Induced fit in RNA–protein recognition

James R. Williamson

Two generalizations can be drawn from the recent rapid progress in understanding RNA–protein interactions. First, there is a great diversity of observed protein and RNA structural motifs. Second, formation of almost every RNA–protein complex that has been characterized involves conformational changes in the protein, the RNA, or both. The role of these conformational changes in the biological function of RNA–protein complexes is not at all clear. Whether or not conformational changes are a critical feature of ribonucleoprotein complex assembly or are an unimportant mechanistic detail, the ubiquity of these changes warrants careful consideration of their implications.

There are many terms in popular usage to describe conformational changes that accompany binding: induced fit, cofolding, mutually induced fit, ligand-induced conformational change, and tertiary structure capture. These terms are used to describe various aspects of local macromolecular folding in the formation of intermolecular complexes. Typically, there is only information available about the structure and relative energy of the free and bound states of the RNA and protein. That is, there is no specific information about the binding pathway, which might include a series of intermediates. Rather, we only know that the free protein and RNA are flexible or have particular conformations, and that they are more rigid or have different conformations in the final RNA–protein complex.

Induced fit in RNA–protein complexes

In the case of RNA–protein complexes, induced fit upon binding can be seen in either the RNA or the protein components[1]. The free protein and/or RNA may have a disordered or flexible region, or simply a different conformation than the bound form. There are three basic classes of induced fit mechanisms for RNA–protein complex formation (Fig. 1). In the first example, the RNA undergoes a conformational change upon protein binding, but the conformation of the protein is relatively unchanged (Fig. 1*a*). In the second example, the converse is true, and the RNA undergoes little conformational change, while the protein undergoes a significant change (Fig. 1*b*). Finally, there is mutually induced fit, or cofolding, where both the RNA and protein components change conformation (Fig. 1*c*). Each of these different cases results in formation of a stable RNA–protein complex that has properties that are different from the free forms of the RNA and protein, and this complex results in a particular biological function.

In the following discussion, I will focus on induced fit observed in four particular RNA–protein complexes. However, the general mechanistic features described apply to many other complex-forming reactions as well. Induced fit is an important feature of tRNA recognition by tRNA aminoacyl synthetases[2–5], and has been extensively studied by structural and enzymological methods. Induced fit is also widely observed in RNA–peptide complexes, such as HIV Tat–TAR[6,7] and Rev–RRE complexes[8,9] and the BIV Tat–TAR complex[10,11]. In addition, most RNA aptamers that have been selected to bind to small molecule ligands also undergo induced fit upon complex formation[12]. Thus, induced fit appears to be common in ligand binding reactions of RNA in particular.

The RNA–protein complexes described below are among the best characterized complexes for which there is information about both the free and the bound states of the protein and RNA.

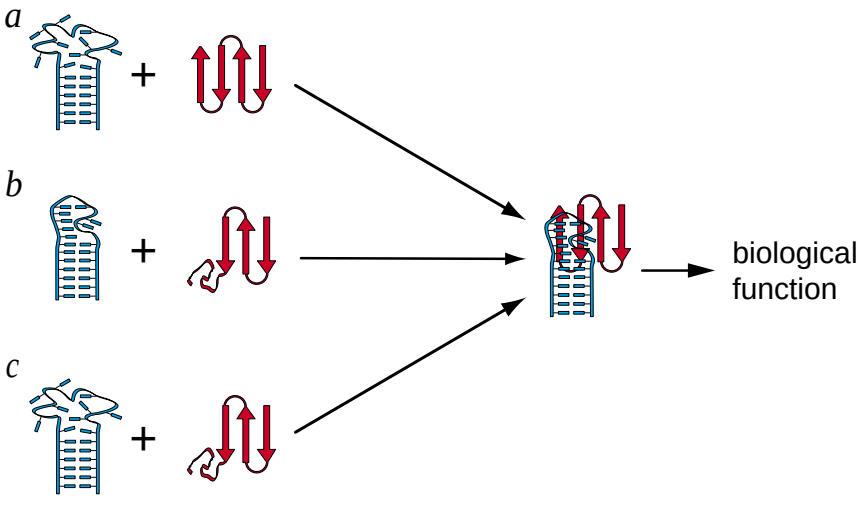

Fig. 1 Possible mechanism for induced fit in RNA–protein complex formation. The RNA (blue) and protein (red) form a complex, which results from ***a***, protein-induced RNA folding, ***b***, RNA-induced protein folding, or ***c***, cofolding. The partially folded RNA is schematically diagrammed with a fixed helical region and a flexible or disordered loop region. The partially folded protein is schematically represented with one structural element flexible, in this case a β-strand. The net overall effect upon formation of the RNA–protein complex is to order the loop of the RNA and to fold the protein, resulting in an intimate RNA–protein interface in the complex. The RNA–protein complex then mediates its particular biological function. Note that this is the overall thermodynamic view of RNA–protein complex formation, which only considers the initial and final states of the RNA and protein. The pathway or mechanism for folding and binding cannot be determined from equilibrium measurements alone.

Department of Molecular Biology and the Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California 92037, USA.

Correspondence should be addressed to J.R.W. *email: jrwill@scripps .edu*

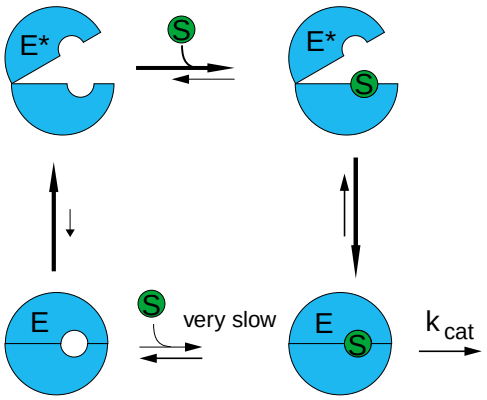

Fig. 2 The classic enzymological view of induced fit. The enzyme (E) exists predominantly in an inactive conformer (E*), which is capable of binding substrate (S). Usually, binding of S directly to E occurs very slowly, or not at all. After binding of substrate, a conformational change occurs from E*S to ES, which is the active enzyme substrate complex. Thus, binding of the substrate 'induces' formation of the complete substrate binding site and the active conformation of the enzyme.

In these examples, induced fit behavior typical of each of the above classes of mechanisms can be seen. Clearly, this is a common feature for RNA–protein complex formation, and it remains an important goal of structural biology to understand the mechanistic and energetic consequences of the mutually induced fit mechanism that is so ubiquitous. Formally, conformational changes that occur upon binding between two macromolecules can be considered in the same thermodynamic framework that was developed for enzyme–substrate complexes, which represents the classic definition of induced fit.

Classic induced fit

This concept was first introduced to explain how some enzymes were activated for catalysis by substrate binding[13]. The free enzyme exists in an inactive state, and upon binding of the substrate, the conformation of the enzyme changes, converting the enzyme to its active form. Hexokinase is a classic example of this, where binding of the sugar substrate activates the active site for reaction with ATP[14]. If the enzyme active site were to hydrolyze ATP in the absence of bound substrate, ATP would be wasted. Requiring the substrate to bind to activate the enzyme avoids this problem.

A pictorial representation of the induced fit mechanism[15] is shown in Fig. 2. The inactive enzyme (E*) exists in equilibrium with an active form (E), but the E* form predominates. The inactive enzyme (E*) has a substrate binding site between the hinge region linking two domains. The substrate can bind to this site to form the inactive Michaelis complex (ES*), which rapidly closes to form the active complex (ES), which then goes on to react. In the case of hexokinase, the free enzyme exists in the open E* conformation, which is incapable of hydrolyzing ATP. The active site is only completely formed after substrate binding and subsequent conformational change to form ES. This in effect guarantees that ATP hydrolysis can occur only in the presence of the substrate. Thus, induced fit ensures that the biological function of the enzyme, catalysis, is not enabled until after a proper substrate is bound.

Protein-induced RNA folding in the S15–rRNA complex

The ribosomal protein S15 is a primary binding protein that is important for nucleating the assembly of the central domain, which forms the platform region of the 30S subunit. S15 binds to a three-helix junction in 16S ribosomal RNA. The structure of the S15–rRNA complex has been solved[16,17] (Fig. 3). X-ray[18] and NMR[19] structures have been determined for the free form of S15 as well. Although these two structures differ in the position of the N-terminal α-helix, the core structure of the remaining three helices is very similar, and is similar to the bound form of the protein. It appears that the protein may be somewhat more flexible or dynamic in the free form than in the bound form[19], but the structures are essentially the same. In the case of the free RNA, transient electric birefringence experiments showed that in the absence of S15 and Mg^{2+} ions, the angle between the three helices is nearly equal at ~120° (ref. 20). In the S15 complex, two of the three helices are coaxially stacked, while the third makes an acute angle with the main part of the S15 binding site, both in solution and in the crystal structures[16,17,20]. In the case of the S15–rRNA complex, the RNA undergoes a major conformational change, while the protein undergoes only minor changes and stabilization upon RNA binding.

RNA-induced protein folding in the L11–rRNA complex

The ribosomal protein L11 forms part of the interaction site with elongation factors (EF-G and EF-Tu) and is part of the binding site for the antibiotic thiostrepton. L11 binds to a four-way junction in 23S rRNA, and the crystal structure of the L11–rRNA complex has been determined[21,22] (Fig. 4). The RNA tertiary structure of the L11 binding site is unstable in the absence of magnesium ions and L11 protein. Although there is no quantitative information on the conformation of the free RNA, the RNA tertiary structure must be properly formed for the protein to bind, and it is likely that the free and bound conformations of the RNA site are quite similar[23]. The NMR structure of part of the L11 protein has been determined in the free

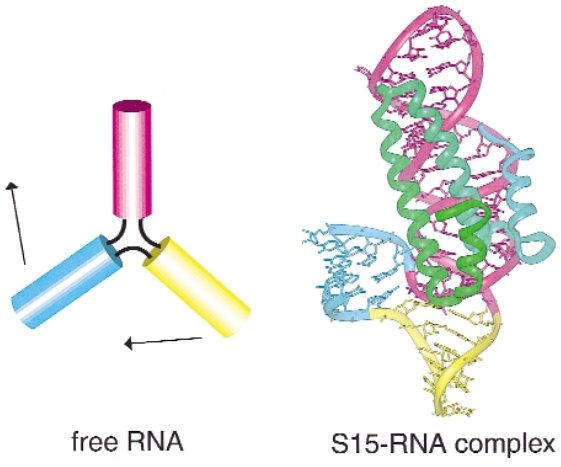

Fig. 3 Protein-induced RNA folding in the S15–rRNA complex. The free and bound conformations of S15 are extremely similar. In contrast, the free form of the RNA is quite different from the bound form, in the absence of magnesium ions. Although no detailed structural information is available on the free form of the S15 binding site RNA, biophysical studies have shown that the angle between the three helices is nearly equal, and the orientation of the helices in the free RNA is represented by the colored cylinders. The conformational change involves co-axial stacking of two helices (magenta/yellow) and hinge movement of the third helix (cyan). The arrows show the direction of the movements required to form the RNA structure in the S15–RNA complex.

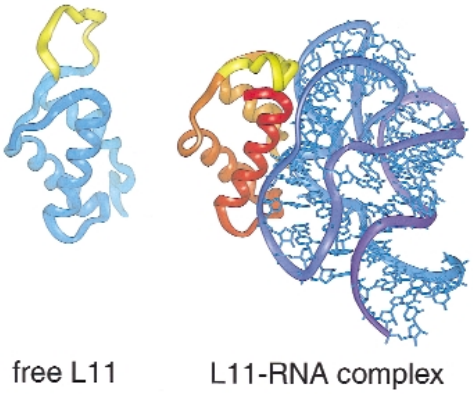

Fig. 4 RNA-induced protein folding in the L11–rRNA complex. One loop in L11 undergoes a significant conformational change and organization upon binding to its ribosomal RNA site. The RNA tertiary structure in the L11 site requires magnesium ions to form, and is required for protein binding. Most of the L11 protein changes conformation very little, but the loop that does change conformation (shown in yellow) makes RNA contacts in the RNA–protein complex.

form[24]. Most of the L11 protein structure is the same in either the free or the bound state, but there is a flexible loop in the free protein that becomes ordered upon binding to the RNA, and this loop is located at the interface with the RNA. In this case, the induced fit occurs in the protein, which changes conformation in order to form additional interactions with the RNA in the complex.

Mutually induced fit in the U1A–UTR complex

The N-terminal RNA binding domain of the U1A protein, which is a classic ribonuclear particle (RNP) consensus domain binds to an internal loop in the 3′ untranslated region of its own mRNA to regulate its polyadenylation. This complex is one of the best characterized in terms of structural[25–28] and thermodynamic[29–31] information, and there are structures available of the free forms of both molecules[26,32] as well as the complex[27,28](Fig. 5). The internal loop of the RNA is highly flexible and somewhat disordered, but exhibits stacking interactions that are different from those observed in the RNA–protein complex. In the free form of the protein, the C-terminal helix is packed against the face of the β-sheet that is used for RNA recognition. In the complex, the helix is moved away from the sheet to allow intimate contact between the RNA and the protein. In this case, the formation of the RNA–protein complex requires significant reorganization of both of the partners in a true case of cofolding.

Consequences of induced fit for structure prediction

One of the important consequences of induced fit is that it can make structure prediction difficult. If the free structure is different in part from the bound structure, it can mean that there is not enough information in the sequence of a protein to completely determine its final structure in a functional complex. An interesting example of this potential problem is the ribosomal protein L30. As is the case for the other RNA–protein complexes discussed above, binding of L30 to its mRNA binding site is accompanied by conformational changes in both the protein and the RNA[33,34] (Fig. 6). The RNA internal loop is flexible and not highly structured, but adopts a highly stacked and hydrogen bonded structure in the complex. One of the helices in L30 is unstable and only partly formed in the free protein, but becomes more ordered upon binding. This region of the protein contains several hydrophobic residues that are exposed on the surface of the protein, but that become buried at the RNA interface upon binding.

The structure of the L30 protein was submitted to the third Critical Assessment of Structure Prediction (CASP3) competition (target T0077)[35,36]. In the tertiary structure prediction trials, coordinates of completely folded predictions were scored against the coordinates of the bound conformation of L30. None of the groups submitted a completely accurate prediction, but many of the closest predictions had the correct secondary structure and tertiary packing for the N-terminal three-fourths of the protein. Interestingly, the portion of the protein that was most poorly predicted was the portion of the protein that undergoes an induced conformational change upon RNA binding. NMR relaxation studies have shown that the free L30 is very flexible in this region, and the secondary structure is not stably maintained. In effect, the sequence of the L30 protein is not sufficient to specify the secondary structure or the tertiary fold in this region, which has two relevant consequences. First, the free protein is flexible, and second, it is very difficult to predict the local structure. The full specification of the protein secondary structure and packing is achieved only when bound to the RNA, which was not used in the prediction. It may well be that such ambiguities in predictions could be a signature of regions of a structure that may be important for intermolecular interactions.

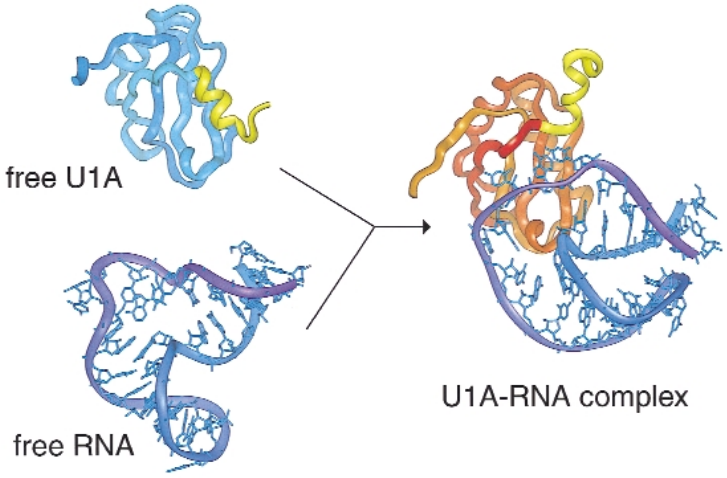

Fig. 5 Cofolding in formation of the U1A–UTR complex[27]. The C-terminal helix of U1A changes its position upon binding of the 3′ UTR RNA binding site (shown in yellow). At the same time, the free UTR RNA is not pre-organized to form the proper structure for U1A recognition. Significant conformational changes occur in both the RNA and the protein upon complex formation — that is, cofolding coupled to binding, or mutually induced fit.

Conclusions

The thermodynamics of DNA–protein interactions have been extensively studied, and it has been clearly shown that local protein folding is coupled to DNA binding in a number of cases. There is a large negative heat capacity change that correlates with local folding

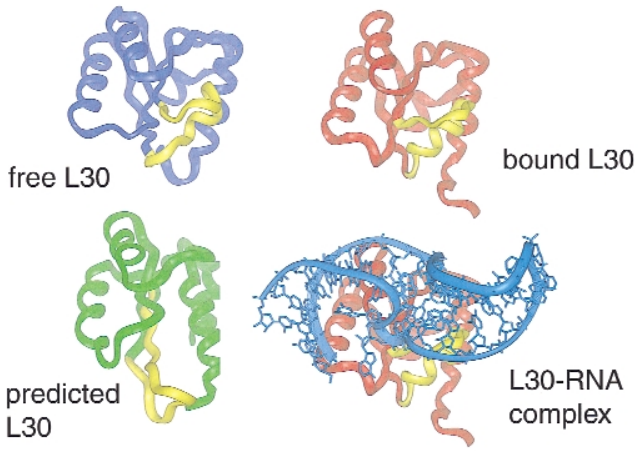

Fig. 6 Induced fit presents an obstacle to protein structure predication. The L30–mRNA complex is another example of cofolding or mutually induced fit: both the RNA and the protein undergo conformational changes upon binding. Most of the changes in the L30 conformation are restricted to one particular β-strand-loop-helix segment at the C-terminus of the protein (shown in yellow). This region is flexible in the free form (blue), and undergoes a conformational change and ordering upon binding the RNA (red). The L30 protein was submitted as a target for the CASP3 structure prediction competition (target T0077). One of the better predictions is shown (T0077TS163_2), where essentially the N-terminal three-quarters of the protein were correctly predicted for secondary structure and overall packing (green). The portion of the structure that agreed least with the real structure contained the region that undergoes induced fit upon RNA binding. The information to form the tertiary structure is not completely contained in the protein sequence, resulting in both flexibility in the free form, and poor predictability. Part of the information to fold the structure lies in the contacts to the RNA, which could not be taken into account in the structure prediction.

upon DNA binding, and the heat capacity change is attributed to the burial of exposed hydrophobic surface upon binding, which is consistent with induced fit[37]. Although progress is being made along similar lines in RNA–protein complexes, there is not yet a clear picture of what the 'thermodynamic signature' of mutually induced fit in RNA–protein interactions will be.

The conformational changes that occur require input of energy to fold the RNA or protein molecules. The energy of binding is used to overcome this energy cost for folding. Thus, induced fit has an intrinsic thermodynamic penalty, and there must be a reason why induced fit is so often observed. If the two components forming the complex were rigid, then no binding energy would have to be expended to pay the cost of folding. Therefore, tighter binding could be achieved using rigid components. Why then waste energy on folding? One possible explanation is that it is not easy to form the intimate interface required for molecular recognition by the docking of rigid components, and that flexibility lowers the energy barrier to complex formation. This is formally equivalent to the hexokinase mechanism (Fig. 2), in which the substrate cannot easily bind to a closed active site in the active form of the enzyme (E).

What is the role of induced fit in RNA–protein complex formation? The answer is certainly complicated. Again, a simple analogy can be drawn to the classic induced fit mechanism for hexokinase (Fig. 2). In this case, the enzymatic activity of the protein (ATP hydrolysis) would be deleterious if the sugar substrate were not bound, and so the active conformation (E) is only populated when the substrate is actually bound. In most cases, RNA–protein complex formation also results in biological activity (Fig. 1) formally analogous to the enzymatic activity in Fig. 2. There is some biological function, call it the 'signal' of the complex, which should not be manifested by either of the components alone. If the structure of the free and bound RNA and protein were the same, there would be the possibility of inappropriate signals. Induced fit, and the more general mechanism of mutually induced fit, ensures that the signal from the RNA–protein complex only results after appropriate conformational changes occur. Apparently, it is generally critical in biology to distinguish free and bound. Induced fit is a convenient way to use the energy of binding to drive conformational changes to ensure that this is the case.

Received 31 July, 2000; accepted 22 August, 2000.

1. Frankel, A.D. *Nature Struct. Biol.* **6**, 1081–1083 (1999).
2. Alexander, R.W. & Schimmel, P. *Biochemistry* **38**, 16359–16365 (1999).
3. Fersht, A.R., Knill-Jones, J.W., Bedouelle, H. & Winter, G. *Biochemistry* **27**, 1581–1587 (1988).
4. Poterszman, A., Delarue, M., Thierry, J.C. & Moras, D. *J. Mol. Biol.* **244**, 158–167 (1994).
5. Ramesh, V., Mayer, C., Dyson, M.R., Gite, S. & RajBhandary, U.L. *Proc. Natl Acad. Sci. USA* **96**, 875–880 (1999).
6. Puglisi, J.D., Tan, R., Calnan, B.J., Frankel, A.D. & Williamson, J.R. *Science* **257**, 76–80 (1992).
7. Aboul-ela, F., Karn, J. & Varani, G. *J. Mol. Biol.* **253**, 313–332 (1995).
8. Battiste, J.L. *et al. Science* **273**, 1547–1551 (1996).
9. Ye, X., Gorin, A., Ellington, A.D. & Patel, D.J. *Nature Struct. Biol.* **3**, 1026–1033 (1996).
10. Puglisi, J.D., Chen, L., Blanchard, S. & Frankel, A.D. *Science* **270**, 1200–1203 (1995).
11. Ye, X., Kumar, R.A. & Patel, D.J. *Chem. Biol.* **2**, 827–840 (1995).
12. Hermann, T. & Patel, D.J. *Science* **287**, 820–825 (2000).
13. Koshland, D.E., Jr. *Proc. Natl Acad. Sci. USA* **44**, 98 (1958).
14. Bennett, W.S., Jr. & Steitz, T.A. *Proc. Natl Acad. Sci. USA* **75**, 4848–4852 (1978).
15. Fersht, A. *Enzyme structure and mechanism* (Freeman and Company, New York; 1985).
16. Agalarov, S.C., Sridhar Prasad, G., Funke, P.M., Stout, C.D. & Williamson, J.R. *Science* **288**, 107–113 (2000).
17. Nikulin, A. *et al. Nature Struct. Biol.* **7**, 273–277 (2000).
18. Clemons, W.M., Jr., Davies, C., White, S.W. & Ramakrishnan, V. *Structure* **6**, 429–438 (1998).
19. Berglund, H., Rak, A., Serganov, A., Garber, M. & Hard, T. *Nature Struct. Biol.* **4**, 20–23 (1997).
20. Orr, J.W., Hagerman, P.J. & Williamson, J.R. *J. Mol. Biol.* **275**, 453–464 (1998).
21. Conn, G.L., Draper, D.E., Lattman, E.E. & Gittis, A.G. *Science* **284**, 1171–1174 (1999).
22. Wimberly, B.T., Guymon, R., McCutcheon, J.P., White, S.W. & Ramakrishnan, V. *Cell* **97**, 491–502 (1999).
23. Blyn, L.B., Risen, L.M., Griffey, R.H. & Draper, D.E. *Nucleic Acids Res.* **28**, 1778–1784 (2000).
24. Hinck, A.P. *et al. J. Mol. Biol.* **274**, 101–113 (1997).
25. Allain, F.H., Howe, P.W., Neuhaus, D. & Varani, G. *EMBO J.* **16**, 5764–5772 (1997).
26. Gubser, C.C. & Varani, G. *Biochemistry* **35**, 2253–2267 (1996).
27. Oubridge, C., Ito, N., Evans, P.R., Teo, C.H. & Nagai, K. *Nature* **372**, 432–438 (1994).
28. Varani, L. *et al. Nature Struct. Biol.* **7**, 329–335 (2000).
29. Kranz, J.K. & Hall, K.B. *J. Mol. Biol.* **275**, 465–481 (1998).
30. Kranz, J.K. & Hall, K.B. *J. Mol. Biol.* **285**, 215–231 (1999).
31. Reyes, C.M. & Kollman, P.A. *J. Mol. Biol.* **297**, 1145–1158 (2000).
32. Avis, J.M. *et al. J. Mol. Biol.* **257**, 398–411 (1996).
33. Mao, H., White, S.A. & Williamson, J.R. *Nature Struct. Biol.* **6**, 1139–1147 (1999).
34. Mao, H. & Williamson, J.R. *J. Mol. Biol.* **292**, 345–359 (1999).
35. Sternberg, M.J., Bates, P.A., Kelley, L.A. & MacCallum, R.M. *Curr. Opin. Struct. Biol.* **9**, 368–373 (1999).
36. Zemla, A., Venclovas, C., Moult, J. & Fidelis, K. *Proteins. Suppl* **3**, 22–29 (1999).
37. Spolar, R.S. & Record, M.T., Jr. *Science* **263**, 777–784 (1994).