G G P Y P Y P E D G T V F E N R E G I L P D C A E G - - Y Y H E Y T V K T P - -
A.o.  E D G E T V G S N S Y P H K X N N Y E G F D F S V S S P - - Y Y - E W P I L S S G D
U.s.  D V A R P D G D N - Y P H Q Y Y D E A S D Q I T L C C G P G S W S E F P L V Y N G P
                90                  100                 110                 120
B.a.  - S G F - R - - - - N S D R I L Y S S D W L - I Y K T T D H Y Q T - F T K I R - - -
B.i.  - S G F - R - - - - N A D R L V Y S S D W L - I Y K T T D H Y A T - F T R I R - - -
S.e.  - S G D D R - - - - G A R R F V V G D G G E - Y F Y T E D H Y E S - F R L T I V N -
A.o.  Y S G P G S - - - - G A D R V V F N E N N Q - L A G V I T H T G A - S G N N F V Q C
U.s.  Y S S R D N Y V S P G P D R V I Y Q T N T G E F C A T V T H T G A A S Y D G F T Q C
```

Fig. 1. Sequence alignment of the extracellular ribonucleases of: B.a., *Bacillus amyloliquefaciens;* B.i., *Bacillus intermedius;* S.e., *Streptomyces erythreus;* A.o., *Aspergillus oryzae;* U.s., *Ustilago sphaerogena*

This is somewhat simpler than the original, but the bias introduced by simply eliminating all gaps in the random sequences is in the direction of underestimating the alignment scores. The probability that a score better than A could occur by chance is then based on an assumed normal distribution of the random scores.

The scores obtained are shown in Table 2, along with the probabilities that such scores, or better, would occur by chance. Also shown are the identity scores, representing the percentages of matching amino acids in each pair alignment. While the alignment scores of about three for the more distant pairs would not be convincing alone, the high scores for all adjacent pairs makes the argument for overall homology overwhelming.

The argument is even stronger if we look at some particular residues. At only five positions do the same amino acids appear in all five sequences. It is important to note, therefore, that chemical modification studies of ribonuclease T1 (Takahashi et al., 1967, Takahashi, 1970, 1976) have identified three of these, Glu-77, Arg-100 and His-116, as participating in the catalytic site. Arg-100 is the only arginine in ribonuclease T1 and the *B. intermedius* enzyme has only the single histidine. The other two conserved residues are Leu-42 and Ser-88.

Table 2. Pairwise alignment scores, identity scores, and the probabilities that alignment scores as good or better would occur by chance. Prokaryotes; 1) *Bacillus amyloliquefaciens,* 2) *Bacillus intermedius,* 3) *Streptomyces erythreus,* Eukaryotes; 4) *Aspergillus oryzae,* 5) *Ustilago sphaerogena*

				Probability, P		
		1.	2.	3.	4.	5.
	1.	–	10^{-24}	4×10^{-11}	10^{-3}	10^{-2}
	2.	19.2 (80.9)	–	4×10^{-11}	10^{-3}	2×10^{-3}
Alignment score, A	3.	6.5 (22.6)	6.5 (26.3)	–	10^{-11}	10^{-3}
(Identity score, I)	4.	3.1 (14.0)	2.9 (16.9)	7.0 (22.2)	–	10^{-16}
	5.	2.7 (12.4)	3.4 (11.7)	2.9 (9.4)	8.6 (26.1)	–

The data I present are relevant to 1. considerations of the mechanism of hydrolysis by these ribonucleases and 2. extension of currently available data indicating homology between similar proteins of prokaryotes and eukaryotes. Yoshida et al. (1976) suggested a connection between ribonucleases T1 and St on the basis of a proposed active site topography but they apparently missed the sequence homology reported here. In view of this homology, only the Arg-100 (present numbering) and His-116 correspondences in their scheme appear to be correct. If we assume a universal active site arrangement, any fundamental role for either lysine or tryptophan can be ruled out, as each is completely absent from one or more of the enzymes. While all are reported to hydrolyse RNA in the same manner, yielding 2',3' cyclic- and 3'-nucleotides, there are differences

in specificity (Egami and Nakamura, 1969). The *Bacillus* enzymes are relatively nonspecific, while St, T1 and U1 are guanine specific and U2 is purine specific. Note also that the *Bacillus* enzymes are basic, while the others, in spite of the necessity of their interacting with RNA, are acidic.

As in other cases of this sort, the question remains as to whether homology is due to common descent of the organism or to some form of horizontal gene transfer. Answers to such questions will require understanding, for a large number of gene products, not only of sequence homologies, but also of structural and functional restraints on their evolution. The stringent requirements for folded stability in proteins as small as these ribonucleases, for example, may have helped to maintain evidence of homology in spite of great phylogenetic distance.

Comparison of the three-dimensional structures of these enzymes with each other, and even with pancreatic ribonuclease, will be awaited with interest. Efforts to solve the structure of barnase have been under way for some time and large crystals of ribonuclease T1 have been reported (Minato et al., 1966). Preliminary X-ray crystallographic analyses have also been reported for two guanine-specific ribonucleases from *Aspergillus clavatus* and *Penicillium chrysogenum* (Bezborodova et al., 1977). Further study of this family of small enzymes may contribute both to our knowledge of early evolution and to the protein chemists' major problem of the mechanism of sequence-directed folding.

References

Aphanasenko, G.A., Dudkin, S.M., Kaminir, L.B., Leshchinkaya, I.B., Severin, E.S. (1979). FEBS Lett. **97**, 77–80

Barker, W.C., Dayhoff, M.O. (1972). Atlas of protein sequence and structure **5**, 101–110. Washington, DC: National Biomedical Research Foundation

Bezborodova, S.I., Beletskaya, O.P., Tishchenko, G.N., Zhuchlistova, N.E., Nuzimova, N.W. (1977). Kristallographiya **22**, 1089–1091

Dayhoff, M.O. (1976). Atlas of protein sequence and structure **5**, Suppl. 2, 1–8 Washington, DC: National Biomedical Research Foundation

Egami, F., Nakamura, K. (1969). Microbial ribonucleases, pp. 29–46. Berlin Heidelberg New York: Springer

Hartley, R.W., Barker, E.A. (1972). Nature New Biol. **235**, 15–16

Hashimoto, J., Takahashi, K. (1974). J. Biochem. (Tokyo) **76**, 1359–1361

Minato, S., Tagawa, T., Nakanishi, K. (1966). J. Biochem. (Tokyo) **59**, 443–448

Sato, S., Uchida, T. (1975). Biochem. J. **145**, 353–360

Takahashi, K. (1965). J. Biol. Chem. **240**, 4117–4119

Takahashi, K. (1970). J. Biochem. (Tokyo) **68**, 659–664

Takahashi, K. (1976). J. Biochem. (Tokyo) **80**, 1267–1275

Takahashi, K., Stein, W.M., Moore, S. (1967). J. Biol. Chem. **242**, 4683–4690

Yoshida, N., Sasaki, A., Rashid, M.A., Otsuka, M. (1976). FEBS Lett. **64**, 122–125

Received December 10, 1979; revised March 10, 1980