TIMELINE

From the first protein structures to our current knowledge of protein folding: delights and scepticisms

Alan R. Fersht

Abstract | Every breakthrough that opens new vistas also removes the ground from under the feet of other scientists. The scientific joy of those who have seen the new light is accompanied by the dismay of those whose way of life has been changed for ever. The publication of the first structures of proteins at atomic resolution 50 years ago astounded and inspired scientists in every field, but caused others to flee or scoff. That advance and every subsequent paradigm-shifting breakthrough in protein science have met with some resistance before universal acceptance. I relate these events and their impact on the field of protein folding.

Well-funded laboratories can today run the gamut of protein studies; from molecular biology, biophysics and computation to X-ray crystallography and nuclear magnetic resonance (NMR). We take it for granted that protein science is underpinned by a multitude of complementary disciplines and methods. But it is easy to forget that each of the underpinning technologies represented a major breakthrough and each had to overcome initial, and sometimes continuing, scepticism to be universally adopted. The progress from the initial determination of protein structures to our present understanding of protein folding typifies this process of simultaneous delight and scepticism about new methods before their eventual acceptance. In this article, I do not track all of the developments in protein science, just some key advances, mainly technological and experimental, that have led to our present understanding of protein folding.

The first 3D structures

For half a century, the three-dimensional structures of proteins were studied indirectly. Details of a mechanism and its relationship to structure could only be inferred from the properties of the protein. In 1958, fifty years ago this year, a landmark paper[1] on the structure of myoglobin by John Kendrew and co-workers showed the first 3D structure of a protein (FIG. 1; TIMELINE). Its resolution was too low to show the atomic details of what seemed to be a multiply bent sausage, and the crude map gave only a hint of the revolution to come. The subsequent low-resolution structure of haemoglobin, which resembled four myoglobin molecules stacked together (FIG. 2), was reported by Max Perutz and colleagues in 1960 (REF. 2) and gave the first intimation of protein families.

> "We take it for granted that protein science is underpinned by a multitude of complementary disciplines and methods."

The floodgates burst open after the publication of the high-resolution structure of hen egg-white lysozyme by David Phillips and his colleagues[3]. The detailed molecular structure at atomic resolution inspired the proposal of a chemical mechanism for the enzymatic reaction that this protein catalysed, parts of which have stood the test of time. Soon afterwards, the structures of α-chymotrypsin[4], ribonuclease[5], carboxypeptidase[6] and *Staphylococcus aureus* nuclease[7] followed, each of which gave new information on the protein structure and its function. Molecular biologists (as structural biologists called themselves then) were convinced that function and mechanism would always spring obviously from structure. It was therefore a "disappointment" to Francis Crick that this was not generally true (F. Crick, personal communication). Regardless, the rules had changed: from then on, mechanistic work on proteins had to be based on atomic-level-resolution structures.

The beauty of the structures, and especially the proposed atomic-level mechanism for lysozyme, inspired a new generation of experimental scientists to test proposals for mechanisms of action that were based on the protein structure with ingenious new experiments. Theoreticians worked in parallel to calculate the interactions within proteins and test ideas, such as whether an enzyme could distort a substrate towards the structure of its transition state. Perhaps the most illuminating work of all came from Perutz, whose subsequent studies on the conformational changes of haemoglobin showed much of the full power of X-ray crystallography combined with both protein chemistry and work on natural mutants, which had been isolated by Herman Lehmann[8].

A wave of enthusiasm engulfed the world of protein science, but not all of the older guard were thus inspired. Fine work had been done over the years on indirectly inferring the structures of active sites of enzymes through experiments on their activity with different substrates and by identifying functional groups from their pH titrations and chemical modification. The combination of protein chemistry and kinetics had laid the foundation stones of modern protein science. Some traditional enzymologists and protein chemists focused their attention on proteins that they thought would never be crystallized in order to avoid the new technocrats, who they thought would put them out of business but who were to become, in fact, indispensable partners. Others complained that crystalline proteins had unrepresentative structures and that protein crystals were unnatural, which, in a small minority of cases, is true. To make matters worse, some crystallographers kept their structure coordinates as closely guarded secrets and regarded the protein chemists as vultures. So, some enzymologists

were indeed frozen out and had little choice but to flee, and it took three decades for every journal to insist on depositing atomic coordinates as a prerequisite for publication.

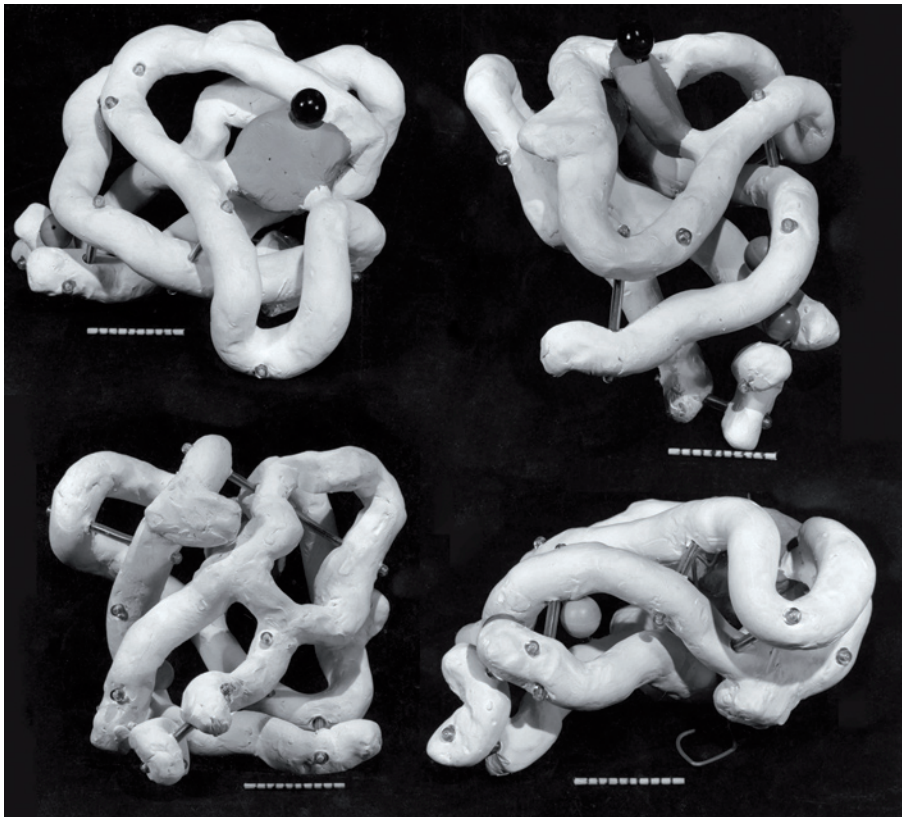


Figure 1 | **Three-dimensional structure of myoglobin.** The low-resolution structure of myoglobin that was published by John Kendrew and colleagues in 1958 (REF. 1). This figure in the *Nature* paper was reconstructed by the author using the original figures in the archives of the Medical Research Council Laboratory of Molecular Biology, Cambridge, UK. Polypeptide chains are in white and the grey disc represents the haem group. The three spheres show positions at which heavy atoms were attached to the molecule (black, Hg of *p*-chloro-mercuri-benzene-sulphonate; dark grey, Hg of mercury diammine; light grey, Au of auri-chloride). The marks on the scale are 1 Å apart.

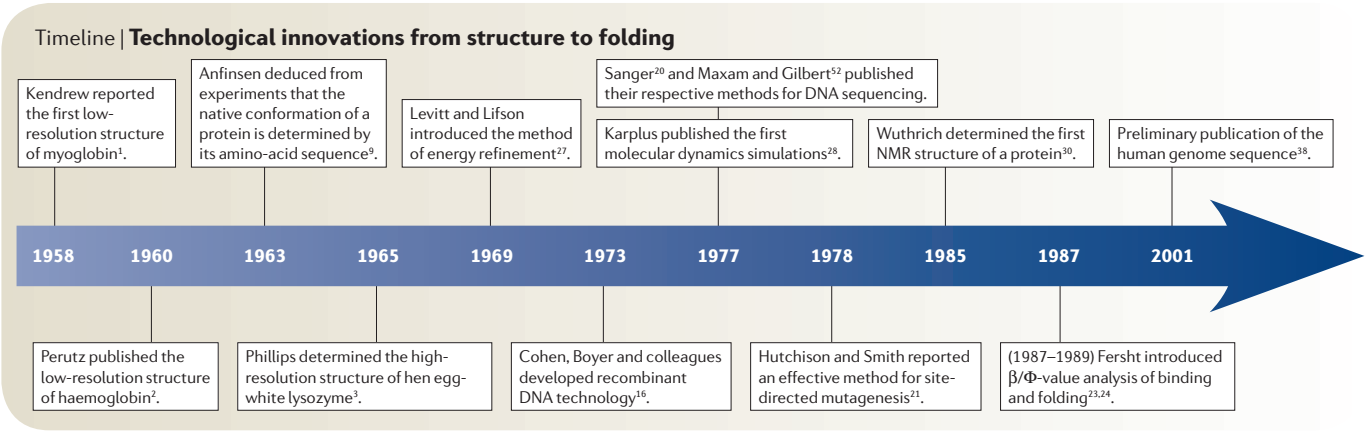
From structure to protein folding

Solving the structures of proteins at high resolution uncovered a new problem and initiated a novel field of research — that of protein folding. It was known from seminal experiments by Christian Anfinsen that small proteins could spontaneously refold from their denatured states, and so the primary structure (sequence) of a protein dictates its tertiary structure[9]. The 'protein folding problem' consists of two parts: first, the tertiary structures of proteins need to be predicted from their primary sequences, and second, the pathway of folding and unfolding must be predicted.

Cyrus Levinthal and others, such as Michael Levitt and Oleg Ptitsyn, wanted to predict tertiary structures by predicting their folding pathways. Levinthal famously pointed out in almost offhand remarks during a meeting in 1969 (REF. 10) that it seemed impossible that an unfolded protein could fold spontaneously by a random process on a biological time scale. Mechanisms were proposed that could overcome the 'Levinthal paradox' by simplifying the folding process and breaking it down into subprocesses that could occur stepwise. Ptitsyn suggested that folding could occur in a hierarchical process with the initial rapid formation of secondary structures, such as α-helices or β-sheets. At each step, the formation of a new layer of structure stabilized the previous one[11]. The framework mechanism was tested by Robert Baldwin[12] and formulated analytically and computationally in the diffusion–collision model of Karplus and Weaver[13]. Ptitsyn further proposed that proteins could undergo rapid hydrophobic collapse to form a 'molten globule', in which native tertiary interactions were rapidly formed that directed the subsequent formation of native secondary structure[14]. Levinthal had suggested a nucleation-growth mechanism in which the slow formation of a nucleus of secondary structure was followed by its rapid growth. Nucleation mechanisms went out of favour because the first proteins to be studied folded through kinetically observable intermediates. This is consistent with the framework and molten globule mechanisms, but not the nucleation mechanism, which implies only high-energy, kinetically silent intermediates.

How do we investigate protein folding mechanisms that involve the formation of a myriad of non-covalent interactions that give states that cannot be isolated and studied directly? One early approach, pioneered by Tom Creighton, was to study proteins that contained disulphide bonds and trap the covalent intermediates that contain both correctly and incorrectly formed disulphide bonds[15]. But modern mechanistic studies have to achieve atomic-resolution information and so it was necessary to develop methods that could analyse folding at the level of individual residues and atoms. It was also essential to extract and interpret the information encoded in a large number of proteins. The solution, albeit partial, of both parts of the folding problem required the introduction of new technologies, including recombinant DNA technology, protein engineering, advanced computer simulation and bioinformatics.

Recombinant DNA technology. The march of X-ray crystallography lost some momentum in the 1970s because studies were restricted to proteins that could be isolated from natural sources in large amounts. The development of recombinant DNA technology reinvigorated the field by providing the resources for making large quantities of previously rare and unknown proteins[16]. However, even Perutz was heard to comment when so many colleagues immediately took up DNA cloning that "they must have been bored with what they were doing to drop it so readily", and the Medical Research Council Laboratory of Molecular Biology in Cambridge, UK — the epicentre of the structural biology revolution where not only Perutz and Kendrew had been working but also Sydney Brenner, Crick, Aaron Klug, Cesar Milstein and Fred Sanger — was slow in adopting the technology. Recombinant DNA technology proved essential for protein folding studies as it provided a source of experimentally tractable proteins[17–19], which replaced an older generation of readily available proteins. The new proteins lacked disulphide bridges, the formation of which dominated the folding pathways of the former generation. The recombinant proteins were amenable to the all-important structure–function studies through protein engineering.



Protein engineering. DNA sequencing, in particular the method that was developed by Sanger[20] and its consequent refinements, profoundly altered the course of biology and protein science, without a murmur of dissent. Until the introduction of site-directed mutagenesis[21], and subsequently protein engineering[22] by the manipulation of genes, protein scientists were just observers and users of proteins. But by changing first individual amino-acid residues and then whole segments of proteins, precise structure–function–activity experiments could be designed and new functional proteins made.

There was a time lag of 5 years between the introduction of the technology for site-directed mutagenesis[21] and it being put into practice, initially to analyse enzyme catalysis[22]. The delay occurred, perhaps, because people were not sure how to use the technology rather than because they objected to it. Even so, there were initial murmurings that point mutations would radically alter the structure and function of proteins, which ill-designed mutations might, of course, do.

Figure 2 | **Max Perutz and John Kendrew.** Max Perutz (left) holding a balsawood model of the structure of haemoglobin solved at 6-Å resolution, and John Kendrew (right) holding a wire model of myoglobin solved at 1.4-Å resolution, which was determined remarkably soon after the low-resolution structure was published in 1958 (REF 1). Photo, which was taken in 1962, courtesy of the Medical Research Council Laboratory of Molecular Biology, Cambridge, UK.

Protein engineering proved indispensable to protein folding studies because it allowed structure–activity studies that monitored the perturbation of the mechanism and kinetics of folding by making tiny changes in the protein structure. In particular, Φ-value analysis (a way of estimating the extent of non-covalent bonding during folding by examining the changes in folding kinetics and equilibria following targeted mutation[23], a method that was first implemented as β-value analysis[23,24]) has given near-atomic-level descriptions of protein folding transition states. Φ-value analysis took more than a decade for its near-universal adoption, and even the concept of transition states in protein folding met with initial scepticism. Nevertheless, protein engineering studies led to the discovery of a widespread and fundamental mechanism of protein folding — nucleation condensation — whereby the protein seems to collapse around a diffuse nucleus with most of the final native interactions being partly formed[25]. The framework and molten-globule mechanisms are extreme manifestations of this basic mechanism[25,26].

Simulation and computer graphics. The first atomic representations of protein structures were built from pieces of screwed-together wire and inter-atomic distances were measured with rulers. In fact, one could buy plastic ball-and-sticks modelling kits. The first attempts at computer graphics were met with some wry amusement before such programs rapidly, and initially very expensively, displaced the old, crude mechanical models.

Computer calculations of energetics within proteins were used originally to refine the structures that were derived from fitting side chains to electron-density maps, by optimising the non-covalent interactions[27]. Molecular dynamics simulation then burst on to the scene[28]. The first attempts at simulation were severely limited by computational power and by approximations in energy functions, which are still imperfect today. Initially disparaged, these methods are now indispensable for understanding the mechanisms and folding of proteins as well as their intricate details because atomistic simulation, benchmarked by experimentation, is the only way of analysing a complete folding pathway and calculating the folding energetics[29].

NMR methods. NMR spectroscopy is such an important experimental procedure for determining structure in both solution and the solid state that it is difficult to conceive the battle that it took to establish its importance following the first determination of the structure of a protein[30]. Part of the problem was that NMR spectroscopy was seen as a rival, rather than complementary, technology to X-ray crystallography. NMR is, however, the only procedure available for studying the structures of disordered, denatured and partly folded states at atomic-level resolution[31]. This is done in solution, in which the proteins display their natural dynamics, which can also be analysed[32].

Stable folding intermediates have been engineered so that their structures can be solved directly by NMR methods[33]. High-energy folding intermediates that are only fractionally populated and are otherwise undetectable can be analysed structurally and their rates of interconversion measured[34]. The use of hydrogen–deuterium exchange between backbone NH groups and solvent water, introduced by Linderstrøm-Lang in

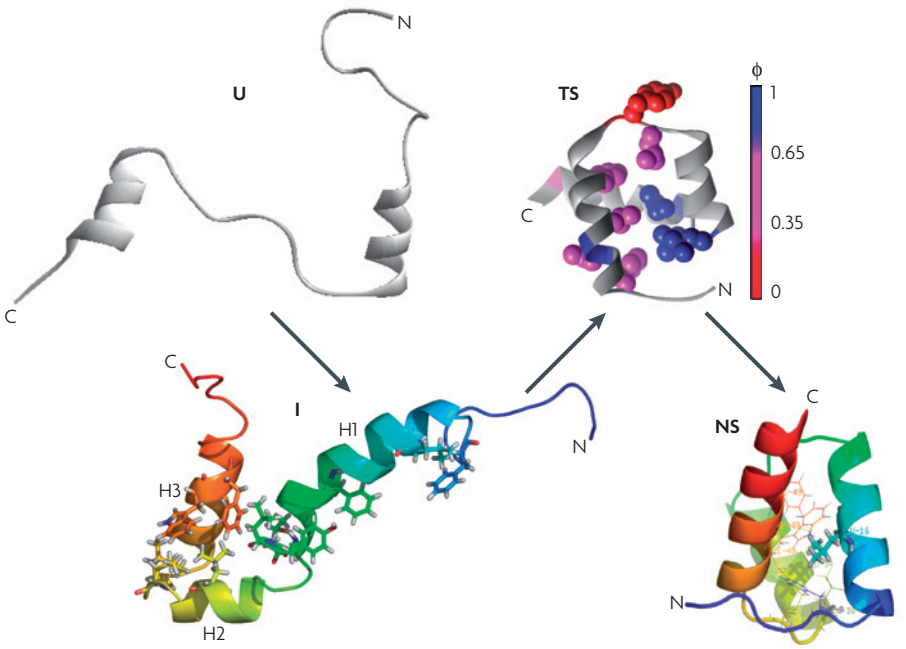


Figure 3 | **The pathway of folding of the Engrailed homeodomain of *Drosophila melanogaster*.** From right to left: the structure of the native state (NS) was solved by nuclear magnetic resonance (NMR) and X-ray crystallography; the transition state (TS) by Φ analysis (colour-coded from red, meaning unstructured, to blue, meaning highly native-like); the folding intermediate (I) was generated as a stable entity using protein engineering and its structure was solved by NMR; the structure of the denatured state (U), under conditions that favour folding, was simulated using molecular dynamics; and the entire unfolding pathway was simulated by molecular dynamics. Nearly 50 years of technological advances were needed to proceed from the structural resolution of Figure 1 to the dynamics and structures in Figure 3. H1, H2 and H3 represent helices 1, 2 and 3, respectively.

the 1950s (REF. 35), is exquisitely exploited by NMR, especially by S. Walter Englander[36], as a residue-specific probe for structural organization in folding intermediates. NMR has, as a result, become indispensable for protein folding studies.

Bioinformatics and structural genomics. In the 1970s, Levitt and Cyrus Chothia showed that proteins could be classified according to their secondary structure[37], which inexorably led to modern bioinformatics studies whereby protein structures are analysed by homology. Such studies require the knowledge of protein sequences. However, establishing one of the finest initiatives of the twentieth century — the determination of the DNA sequence of a human genome[38] — was highly controversial. Here, the arguments centred on 'small' science versus 'big' science, the possible diversion of funds from individual projects and the usefulness of the whole initiative. The huge amount of data that stemmed from the Human Genome Project and the technology that was introduced for rapid sequencing has been invaluable for areas as diverse as bioinformatics and personalized medicine. The two major repositories of the information that has been obtained are the National Center for Biotechnology Information (NCBI), and the European Bioinformatics Institute (EBI; see Further information). The databases are used to look for structural homology and so aid in the prediction of protein structure by analogy, and also provide libraries of local structural elements to be used as analogues or for calculating empirical energy functions.

> "Several protein folding pathways are now known in detail at atomic resolution[29] thanks to the application of a combination of technologies..."

The arguments regarding the Human Genome Project have been repeated for the Protein Structure Initiative. This multinational structural genomics project aims to determine the 3D structures of all proteins by organizing known protein sequences into families, solving the structures of representative examples by X-ray crystallography or NMR spectroscopy and building models of the other proteins by homology. It is hoped that just 2,000 structures are required to be able to model every protein. Some of the arguments against the initiative almost reprise those that opposed the relevance of crystal structures 40–50 years ago, and both sides are worth reading for the discussion of the issues involved[39,40]. Even if 2,000 structures turn out to be insufficient, the structural information will be invaluable.

Current status of protein folding

Despite the huge advances in computational power, it is still not possible, in general, to predict the structures of proteins *de novo*. What has been possible is to harness bioinformatics and use databases of known structural elements to compute the structures of small proteins with high precision and give good models of multidomain proteins. The procedures rely, fundamentally, on the accumulated experimental information that has stemmed from structural biology, using local and global homology modelling and empirical equations that describe the non-covalent interactions between atoms. The progress in prediction methods is assessed every two years (see REF. 41 for the latest assessment).

The pathway of protein unfolding can be calculated by molecular dynamics simulation from the known 3D structure[42]. Folding simulations are more difficult, but have been aided by the discovery of ultra-fast-folding proteins, which fold a million times faster than those that prompted the Levinthal paradox 40 years ago — these proteins fold within a few microseconds on the time scale that is accessible to full atomistic simulation. Such pathways are benchmarked by experimentation, with Φ-value analysis of transition states and NMR spectroscopic structural determination of intermediates and analysis of denatured states. The pathway of folding and unfolding of the Engrailed homeodomain from *Drosophila melanogaster*, for example, has been solved at atomic resolution using this whole gamut of techniques[33,43] (FIG. 3). Simulations using Φ values as constraints can be used to construct transition states[44]. In addition, we can identify from databases structurally homologous proteins of vastly different amino-acid sequence to study the change of folding mechanism with structure and derive more general principles about the mechanism[45]. Several protein folding pathways are now known in detail at atomic resolution[29] thanks to the application of a combination of technologies, all of which initially met with some scepticism.

Challenges ahead

After half a century of structural studies on beautifully folded globular proteins, it is perhaps a shock to discover that up to some 40% of the proteins in the human proteome are estimated to be intrinsically disordered and become fully or partly structured on binding to binding partners in the cell[46]. Analysing the structures of disordered or partly disordered proteins, especially in the context of the cell, is a real challenge and requires a combination of structural approaches[47]. The concept of functional disorder is only slowly percolating through the scientific community. The role of protein instability, misfolding and aggregation, which might lead to novel structures and disease pathogenesis, is a further challenge for therapy[48] and structural studies.

The ability to design a functionally active protein *de novo* is still in its infancy despite 40 years of atomic-level structural studies and 25 years of protein engineering. Our most effective strategy is still to weave themes on what nature has already provided, such as producing humanized antibodies[49], which are now used in the clinic. We can routinely increase the thermostability of proteins, not only for biotechnology applications but also to facilitate structural studies. We can even change specificity, although the design of catalytic function is still extremely difficult. The first enzymes to be designed *de novo* catalyse reactions that require only minimal catalysis[50,51], but this achievement remains very impressive. Protein design and the structural analysis of proteins within the context of the cell are two of the great future challenges. Given the staying power of scientists in the face of difficulty and opposition, we can be optimistic that those challenges will be met.

Alan Fersht is the Herchel Smith Professor of Organic Chemistry at the Department of Chemistry, Lensfield Road, Cambridge, CB2 1EW, UK, and is the Director of the MRC Centre for Protein Engineering, MRC Centre, Hills Road, Cambridge, CB2 0QH, UK.
e-mail: arf25@cam.ac.uk
doi: 10.1038/nrm2446
Published online 25 June 2008

1. Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature* **181**, 662–666 (1958).
2. Perutz, M. F. *et al.* A three-dimensional fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis. *Nature* **185**, 416–422 (1960).
3. Blake, C. C. *et al.* Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Å resolution. *Nature* **206**, 757–761 (1965).
4. Matthews, B. W., Sigler, P. B., Henderson, R. & Blow, D. M. Three-dimensional structure of tosyl-α-chymotrypsin. *Nature* **214**, 652–656 (1967).
5. Wyckoff, H. W. *et al.* The structure of ribonuclease-S at 3.5 Å resolution. *J. Biol. Chem.* **242**, 3984–3988 (1967).
6. Lipscomb, W. N., Hartsuck, J. A., Quiocho, F. A. & Reeke, G. N. Jr. The structure of carboxypeptidase A. IX. The x-ray diffraction results in the light of the chemical sequence. *Proc. Natl Acad. Sci. USA* **64**, 28–35 (1969).
7. Arnone, A. *et al.* A high resolution structure of an inhibitor complex of the extracellular nuclease of *Staphylococcus aureus*. I. Experimental procedures and chain tracing. *J. Biol. Chem.* **246**, 2302–2316 (1971).
8. Perutz, M. F. Stereochemistry of cooperative effects in haemoglobin. *Nature* **228**, 726–739 (1970).
9. Epstein, C. J., Goldberger, R. F. & Anfinsen, C. B. The genetic control of tertiary protein structure. Model systems. *Cold Spring Harb. Symp. Quant. Biol.* **28**, 439–449 (1963).
10. Levinthal, C. in *Mossbauer Spectroscopy in Biological Systems* (eds Debrunner, P. *et al.*) 22–24 (University of Illinois Press, Urbana, Illinois, 1969).
11. Ptitsyn, O. B. Stages in the mechanism of self-organization of protein molecules. *Dokl. Akad. Nauk SSSR* **210**, 1213–1215 (1973) (in Russian).
12. Kim, P. S. & Baldwin, R. L. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* **51**, 459–489 (1982).
13. Karplus, M. & Weaver, D. L. Diffusion-collision model for protein folding. *Biopolymers* **18**, 1421–1437 (1979).
14. Dolgikh, D. A. *et al.* α-Lactalbumin: compact state with fluctuating tertiary structure? *FEBS Lett.* **136**, 311–315 (1981).
15. Creighton, T. E. The two-disulphide intermediates and the folding pathway of reduced pancreatic trypsin inhibitor. *J. Mol. Biol.* **95**, 167–199 (1975).
16. Cohen, S. N., Chang, A. C., Boyer, H. W. & Helling, R. B. Construction of biologically functional bacterial plasmids *in vitro*. *Proc. Natl Acad. Sci. USA* **70**, 3240–3244 (1973).
17. Shortle, D. A genetic system for analysis of staphylococcal nuclease. *Gene* **22**, 181–189 (1983).
18. Perry, L. J., Heyneker, H. L. & Wetzel, R. Non-toxic expression in *Escherichia coli* of a plasmid-encoded gene for phage T4 lysozyme. *Gene* **38**, 259–264 (1985).
19. Hartley, R. W. Barnase and barstar. Expression of its cloned inhibitor permits expression of a cloned ribonuclease. *J. Mol. Biol.* **202**, 913–915 (1988).
20. Sanger, F. *et al.* Nucleotide sequence of bacteriophage φ X174 DNA. *Nature* **265**, 687–695 (1977).
21. Hutchison, C. A., *et al.* Mutagenesis at a specific position in a DNA sequence. *J. Biol. Chem.* **253**, 6551–6560 (1978).
22. Winter, G., Fersht, A. R., Wilkinson, A. J., Zoller, M. & Smith, M. Redesigning enzyme structure by site-directed mutagenesis: tyrosyl tRNA synthetase and ATP binding. *Nature* **299**, 756–758 (1982).
23. Matouschek, A., Kellis, J. T. Jr., Serrano, L. & Fersht, A. R. Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **340**, 122–126 (1989).
24. Fersht, A. R., Leatherbarrow, R. J. & Wells, T. N. Structure–activity relationships in engineered proteins: analysis of use of binding energy by linear free energy relationships. *Biochemistry* **26**, 6030–6038 (1987).
25. Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.* **254**, 260–288 (1995).
26. Daggett, V. & Fersht, A. R. Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* **28**, 18–25 (2003).
27. Levitt, M. & Lifson, S. Refinement of protein conformations using a macromolecular energy minimization procedure. *J. Mol. Biol.* **46**, 269–279 (1969).
28. McCammon, J. A., Gelin, B. R. & Karplus, M. Dynamics of folded proteins. *Nature* **267**, 585–590 (1977).
29. Daggett, V. & Fersht, A. The present view of the mechanism of protein folding. *Nature Rev. Mol. Cell Biol.* **4**, 497–502 (2003).
30. Williamson, M. P., Havel, T. F. & Wuthrich, K. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ^{1}H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.* **182**, 295–315 (1985).
31. Baum, J., Dobson, C. M., Evans, P. A. & Hanley, C. Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig α-lactalbumin. *Biochemistry* **28**, 7–13 (1989).
32. Bax, A. & Grishaev, A. Weak alignment NMR: a hawk-eyed view of biomolecular structure. *Curr. Opin. Struct. Biol.* **15**, 563–570 (2005).
33. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M. & Fersht, A. R. Solution structure of a protein denatured state and folding intermediate. *Nature* **437**, 1053–1056 (2005).
34. Korzhnev, D. M. & Kay, L. E. Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding. *Acc. Chem. Res.* **41**, 442–451 (2008).
35. Hvidt, A. & Linderstrøm-Lang, K. Exchange of hydrogen atoms in insulin with deuterium atoms in aqueous solutions. *Biochim. Biophys. Acta* **14**, 574–575 (1954).
36. Englander, S. W., Mayne, L., Bai, Y. & Sosnick, T. R. Hydrogen exchange: the modern legacy of Linderstrøm-Lang. *Protein Sci.* **6**, 1101–1109 (1997).
37. Levitt, M. & Chothia, C. Structural patterns in globular proteins. *Nature* **261**, 552–558 (1976).
38. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
39. Petsko, G. A. An idea whose time has gone. *Genome Biol.* **8**, 107 (2007).
40. Banci, L. *et al.* An idea whose time has come. *Genome Biol.* **8**, 408 (2007).
41. Moult, J. *et al.* Critical assessment of methods of protein structure prediction — Round VII. *Proteins* **69** (Suppl. 8), 3–9 (2007).
42. Daggett, V. & Levitt, M. Protein unfolding pathways explored through molecular dynamics simulations. *J. Mol. Biol.* **232**, 600–619 (1993).
43. Mayor, U. *et al.* The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* **421**, 863–867 (2003).
44. Vendruscolo, M., Paci, E., Dobson, C. M. & Karplus, M. Three key residues form a critical contact network in a protein folding transition state. *Nature* **409**, 641–645 (2001).
45. Gianni, S. *et al.* Unifying features in protein-folding mechanisms. *Proc. Natl Acad. Sci. USA* **100**, 13286–13291 (2003).
46. Fink, A. L. Natively unfolded proteins. *Curr. Opin. Struct. Biol.* **15**, 35–41 (2005).
47. Wells, M. *et al.* Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl Acad. Sci. USA* **105**, 5762–5767 (2008).
48. Chiti, F. & Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333–366 (2006).
49. Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. & Winter, G. Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature* **321**, 522–525 (1986).
50. Jiang, L. *et al. De novo* computational design of retro-aldol enzymes. *Science* **319**, 1387–1391 (2008).
51. Rothlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).
52. Maxam, A. M. & Gilbert, W. A new method for sequencing DNA. *Proc. Natl Acad. Sci. USA* **74**, 560–564 (1977).

Acknowledgements
I thank the Medical Research Council for 40 years of funding, without which this article would not have been written.

DATABASES
UniProtKB: http://ca.expasy.org/sprot
haemoglobin | hen egg-white lysozyme | myoglobin | nuclease

FURTHER INFORMATION
Alan Fersht's homepage:
http://www.ch.cam.ac.uk/staff/arf.html
European Bioinformatics Institute: http://www.ebi.ac.uk
National Center for Biotechnology Information:
http://www.ncbi.nlm.nih.gov
Protein Structure Initiative:
http://www.structuralgenomics.org

ALL LINKS ARE ACTIVE IN THE ONLINE PDF