Proteins, RNAs and chaperones in enzyme evolution: a folding perspective

Peter Csermely

The present roles of RNA molecules as templates and of proteins as cellular catalysts may not have always been so clearly defined during evolution. Recent work on ribozymes shows that the catalytic activity of early RNAs may have been more general than previously thought. How did evolution select proteins, not RNA, to be the major biological catalysts? Why were chaperones necessary for the evolution of modern protein enzymes?

RECENT ADVANCES IN ribozyme biochemistry suggest that RNA may have been an even more general early catalyst than previously thought, able to catalyse the formation of C–N bonds and amino acid transfer reactions[1]. However, evolution preferred proteins as cellular catalysts, while enzymatic functions of ribonucleic acids became fairly specific. The general explanation of this phenomenon is centered around the multifunctionality of proteins consisting of 20 amino acids compared to the paucity of catalytically active segments of the four nucleotides in RNA[2]. Detailed analyses of the enzyme reactions catalysed by proteins and RNAs, however, show that the real differences in the catalytic potential of the two macromolecules are not so extensive[1,2]. If we take into account those hypotheses that propose a smaller number of proteinaceous amino acids and a larger number of ribonucleotides at the early phase of molecular evolution[2–6], the above explanation of the present-day major duties of RNAs as templates and of proteins as catalysts seem to be even less satisfactory.

Differences between proteins and RNAs: folding and stability

What was the additional driving force that helped to specialize proteins to be the major source of catalytic power in the present living organisms? The above examples may indicate that composition alone is not enough, a further clue must lie in the structure of these molecules. Protein folding is a highly cooperative process regardless of whether it proceeds via an initial collapse and then restructuring, as is suggested for smaller proteins, or if the final structure is formed via folding nuclei having a less-pronounced helical/β-sheet content as proposed for larger proteins[7–9]. RNA folding mainly consists of the formation of canonical base-pairing complemented with non-canonical and 'tertiary' interactions. By contrast to proteins, RNA secondary structures are extremely stable in the absence of tertiary structure. Sidechains are on the inside of RNA double helices, while amino acid sidechains face outwards in a protein secondary structure, allowing cooperative hydrophobic interactions to form tertiary structure, which is the major governing force in protein folding. In RNA, base-stacking and divalent cations stabilize the structure instead. RNA base-pairing also proceeds by cooperative interactions, but this cooperativity is far less than that observed in proteins. Consequently, the folding autonomy of secondary structure elements is much higher in RNA molecules than in proteins. In the reverse process, most RNAs unfold in discrete, sequential steps, while protein denaturation is often a concerted process. Therefore, interactions in RNAs are few, strong, complementary and non-cooperative, while interactions in proteins are plenty, weak, non-complementary and cooperative[10–12].

P. Csermely is at the Department of Medical Chemistry, Semmelweis University, H-1444 Budapest, PO Box 260, Hungary.
Email: csermely@puskin.sote.hu

Catalytic power of RNA and proteins

How do these differences in RNA- and protein folding affect the catalytic power of the structured, native macromolecules? A detailed answer clearly requires much more experimental and theoretical work. However, some initial remarks can be made. Ribozymes often have mismatches and other structural irregularities at their catalytic core, which makes the core secondary structure less stable. Protein enzymes inherently have this destabilizing force owing to the weak intrinsic stability of their secondary structures. Differences in the complementarity and cooperativity of folding may also contribute to the slow product release of some multiple-turnover ribozymes compared to protein enzymes[10,13].

It is difficult even to estimate how close the present (natural or artificial) forms of protein enzymes and ribozymes have got to the theoretical limits of their catalytic repertoire. It is quite possible that RNAs have a larger 'evolutionary unexplored space' than proteins (Fig. 1). One step in protein evolution (a change of a single amino acid) usually destroys the catalytic power of that enzyme much less than the change of one nucleotide affects the efficiency of a ribozyme[12–14]. However, this 'evolutionary disadvantage' of RNA is also derived from differences in complementarity and cooperativity between RNA and protein folding.

Evolution of rigidity followed by evolution of flexibility

The proper arrangement of catalytic residues requires a stable structure. Small peptides can rarely acquire a stable fold. Therefore, the 'first phase' of enzyme evolution was likely to increase the size in order to achieve more rigidity. However, a structure that is too rigid does not allow correct proximity and positioning of key residues, which would result in little catalytic activity and poor selectivity. Consequently, the 'second phase' of enzyme evolution increased the size further to develop more flexible structures with the allosteric properties[9,14–16] (Fig. 1). The ribose-phosphate backbone of ribozymes is inherently more flexible than the peptide backbone of protein enzymes. For this reason, RNAs may need to be larger than proteins to achieve the same rigidity[13]. The reasonable limits for size-growth, the inherent differences between protein and RNA structure and the higher risk of individual mutations of RNAs than that of proteins may have allowed ribozymes to reach only the relatively 'rigid' and not the 'more flexible, allosteric' phase of enzyme evolution (Fig. 1).

The price of efficient catalysis: the folding problem and chaperones

Efficiency has its price. Many of the small proteins studied so far require only

Copyright © 1997, Elsevier Science Ltd. All rights reserved. 0968-0004/97/$17.00 PII: S0968-0004(97)01026-8

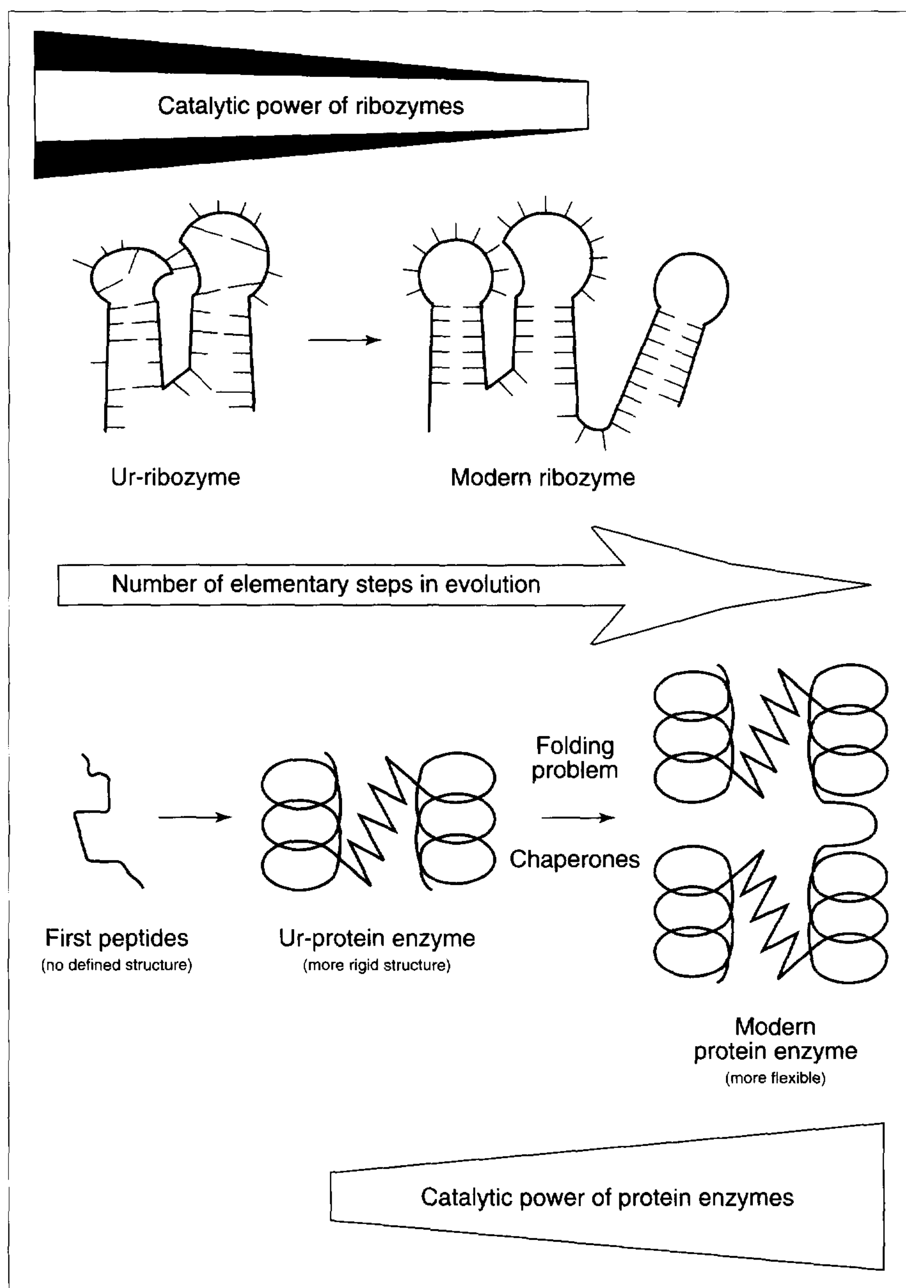

Figure 1
Cartoon depicting a hypothetical scheme for the development of modern ribozymes and protein enzymes. High cooperativity and weak, non-complementary interactions in protein folding may significantly contribute to the fact that molecular evolution selected ancient pre-proteins as major catalysts and probably increased the number of their constituent amino acids to perform this function even better. Orange segment of the catalytic power of ribozymes (the catalytic repertoire of the total ribozyme population of Earth) illustrates the purely speculative development of Ur-RNAs loosing some of their nucleotides and reducing their catalytic power, but increasing their fidelity as templates.

a few milliseconds to fold and need little, if any, help to reach a relatively good yield of their final structure[8–10]. The evolution of more efficient, larger enzymes was paralleled by the emergence of the 'protein-folding problem' as a consequence of the increasing complexity of the folding landscape[17]. This required the help of chaperones to protect the larger proteins from aggregating during their extended folding, and to unfold misfolded proteins[18].

Molecular chaperones were probably also necessary to prevent the aggregation of primitive, small proteins. We have vague estimates, however, on the protein concentration of the putative 'primordial soup' or in the cytoplasm of the first living organisms. In ancient, dilute conditions, more help would be required in unfolding than in the prevention of aggregation.

The emerging examples seem to suggest that 'a good fold is a rare fold', which is preserved when it has been found during evolution[19]. In more complex proteins 'good folds' may behave as folding nuclei, which are surrounded by more flexible structures that allow better enzyme activity and that probably developed later in evolution. Folding of these proteins is an especially error-prone process that requires even more help from molecular chaperones.

Is the appearance of chaperones a miracle of evolution on Earth? In fact, chaperoning might be a common feature of several ancient proteins and RNAs. If monomers are polymerized in the presence of their polymers, they acquire at least a low-affinity binding of the polymer[20]. Thus, ancient enzymes might themselves have possessed weak chaperone activity. Primitive archaic chaperones could well have been present even in the putative 'primordial soup'. From these inefficient predecessors, the complexity of proteinaceous chaperones we observe today has developed[21]. RNA-based protein chaperones have not been discovered yet, but there are numerous examples for the stabilization of protein structure by RNA in RNA–protein complexes[22].

Conclusions

High cooperativity and weak, non-complementary interactions in protein folding may significantly contribute to the fact that molecular evolution selected ancient preproteins as major catalysts. Later in molecular evolution, folding of the more-efficient, more-flexible, larger protein enzymes required the help of molecular chaperones, which may have been mandatory for the evolution of our present-day catalysts.

Acknowledgements

The author is an International Research Scholar of the Howard Hughes Medical Institute and acknowledges the stimulating ideas of D. Draper (Department of Chemistry, Johns Hopkins University, Baltimore, MD, USA), the useful comments of A. Finkelstein (Institute of Protein Research, Russian Academy of Sciences, Pushchino, Russia), D. Herschlag (Department of Biochemistry, Stanford University, Stanford, CA, USA), G. H. Lorimer (Du Pont de Nemours, Wilmington, DE, USA) and the two anonymous referees of this manuscript.

References

1 Lochse, P. A. and Szostak, J. W. (1996) *Nature* 381, 442–444
2 Smith, J. M. and Szathmary, E. (1995) *The Major Transitions in Evolution*, W. H. Freeman

3 Brack, A. and Orgel, L. E. (1975) *Nature* 256, 383–387
4 Wachteshauser, G. (1988) *Microbiol. Rev.* 52, 452–484
5 Joyce, G. F. (1989) *Nature* 338, 217–224
6 Piccirilli, J. A. *et al.* (1990) *Nature* 343, 33–37
7 Finkelstein, A. (1994) *Curr. Opin. Struct. Biol.* 4, 422–428
8 Dill, K. A. *et al.* (1995) *Protein Sci.* 4, 561–602
9 Dinner, A. R., Sali, A. and Karplus, M. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 8356–8361
10 Cech, T. R. (1993) in *The RNA World* (Gesteland, R. F. and Atkins, J. F., eds), pp. 239–269, Cold Spring Harbor Laboratory Press
11 Herschlag, D. (1995) *J. Biol. Chem.* 270, 20871–20874
12 Draper, D. E. (1996) *Trends Biochem. Sci.* 21, 145–149
13 Narlikar, G. J. and Herschlag, D. (1997) *Annu. Rev. Biochem.* 66, 19–59
14 Koshland, D. E., Jr (1976) *Fed. Proc.* 35, 2104–2111
15 Monod, J., Wyman, J. and Changeux, J-P. (1965) *J. Mol. Biol.* 12, 88–118
16 Kacser, H. and Beeby, R. (1984) *J. Mol. Evol.* 20, 38–51
17 Bryngelson, J. D. *et al.* (1995) *Protein Struct. Funct. Genet.* 21, 167–195
18 Todd, M. J. *et al.* (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 4030–4035
19 Orengo, C. A. *et al.* (1994) *Nature* 372, 631–634
20 Pande, V. S. *et al.* (1994) *Proc. Natl. Acad. Sci. U. S. A.* 91, 12976–12979
21 Hartl, F. U. (1996) *Nature* 381, 571–580
22 Allain, F. H. *et al.* (1996) *Nature* 380, 646–650

Kinetic data reliability

In a recent *Talking Point* article[1], Schuck and Minton propose two simple consistency tests to ascertain the reliability of kinetic measurements performed with surface plasmon resonance (SPR) biosensors. We agree totally with their suggestions and the necessity to perform these tests. However, we believe that the inclusion of some of our published experiments[2] as examples of data that fail these tests is inappropriate.

The interaction we studied was characterised by substantial re-binding of dissociating material to the surface. This is a common problem in binding studies, especially for interactions with fast association rates, and can be overcome either by 'infinite' dilution (which is not practical) or by the addition of an excess of competing ligand during the dissociation phase to prevent re-binding[3] (see Fig. 2a in Ref. 2). Schuck and Minton wrongly compare the k_{diss} values obtained in this way (and shown in our Table I) with the k_{obs} calculated from Fig. 1d. The k_{obs} should be and is consistent with the 'apparent' low k_{diss} value obtained in buffer flow (Fig. 2a), not with the high k_{diss} value obtained in the presence of the competing peptide.

Concerning consistency test 1, Schuck and Minton somehow calculate a value of approximately 1.3 nM for the equilibrium constant K_D from our data shown in Fig. 1a. These are six points with the first and last differing by tenfold. We are sure that Schuck and Minton would agree that, in order to obtain an accurate value, a range of concentrations differing by at least several hundred-fold has to be used, including points close to saturation[3]. It is also obvious from the same figure that true equilibrium was not reached during the time course of these experiments, again making these data unsuitable for a K_D calculation. Moreover, Schuck and Minton compare the value of 1.3 nM with the K_D obtained using the 'true' off rate from Table I, not the 'apparent' one, as discussed above. We have recently repeated these experiments with recombinant proteins that are not fused to GST. This diminishes the re-binding problem and equilibrium is reached much faster than with the GST-fusion proteins. Using this approach, both the association and dissociation phases can be fitted accurately with simple kinetic models, confirming our published finding of fast association and very fast dissociation rates for SH2 domain–phosphopeptide interactions. The calculated K_D values also match those defined by equilibrium binding assays. We conclude that our published data are not 'self-inconsistent' and that, provided experimental conditions and materials are chosen carefully, SPR biosensors using continuous flow (BIAcore) can be reliably used for calculation of kinetic and equilibrium constants.

References

1 Schuck, P. and Minton, A. P. (1996) *Trends Biochem. Sci.* 21, 458–460
2 Panayotou, G. *et al.* (1993) *Mol. Cell. Biol.* 13, 3567–3576
3 Limbird, L. E. (1986) in *Cell Surface Receptors: a short course on theory and methods*, Martinus Nijhoff Publishing

GEORGE PANAYOTOU AND MIKE WATERFIELD

Ludwig Institute for Cancer Research, 91 Riding House Street, London, UK W1P 8BT.

Copyright © 1997, Elsevier Science Ltd. All rights reserved. 0968-0004/97/$17.00 PII: S0968-0004(97)01037-2

Reply to Panayotou and Waterfield

Panayotou *et al.*[1] have evaluated the dissociation rate constant k_- via analysis of a dissociation experiment conducted in the presence of a large excess of competing peptide. We agree that this value of k_- is much more likely to reflect the true chemical rate constant for dissociation of peptide from SH2 domain than the value obtained from analysis of the association experiment in the context of the elementary 1:1 association model. However, the use of a value of k_+ obtained from the same analysis of the same association experiment with the same oversimplified model only compounds the internal inconsistency of the calculation of the association equilibrium constants K_a (mistakenly labeled dissociation constants) in Table I of Panayotou *et al.*[1]. We have shown elsewhere[2] that mass transport effects can result in quasi-linear plots of dR/dt vs R, analysis of which via the elementary association model (neglecting mass transport) yields apparent rate constants k_+ and k_-^{ass} that both are far below the actual intrinsic chemical rate constants for binding. If one calculates an apparent equilibrium association constant by dividing an artifactually low estimate of k_+ by a realistic estimate of k_-, then the resulting estimate of K_a will be depressed by the same factor as k_+.

The internal inconsistency inherent in the analysis employed by Panayotou and colleagues[1] is evident in the limiting long-time behavior of the association experiments plotted in Fig. 1a, c. One may estimate the equilibrium response R_∞ of the system from Fig. 1a, d, as described in our note (Table I caption) or, equivalently, by extrapolating the straight lines plotted in Fig. 1c to the x-intercept (dR/dt = 0). The resulting dependence of R_∞ on free ligand concentration can be modeled via our Eqn 4 to obtain a reasonable estimate of K_D (or K_a). We emphasize that this calculation is not subject to the influence of mass transport. When done, one obtains a value of K_a approximately equal to 1 nM^{-1} (Ref. 1) for the particular set of data plotted in Fig. 1. It is quite impossible to describe the long-time limit of these data using the value of 0.0237 nM^{-1} (Ref. 1) for the association equilibrium constant as reported by the authors.

References

1 Panayotou, G. *et al.* (1993) *Mol. Cell. Biol.* 13, 3567–3576
2 Schuck, P. and Minton, A. P. (1996) *Anal. Biochem.* 240, 262–272

PETER SCHUCK AND ALLEN P. MINTON

Laboratory of Biochemical Pharmacology, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA.

Published by Elsevier Science Ltd. 0968-0004/97 PII: S0968-0004(97)01038-4