Cell. Mol. Life Sci. 63 (2006) 207–219
1420-682X/06/020207-13
DOI 10.1007/s00018-005-5236-7

Cellular and Molecular Life Sciences

Review

Protein flexibility: its role in structure and mechanism revealed by molecular simulations

G. Dodson[a, b] and C. S. Verma[c, *]

[a] Division of Protein Structure, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA (United Kingdom)
[b] York Structural Biology Laboratory, University of York, York YO10 5YW (United Kingdom)
[c] Bioinformatics Institute, 30 Biopolis Way, #07-01 Matrix (Singapore) 138671, Fax: + 65 6478 9047, e-mail: chandra@bii.a-star.edu.sg

Received 29 May 2005; received after revision 23 August 2005; accepted 21 October 2005
Online First 2 January 2006

Abstract. Computer simulations at the atomic level have arrived at a stage where they provide realistic modeling of flexibility in proteins (and the mobility of their associated solvent) that is important in understanding the nature of molecular motions. This can now be extended to the molecular and atomic motions that are associated with protein mechanisms. Moreover, the derived data agree reasonably accurately with experimental measurements of several kinetic and thermodynamic parameters. Fundamental insights emerge on the roles that this intrinsic flexibility plays in the thermodynamic characteristics of macromolecules in solution; these equip the investigator to probe the consequences of cognate interactions and ligand binding on entropy and enthalpy. Thus simulations can now provide a powerful tool for investigating protein mechanisms that complements the existing and the emerging experimental techniques.

Key words. Protein structure-function mechanism; flexibility; atomistic simulation; buried water; lipase; chaperone; prion.

Introduction

The enormous advances in the molecular sciences over the last 30 years have had an immense impact on biological science, indeed they have played a large part in defining it. In particular, the scale and scope of the crystallographic analysis of macromolecules has followed advances in synchrotron radiation technology, in molecular biology and, of course, relevant to this review, in computational methods and computing technology. At the same time nuclear magnetic resonance (NMR) spectroscopy is now making a significant contribution to structure determination and to characterizing protein mobility. In the last 15 years or so, structural research has led to accurate, sometimes very accurate, three-dimensional structures of a wide range of biochemically and biologically active proteins. This expanding avalanche of structural information, all available from the Protein Data Bank [1], includes details of proteins and macromolecular organization, of protein interactions with nucleic acids, other biological molecules and ligands and, again relevant to this review, of conformational behavior. Thus, structural data have provided the essential framework for characterizing mechanisms, for analyzing evolutionary relationships and illuminating the understanding of function. Quite clear now, however, is that the flexibility in proteins, orchestrated with the chemical requirements at each stage of their reactions, is a major component in creating their specificity and catalytic power. Indeed, a very useful database of macromolecular movements is now available

* Corresponding author.

that demonstrates nicely the extent of the inherent flexibility of proteins [2] (http://www.molmovdb.org).

X-ray analysis has laid crucial foundations in modern biology, but while it provides an image of unparalleled detail, this is a static image, the molecules are trapped in a lattice. If the analysis is carried out at a reasonably high resolution, 2.5 Å or better, then the crystal structure does give usefully accurate thermal, or vibrational parameters. But owing to the crystal contacts, these data do not necessarily allow one to predict the dynamic character of macromolecules. For example, the presence of substantial >1 Å breathing motions in proteins has been recognized since Wuthrich's observations on the flipping of the buried aromatic residues in pancreatic trypsin inhibitor [3]. The breathing motions required to accommodate the flipping are not deducible from the X-ray crystallographic diffraction data. Nonetheless, there have been calculations in which the crystallographic thermal parameters have been used to derive overall molecular motions. Those of Phillips in 1979 on lysozyme [4] and those of Karplus and colleagues [5] on several proteins address the potential for extrapolating from thermal parameters to intrinsic molecular motions.

Most importantly, Artymiuk et al. [6] identified distinctly higher atomic mobility at the lysozyme active site and argued this reflected a functional property. Moreover, they deduced correctly from other experimental data that it was possible to allocate a component in the amplitudes of atomic vibration that derived from the overall global motion of the protein. Incoherent scattering is another source of information about molecular motions. This is rather often detected in proteins and there have been some careful and detailed analyses of the phenomenon, such as that by Caspar and colleagues on the insulin hexamer [7]. However, the general difficulty in extracting the appropriate atomic motions in macromolecules from X-ray diffraction data is very much more serious when one considers structural behavior in protein mechanisms – which is the subject of this review. For example, in the case of insulin in the hexameric organization, the N-terminal B chain residues assume two conformations: extended and helical [8]. This conformational flexibility is possibly exploited in receptor binding but this structural capacity is not evident from the thermal parameters in a number of structures in which these different N-terminal B chain conformations occur. It is important to recognize this limitation in crystallographic analysis.

The nature of the motions in protein and macromolecular mechanisms clearly depends on local, and probably global, fluctuations, but these generally appear to be modulated by interactions between the protein concerned and its ligands or cognate partners. Solution structures derived from NMR [9] can reveal some aspects of protein flexibility but rather special techniques are required to characterize their specific features and, as with crystallography, the NMR approach faces limiting technical problems in determining the structures or dynamic equilibria associated with interactions and chemical events [10]. In addition there are fundamental molecular properties that are significantly affected by the flexibilities in the macromolecules under consideration, such as the structures of transition states, ligand affinities and the effects of mutations on reactivity, to say nothing of the complexities in thermodynamic parameters and the important influence of solvation. However, encouragingly, recent advances in optical spectroscopy, particularly time-resolved spectroscopy [11, 12], promise fundamental benefits for the applications of simulation methods. Another technique, FRET, is now allowing intramolecular interactions to be identified [13]. When these interactions are generated by conformational change then, of course, one has a measure of the real-time dynamics, but not of the structural changes themselves. Such observations, nonetheless, are going to be invaluable, since they place a timescale on the structural changes and the associated mechanisms.

Over the last 10 years, a major impetus to understanding the links between flexibility and function has come from the applications of computer 'experiments,' based on molecular dynamics (MD) simulations. Modern MD simulations have roots in the seminal work of Alder and Wainright [14], Lifson and Warshel [15] and Rahman and Stillinger [16]. However applications to biological systems were essentially pioneered (and continue to be so) in the work of Martin Karplus and colleagues [5, 17, 18]. A combination of improvements in computer power, and in the representations of the parameters that define the physicochemical properties of biological systems, has meant that detailed simulations of thousands of atoms involved in various complex physical and chemical reactions are now routinely carried out. These yield details of atomic motions, flexibility, titration characteristics, effects of mutations and pathways of conformational changes on the one hand, and complex chemical processes on the other, particularly when combined with quantum mechanical methods. One also notes an encouraging close agreement between the root-mean-squared fluctuations and motions from simulations and those observed in high-resolution static and time-resolved rapid crystallographic experiments; this validates the individual MD trajectories in a particularly convincing manner [19, 20]. Several reviews [21–26] have covered various aspects of the field, including one extensive review that has been provided by Norberg and Nilsson [27]. Regular updates on advances in methodology are found in the journal *Current Opinions in Structural Biology* (for example the April 2005 issue). On the future of this technology, Norberg and Nilsson state [27]: "On the horizon one may find MD simulations covering the dynamics of viruses, the protein-folding process, and various large

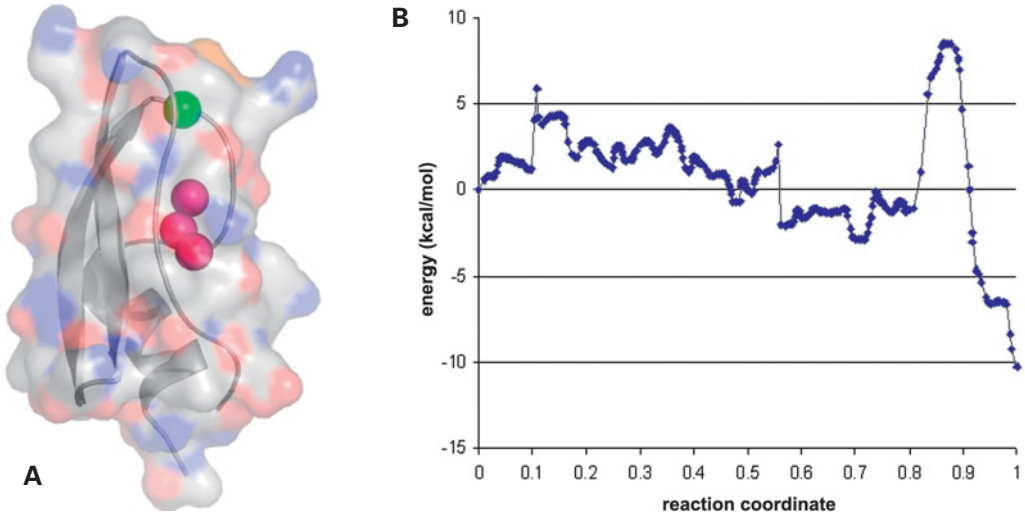

Figure 1 (*A*) The secondary structure of bovine pancreatic trypsin inhibitor (BPTI) with the molecular surface and the locations of four buried water molecules shown as spheres. The most buried water is shown in green. (*B*) The energy landscape associated with one of the several pathways available for a water molecule to enter an enclosed cavity in the protein BPTI.

protein complexes making up the molecular machines that we now begin to perceive in the 'biological cell'." The following review focuses on local structural movements in proteins derived by molecular simulations. These are associated with the mechanism of action of the protein concerned and we examine them in the context of the global oscillations. The proteins chosen here were selected to illustrate the character of functional motions, in particular their dependence on water or particular amino acid residues or the presence of ligand.

Protein flexibility and buried water molecules

The first recognition that protein molecules were intrinsically mobile and that these motions had coherence came from the NMR observations reported by Wuthrich and colleagues in 1976 [28]. The concept of inherent and correlated motions was a landmark in protein physics and structural biology. These ideas and the associated recent developments by Halle and colleagues [29] now also underlie our understanding of buried water and ligand binding in proteins. Buried waters, which occur in most proteins, are generally seen to be serving a structural role in filling space and in completing H bond potentials when the requirements of the protein fold lead to some inefficient packing of the non-polar core. Usually, these buried water molecules are in a locally polar, or mostly so, environment, and because protein interiors contain mostly non-polar side chains, their H bonds generally include mainchain NH and CO atoms. The binding and departure of buried water can tell us a great deal about protein flexibility, possibly more than ligand exchange, since the extent of the protein atom fluctuations is usually larger, and the necessity to preserve chemically and thermodynamically appropriate pathways is more demanding.

The long-known example of a buried water molecule in the small protein bovine pancreatic trypsin inhibitor (BPTI) has proved to be a key system conceptually and experimentally. This water molecule is completely hidden from the surface (fig. 1A) but Halle and colleagues [29], using refined NMR methods, showed that this water does exchange at a frequency of microseconds. Exchange with surface molecules at the microsecond timescale even with a degree of flexibility in the protein looked unlikely in the extreme, since this property implied large and concerted fluctuations in the host protein. These fluctuations were, however, detected by molecular calculations [30, 31]. The calculations revealed first that the water while inside the molecule underwent librational motions of about 10 degrees in which its H bonding contacts were maintained, but stretched; these motions were, however, punctuated by the occasional C2 flip, through which the water H bond partners switched at a microsecond frequency (see www.bii.a-star.edu.sg/~chandra/CMLS-Movie1). The pathway(s) by which the buried water escapes from and arrives at the buried site, and the accompanying protein motions, have been modeled [30]. Striking in these calculations was that as the water migrated from its internal site (see www.bii.a-star.edu.sg/~chandra/CMLS-Movie2), its H bond forming oxygen, hydrogens and the empty orbitals formed successive branched interactions, most reminiscent of Tarzan (water) swinging his way with hands and feet through the jungle (protein) on successive vines (H bonds).

There is another important aspect to these calculations: the energy landscape associated with these linked transient H bonding structures has low values (fig. 1B). The activation energy required to remove the water from the

BPTI interior with little internal flexibility is about 25 kcal/mol; in contrast, the maximum value of the profile generated by including concerted protein motions is 13–15 kcal/mol, while the individual maxima associated with each structural step are 1–2 kcal/mol. Thus the computations reveal both a structural mechanism of the water escaping from and returning to its buried site and an energy barrier profile consistent with the observed timescale of exchange and the thermodynamic parameters [29].

The studies on BPTI extend to the analysis of the thermodynamic parameters of ligand binding. For many years, it has been widely assumed, for intuitive reasons, that ligand binding generally increased the order in the ligand and host molecules, leading to a drop in the entropy of the system and a corresponding increase in the entropy of the solvent displaced by binding. Crystallography provided support for this concept, since in many cases, the complexes were better ordered in the crystal than in the free protein. The issue of the entropic state of a protein on ligand binding was raised by Dunitz [32]. A series of detailed calculations [33–35], however, showed that the effect of ligand binding (in this case a buried water) is to decrease the entropy locally, which we would expect, but that the low-order modes of vibration (those associated with large-scale fluctuations) of the protein increased significantly. This behavior increases the total entropy associated with binding. Given that ligand binding lies at the heart of biochemical and biological processes, it is vital that there is a secure comprehension of the physicochemical basis for ligand-host interactions that includes the dynamics of the system.

The analysis of the BPTI structure was further refined by calculations on BPTI mutants. One of these, Gly36Ser, in which the Ser side chain replaces the bound water in the cavity, has been shown to reduce inhibitor activity [36]. Calculations [31] suggest that the reduced activity of the mutant is a consequence of reduced fluctuations and flexibility in the molecule, the latter leading to a significant reduction in the entropy. In particular, there is a correlation between the number of hydrogen bonds formed within the cavity and binding affinity, a reduction in hydrogen bond numbers leading to a reduction in activity. The obvious conclusion is that the loss of hydrogen bonding in the cavity is modifying the overall flexibility of the molecule away from that which has evolved for optimal binding.

Solvent structure in large cavities

A number of larger cavities with room for more than one water occur in proteins, and these sometimes exhibit no well-defined electron density. The most likely explanation for this phenomenon has always been that the water molecules in the cavity are very mobile and populate several sites – thus their electron density is smeared out. One buried volume of particular interest occurs in interleukin-1β; it is considered here because its internal surface is non-polar and thus H bonding can exist only between buried water molecules. In the X-ray analysis, there is no evidence for well-defined electron density that would identify stable water molecules in the cavity [37]. An NMR study, by contrast, identified the presence of water molecules in this cavity but was unable to characterize their number, or the extent, of their mobility [38]. A later very detailed crystallographic study by Yu et al. [39], in which all the low-order diffraction amplitudes were measured, showed, however, convincing evidence for poorly defined, partially occupied water positions.

Computations [S. Somani, C.-P. Choon, C. S. Verma, unpublished data] carried out on this cavity populated by up to four water molecules revealed that water clusters in varying formations can exist in this environment for periods exceeding 1 ns (fig. 2), the lower limit for detectable nuclear overhauser effect (NOEs) [38]. Moreover, consistent with their not being detected by the initial X-ray analysis, the water molecules undergo large-scale fluctuations with no positions being highly populated. The complementary X-ray crystallographic, NMR and simulation studies on interleukin provide a convincing demonstration of how the water molecules behave in the limited volume. The matches between the low-resolution electron density, the computation of the most populated states and the residence times (>1 ns) are very promising (fig. 3). The simulations suggest that there are more entry-exit pathways available to the water than suggested by the X-ray analysis. This is not surprising, since the crystal lat-

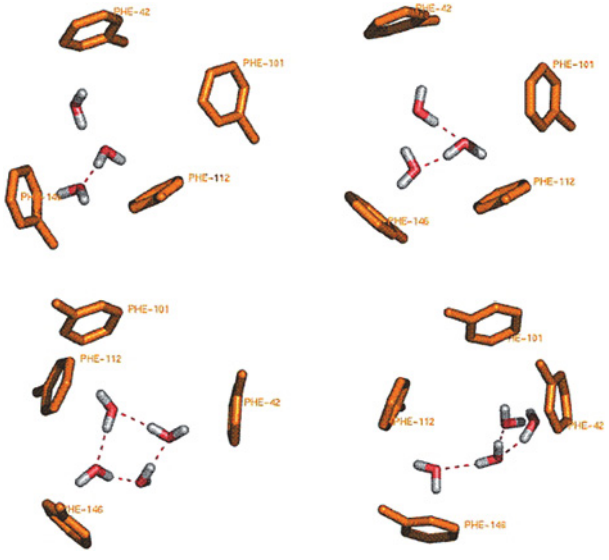

Figure 2. Patterns of hydrogen-bonded clusters of water molecules that are observed in the hydrophobic cavity of interleukin-1β as seen in MD simulations.

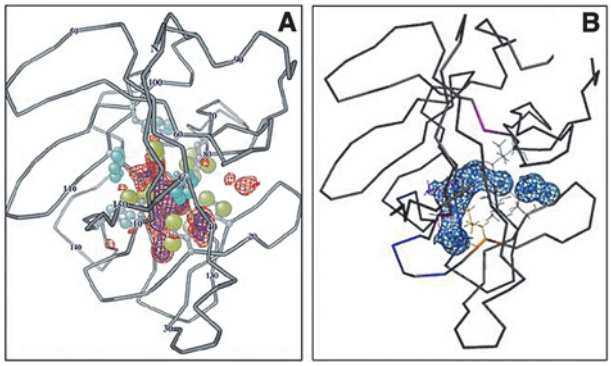

Figure 3. Similarity in the experimentally derived and the computed profiles for the distribution of water molecules in the hydrophobic cavity of human interleukin-(IL)1β. (*A*) The experimental difference electron density map distribution within the cavity region of the human IL-1 molecule within a sphere of 10-Å radius is displayed inside the C trace of the human IL-1 molecule viewed from the side containing the C terminus (center) and N terminus (top), with ball-and-stick representations of the cavity-forming residues. The methyl groups of the aliphatic residues are drawn as large tan spheres, and the aromatic carbons are smaller, light-blue spheres. Every tenth C atom is labeled, except 120, which is obscured behind the cavity density map. The red contours contain 70% of the total of 18 solvent electrons integrated in the cavity region, and the blue contours contain 50% of these electrons. Figure taken from Yu et al. [39] with permission; copyright (1999) National Academy of Sciences, USA. (*B*) The figure illustrates the computed density of waters in the cavity derived from the MD simulation. The region in blue is the binding loop that interacts with the receptor. Regions in yellow and purple are sites of mutations that affect receptor binding

tice will introduce elements of rigidity that are likely to inhibit diffusion of water molecules.

The buried waters in interleukin are apparently also playing a critical role in relaying the intrinsic modes of vibrations across the molecule. Mutations that are 8–20 Å from the functional loop that binds to its receptor are known to interfere with binding [40]. These mutations may alter the character of the global motions of the molecule. But the key observation is that the simulations show that in the absence of water in the cavity, there are no correlated motions that couple the sites of mutations and the binding loop. This observation highlights the complex and subtle requirements needed by proteins for their natural vibrational behavior, and it is not surprising that such properties can be incorporated into the functional behavior of proteins.

Local flexibility and its role in mechanism in selected proteins

The intrinsic flexibility of the polypeptide chain is disciplined in proteins by its organization into secondary, tertiary and where this occurs, quaternary, structures. Polypeptide chains not in secondary-structure elements are, however, often flexible and such structural variability has been exploited in a number of regulatory systems in which the protein is activated by a structural element changing conformation, often after a chemical modification. We now examine a selected set of studies where simulations reveal loop conformational behavior and have provided invaluable insights into experimental observations.

Liganded calmodulin

Calmodulin is a protein involved in the regulation of muscle and a number of other mechanisms. It consists of an N- and C-terminal domain; each homologous and each made up of two smaller, and again homologous, calcium-binding domains. The domains consist of well-defined secondary-structure elements linked by peptide segments. In the presence of calcium ions, the domains form a tighter structure that facilitates binding to the cognate peptide ligands of the protein with very high affinity and activating the system concerned. Calmodulin thus acts as a molecular switch, controlled by calcium levels. The molecule is remarkable in that it binds many peptide fragments with very different sequences and with very high affinity [41, 42]. Exactly how the same sequence can adapt to the varying sequences of the substrate peptides is an interesting and important question. On binding to calmodulin, the cognate peptide usually develops into a helix about which the four component calcium-binding domains are wrapped. The different sequences of the bound peptide segments generate helices with distinctly variable surface properties. Comparison of different calmodulin-peptide complexes shows that the individual helix of the bound ligand and the individual helices in camodulin preserve their structure. For example, in the complexes between the target peptide in skeletal muscle, myosin light-chain kinase and the phosphofructose kinase, the Calpha atom positions in the helical structures agree to within ca 0.3 Å. [P. Bayley, unpublished data]. The completely different helical ligand surfaces are partly accommodated in the complex by movements in the very flexible loops of calmodulin between the N&C domain pairs, in the short links between the individual Ca-binding helical segments within the N&C domains and in the Ca-binding motifs. These flexible segments position the four domains and their helices appropriately, allowing fine-tuned adjustments by the individual side chains on the host and substrate. The nature of the interactions in different complexes has been investigated both by mutations and by some computational analysis [43]. The distinct character of the two calmodulin domains is reflected in different calcium and target interaction affinities, with the C-terminal domain generally showing the higher affinities. From this sample of conformations, multiple normal-mode analyses revealed the vibrational and thermodynamic properties in the molecule. We see

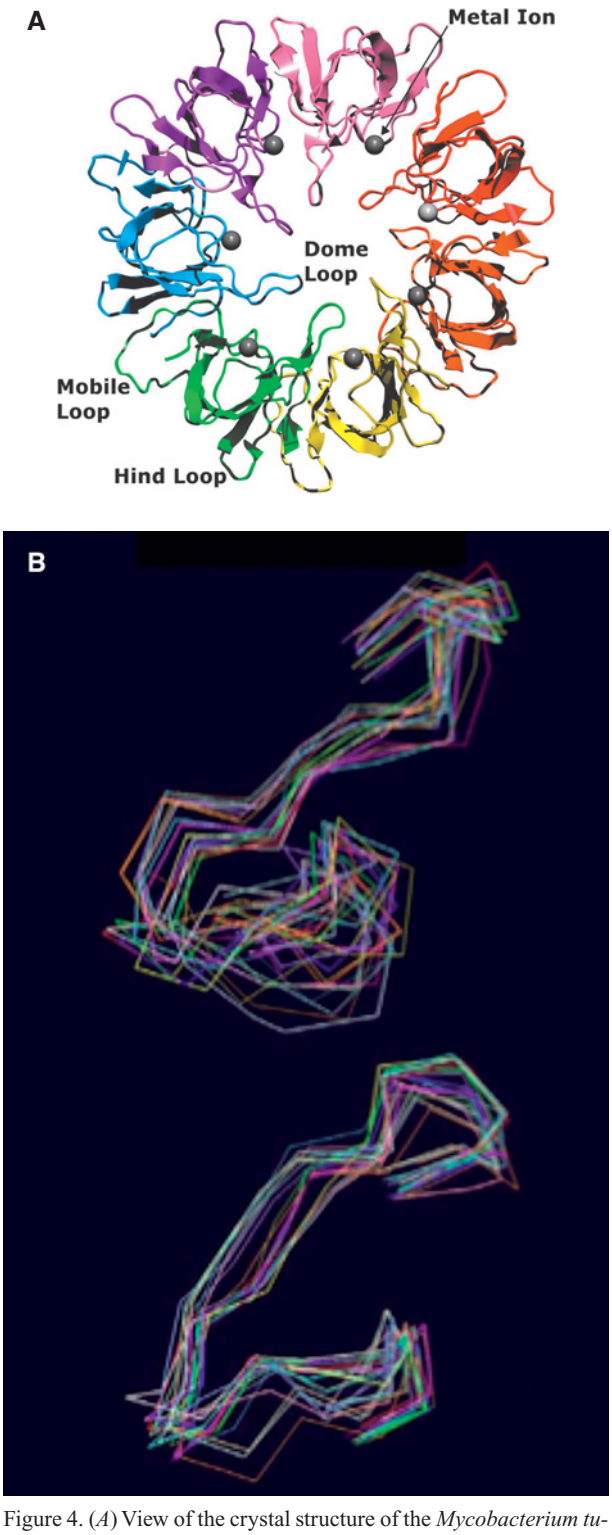

Figure 4. (*A*) View of the crystal structure of the *Mycobacterium tuberculosis* Cpn-10 with metal ions. The ribbon diagram was drawn with MOLSCRIPT. Each subunit is shown in a different color and the metal ions are colored in white. The loop regions have been marked accordingly. (*B*) A detailed picture of the various conformations sampled by the mobile loop of subunit A during the last 1.8 ns of both MD simulations. From the crystal structure taken as the starting conformation, snapshots of the mobile loop were taken every 100 ps and superimposed. *A*, in the presence of calcium ions; *B*, in the absence of calcium ions.

higher intrinsic flexibility of the C domain compared with that of the N domain. Furthermore, in a simulation of a calmodulin-target peptide complex, the C domain conformation maintains its conformation better and has lower atomic root mean squared (RMS) fluctuations than the N domain. These results tie in with the observed differentiation of roles of the calmodulin domains [44]. The system does present computational difficulties partly because of the large conformational changes that occur upon calcium binding and to a further degree upon peptide binding. However, there are extensive NMR data on the system and these, and some thermodynamic parameters, have been reasonably matched by computations [42, 45]. This is one system where research in the future is undoubtedly going to be productive.

Chaperonin-10

Chaperonins are an important class of proteins which help manage the folding process in certain systems. Conformational changes will obviously be involved in their mechanisms and they have been extensively investigated by biochemical, genetic, structural and computational experiments [46, 47]. Recently, some structural and computational studies have been carried out on GroES alone [48]. This ubiquitous heptameric protein, also referred to as Chaperonin-10, exploits metal ion binding, an event which modulates the flexibility and specificity in its functionally important binding loop. X-ray crystallographic analysis of the Chaperonin-10 assembly shows that the segment, identified as a GroEL binding determinant loop, is mobile [49]. Small-angle scattering suggests that the perimeter of the dome, and the orifice connected to this, are highly plastic [50]. Fluorescence spectroscopy experiments demonstrated that calcium ions interact with the molecule bringing about a reversible change in the Trp environment [51].

These structural results stimulated detailed MD simulations in which the affects of calcium ions on the atomic parameters, derived from the crystal structure, were explored [48]. The calculations showed that the binding of calcium ions led to a widening of the central orifice, an increase in mobility in the GroEL-binding loop, and the generation of concerted motions in the seven subunits. The simulations also suggest interplay between two distantly spaced loops, the metal-binding 'dome loop' and the GroEL-binding 'mobile loop,' implying an important cation-mediated role in the recognition of Chaperonin-60 (fig. 4A). In the presence of cations, the mobile loop appears poised ready to dock onto the Chaperonin-60 structure (fig. 4B). Analysis of correlated motions reveal that in the presence of calcium ions, there is a significant increase in correlated motion, in the heptamer as well as in each monomer.

The enhanced flexibility in the presence of ions has been postulated elsewhere [52] and could well turn out to be a

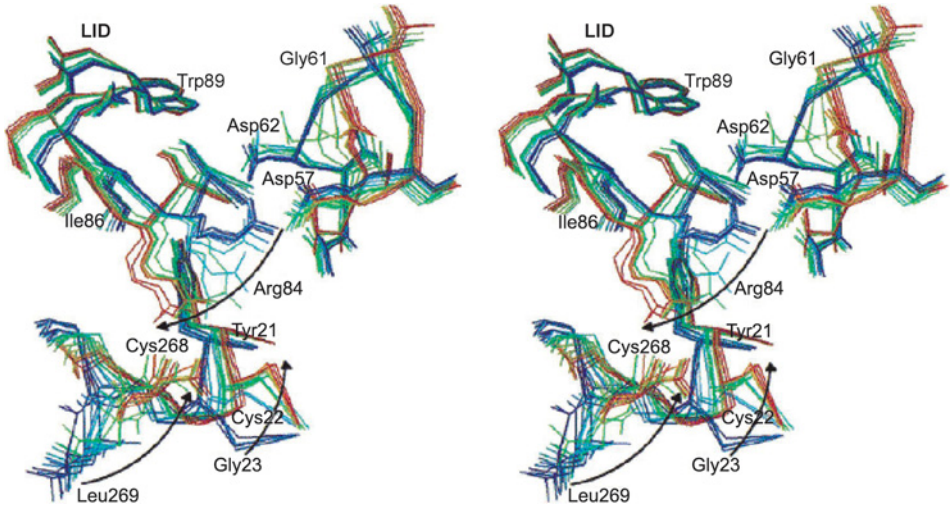

Figure 5. Pathway of conformational change in the *Thermomyces* (*Humicola*) *lanuginosa* lipase showing the hinge at Arg84 associated with the lid helix movement. The figure shows 19 computed states that occur between the active and inactive forms. Conformational states are shown with the order of events from dark blue (first three structures) to dark red (last three structures), and by the direction of the arrows; key residues are labeled (divergent stereo). Taken from Brzozwski et al. [55] with permission.

general mechanism to assist a disorder-to-order transition by creating a large population of structures from which the binding structure is selected, or in the thermodynamic view, by lowering the barriers for this conformational change. Increased correlated motions in the presence of metal suggests that Chaperonin-10 may exploit breathing modes to bind to Chaperonin-60 and to carry out its folding and allosteric functions efficiently. The tight coupling between loops also suggests that metal ions may assist in the transmission of signals across the molecule via the global oscillations.

Lipases

Lipases are enzymes that cleave triglycerides; their reactivity depends crucially on conformational flexibility which is governed by the nature of the solvent conditions. The enzymes are generally inactive in aqueous conditions and are activated by an organic phase, a process referred to as interfacial activation [53]. Crystallographic analysis of the lipase from *Thermomyces lanuginosa* has shown that there is a helix segment that covers its catalytic site in polar conditions (closed state), and that this is displaced on interfacial activation, revealing the catalytic residues to the glyceride substrate (open state) [54]. X-ray analysis of the native enzyme in the presence of increasing levels of lipid showed that there were three well-defined positions for the protecting helix, a closed inactive state, a partially closed low-activity state and an open fully active state [55]. This suggests that, in solution, there is a dynamic motion of this helix, covering the active site, between at least three states identified in the crystallographic studies (fig. 5). The simulations showed that the three states were separated by low (ca 11 kcal/mol) energy barriers and that exchange occurred at the microsecond timescale.

The plasticity in the lipase specificity pocket plays an important part in the recognition specificity and products of the enzyme. Because lipases are able to carry out very difficult chemistry, they are often used in synthesis; for example, in the monoacylation of sucrose, a regioselective reaction. Depending on conditions, lipases produce different positional isomers. The lipase from *T. lanuginosa* displays a high regioselectivity for the 6-hydroxyl position in the acylation of sucrose [56] while that from *Candida antarctica* B yields a mixture of two monoesters (the 6-hydroxyl and the 6′-hydroxyl position) [57]. Investigation into general lipase motions in the *C. antarctica B* and *T. lanuginosa* species identifies modes of molecular vibrations that can be associated with the enzyme specificites. These are linked to the movements that protect and expose the catalytic and binding site [58].

To understand the observed regioselective transesterification, computational analyses of putative transition states in the reaction of sucrose with vinyl laurate catalyzed by lipases from *C. antarctica B* and *T. lanuginosa* have been carried out (fig. 6). The binding pocket of lipase from *Candida* is large and plastic, and it adapts to the conformational variability of the ligand substrates by small, rapid local adjustments. This leads to the identification of one large pocket in *Candida* that accommodates both the sucrose and the lauroyl moieties of the transition state leading to a broad specificity and mixed products. In contrast, the smaller pocket of the *Thermomyces* lipase is divided into two volumes, one polar and the other nonpolar. There are no significant local motions detected in

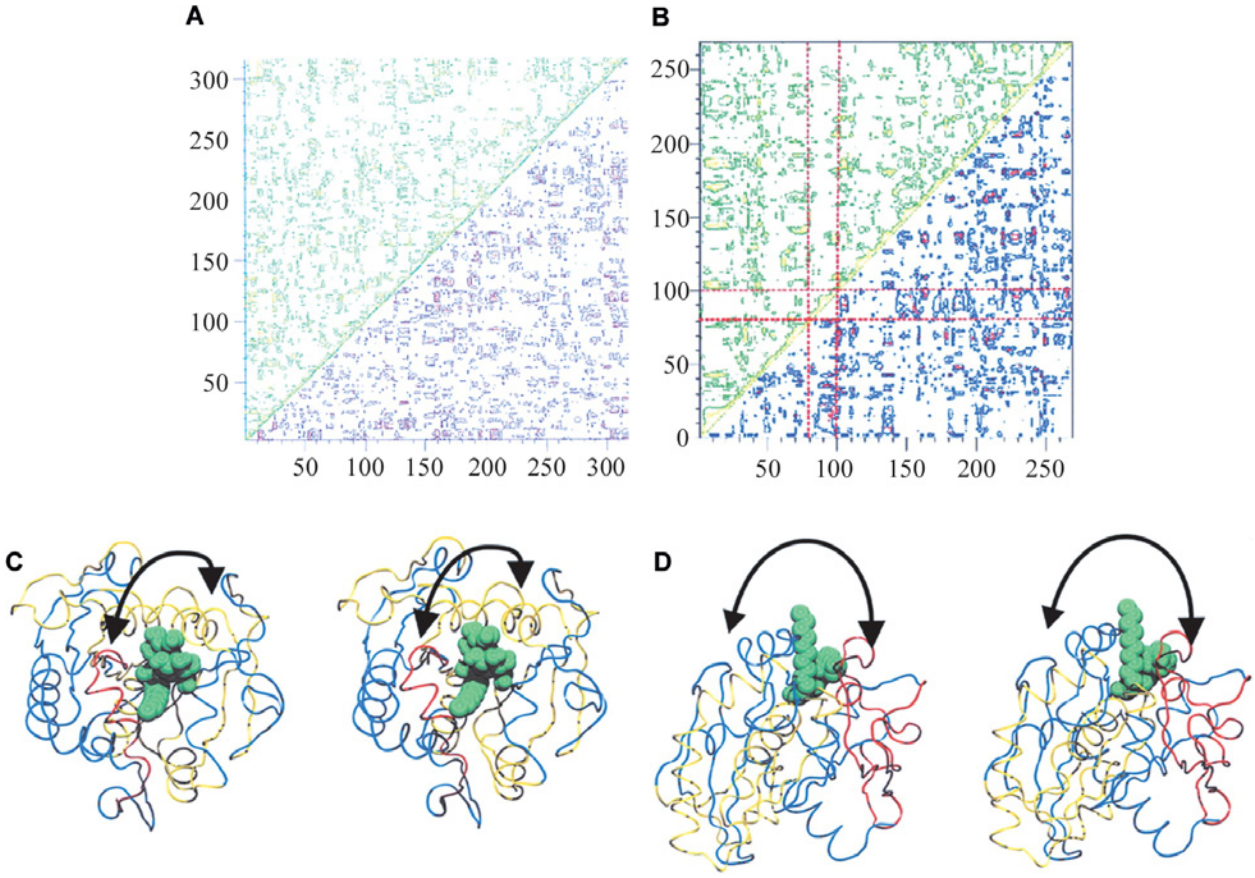

Figure 6. Contour plots of the correlations between distributions of the Calpha positions in lipases. Positive values are shown above the diagonal and negative values are shown below the diagonal. The contours correspond to correlations of –0.6, –0.3, 0.3 and 0.6 in lipases from *C. antarctica B* (*A*) and *T. lanuginosa* (*B*); the dashed lines show the location of the lid region. (*C*) Stereoview of the Calpha trace of *T lanuginosa* in blue ribbon with ligands shown in green spheres; the lid region in red is anti-correlated in motions with the regions in yellow and the black arrows depicts this motion. (*D*) Stereoview of the Calpha-trace of *T. lanuginosa* in blue ribbon with ligands shown in green spheres; the 200–250 subdomain region in red is anti-correlated in motion with the regions in yellow, and the black arrow depicts this motion. *C*, *D* are viewed in two roughly perpendicular directions. Figure taken from Fuentes et al. [58] with permission.

these two pockets which impose more specific constraints on the ligand. Thus the different capacity to adapt to varying conformations of the ligands in the two enzymes occurs as a result of the specificity pocket being a single mobile cavity in one case and a split but more rigid cavity in the other. The relatively well-defined structure of the two pockets in the enzyme from *T. lanuginosa* partly rationalizes its tighter specificity.

Prions

Prions have very unusual properties that present both experimental and conceptual problems to protein science [59–61]. The protein is abundant and thought to be involved in diurnal rhythms [62] as well as, more recently, as a genetic element [63], thereby challenging the long-held central dogma of molecular biology [64] '…all genetic information is stored and transferred digitally through DNA – the only possible explanation of how life evolved. Indeed, it seems that we see growing evidence that information can be transferred genetically in an analogous way through the prion'. Prions have a strong tendency to form amyloid fibers in which the native structured prion, PrPc, is altered in conformation, evidently containing more beta strand and sheet structure and probably in part unfolded, to form the so-called PrPSc. PrPSc is associated with transmissible spongiform encephalopathies, an infectious condition in which the patient suffers major lesions in the brain leading to incapacity and inevitable death. The infectious mechanism is not understood but it is evidently based on the molecular properties of the prion in which PrPSc interacts with PrPc and generates more PrPSc until cell damage results. These events represent a new disease mechanism and present protein science with a fascinating and fundamental problem. Attempts to investigate the phenomenon by experimental methods have however been frustrated by the so-far intractable difficulties associated

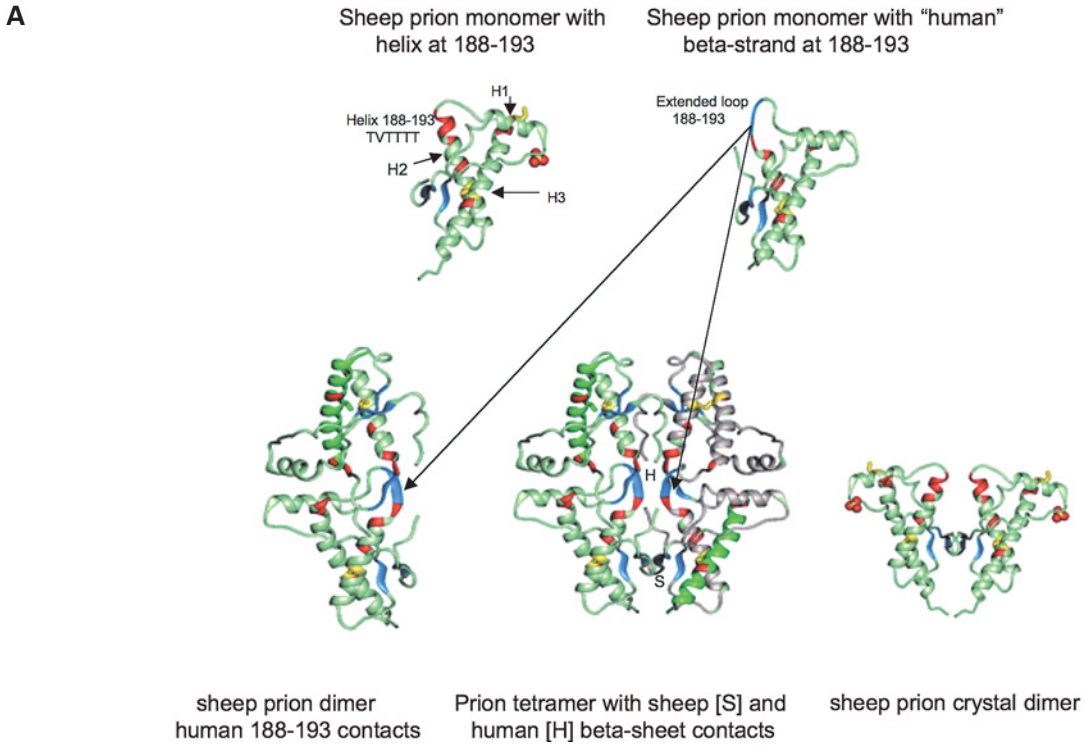

Figure 7. (*A*) A possible model for the development of aggregates in prion proteins using the crystal structures of sheep and human prions. (*B*) Simulated conversion of Syrian hamster D147N PrPC to PrPSc at low pH. *a*, the wild-type NMR structure is shown (left) with the helices and strands labeled. A representative PrPSc-like structure (8-ns snapshot) is shown (right). *b*, hydrogen-bonding network of the PrPC strands S1 and S2 and the in silico PrPSc sheet E1–E3. Taken from De Marco and Dagget [71] with permission.

with a complex pattern of oligomeric intermediates and insoluble fibers [65, 66].

The crystal structures of the sheep and human C-terminal prion domain have revealed first that the molecule is able to undergo a spectacular helical rearrangement that includes both reduction and oxidation of buried disulfide bonds [67, 68]. This behavior illustrates just how extensive the conformational properties of proteins can be – this flexibility in the prion is likely a factor in its enhanced ability to form fibers. Secondly, in both dimers there are anti-parallel beta sheet interactions, in the sheep construct at 129–131 and 161–163, and in the human at 189–193 and its twofold equivalent in the other monomer. The segment 189–193 contains four Thr residues. In the sheep prion C-terminal domain, the crystal structure contains further beta strand interactions at 129–131, forming an intramolecular anti-parallel beta sheet with 161–163. Its extension in the crystal lattice to a four-stranded sheet around the crystal twofold axis is an indication of a possible pathway toward an amyloid structure.

In the human prion C-terminal dimer, helix 3 swaps with its partner with a concomitant reduction and oxidation of

the disulfide bond 179–214. The beta strand 189–193 develops from a helix and this indicates that there is a potential in this particular sequence for beta strand transition. Figure 7A illustrates the sheep prion monomer with the conformation seen in the crystal and the sheep prion monomer with the conformation modeled from the human dimer. This modified monomer has a beta strand between 189 and 193 instead of a helix. The sequence in this region is most unusual consisting of TVTTTT. In the simulations, between the two conformations, it could be seen that the Thr OG was satisfying the hydrogen-bonding potential of the main chain as it switched from helix to strand. This property reduced the activation energy for the transition and the question now arises as to whether this behavior is a factor in the instability at the C-terminal residues of helix 2 identified by Dima and Thriumalai [69]. These calculations are based on an analysis of helical sequences and charge in the PDB which suggested that there were unusual characteristics in the prion helices, especially 2 and 3. By concentrating their simulations on segments of helical instability, Thriumalai and his colleagues propose that helix 1 is not involved but that the second half of helix 2, and some of helix 3 lose helicity and form either beta strands or random coil. This finding is consistent with the development of beta strand structure at 188–193 in the human prion dimer observed in the crystals and does suggest that this region of the prion is plastic [68]. This encouraging picture is however weakened by the findings that an Fab, whose epitope is at the helix 2 C terminus 190–193, binds to both PrPc and PrPSc, suggesting that this element of structure does not undergo conformational change in amyloid formation [71]. This conflict is a reminder of the primitive state of prion fiber research.

The experimental frustrations and the advances in computational power and methods have encouraged MD simulations not just of isolated conformational events but also of the PrPc→PrPSc conversion. Even though the outcomes of such calculations are unlikely to be correct in any detail, they have real value in identifying conformational potential in the molecule which can aid our understanding of the phenomenon of amyloid fiber formation. The simulations by DeMarco and Daggett on PrPc→PrPSc have generated extra beta strands (fig. 7B), especially at low pH, and suggest possible interactions that lead to the development of amyloid structure [71]. The authors have built a model for the fiber which they have mapped on to electromicroscope images and in which the sugar-binding sites have been successfully included, which is encouraging. However, this model differs from other proposals [69], an obvious indication of present inadequacies in treating (un)folding processes.

Evidence that solvent-protein interactions are critical in fiber formation [67] has prompted calculations on the behavior of solvent at prion surfaces. Fernandez's analysis of the prion PrPc H bonding pattern [72] reveals that the prion C-terminal domain has a unusually low ratio of protected and unprotected main chain H bonds. These regions turn out to be populated by polar residues with the capacity to H bond to water molecules and, one presumes, stabilize them locally to some degree. Comparison with D-exchange by NMR methods showed an excellent correlation with the predicted unprotected main chain H bonds which encouraged a full MD simulation of the pattern of solvation at the surface [73]. These calculations reveal that at the unprotected sites, the water is extremely mobile, essentially like bulk water. Thus the side chain polar groups are not stabilizing water molecules but serving to create a highly dynamic population which, in the event of local fluctuations of the main chain H bonds, will be well placed to invade the protein and favor unfolding and fiber formation. Characterization of the solvent-protein interface is likely to be a critically important factor in understanding stability, conformational change, denaturation and the formation of fibers.

Discussion

There is obvious scope for more research in developing techniques. In addition to the analyses we have described, there have been other very penetrating studies on systems such as myoglobin, alpha-lytic protease, cysteine proteases and dihydrofolate reductase, all of which demonstrate the important connections between molecular motions and reactivity [74].

The rate of progress in molecular calculations is making feasible analyses of larger and more complex dynamic systems, for example the ribosome [75]. By contrast, another and very different example where insight is still needed is the nature of the patterns of interactions made by water and counterions at the protein surfaces. These interactions are important if we are to understand the nature of the structural, chemical and thermodynamic factors that govern protein stability, leading from this to fiber formation, one of the most important unresolved issues in protein science.

The knowledge of the important role particular amino acid side chains can play in molecular motion provides a wider reference for the involvement of amino acids in protein structure and function. Thus global modes of vibration can provide a very satisfactory explanation for how mutations distant from active sites can affect activity [76]. Consequently, we should emphasize that selection pressure operates just as much on molecular motions of amino acid side chains as on amino acid chemistry, and one of the beauties of molecular calculations is that effects of mutations on protein function, both positive and negative, can be so effectively explored.

These patterns of molecular flexibility are clearly intrinsic to proteins and, as Miller and Agard [77] say, "have two possible implications. The first pertains to protein engineering, and is the idea that desired patterns of flexibility might be 'designed into' a protein by specific mutation, as a means of either selecting or regulating a given function. The second pertains to rational drug design. If dynamics are important for enzyme function, and if the protein regions most responsible for the motion can be identified, they might present a potential target for small-molecule drug binding and inhibition." Finally, it must be remembered that calculations on protein motions are still in an early stage and better approximations of the complexities in macromolecular systems together with further advances in computer power are needed. We also need much more refined analyses of the effects of mutations both near and distant to the active sites on reaction rates and specificities.

Conclusion

The most important development in the field is that the validity of simulated models of molecular behavior is now being established convincingly through their ability to match reasonably well the spectroscopic, thermodynamic, kinetic and other experimental observations. These experimental data are the essential test which the computational models have to satisfy. As the examples discussed above, and increasingly others in the literature show, advanced MD simulations can generate satisfactory values for the reaction kinetics, residency times and thermodynamic parameters for protein mechanisms, in a range of protein systems. This ability of simulations to model atomic and molecular events matters because it brings in a new dimension to structural biology. It extends the understanding of biological mechanisms from local structural and chemical events to include specific molecular motions. In this connection, it seems to us that the catalytic process in enzymes is likely to exploit the energy available from local and global molecular motions just as much as from the particular chemical environment. With the development of computing there is little doubt that we shall see many more such studies, and very valuable they will be in providing a new, and much needed level of structural and mechanistic insight. Whatever the problems still ahead, the extent of progress in molecular simulations is remarkable and the long-standing skepticism of the experimentalists is no longer justified.

Acknowledgements. C. S. V. is adjunct faculty at the Department of Biological Sciences, National University of Singapore; BII is an A-STAR institute. Thanks to A. Suresh, S. Somani and M. O. Lombardia for help with figures.

1 Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H. et al. (2000) The Protein Data Bank. Nucleic Acids Res. **28:** 235–242
2 Nichols E., Milburn D. and Gerstein M. (2003) MolMovDB: analysis and visualization of conformational change and structural flexibility. Nucleic Acids Res. **31:** 478–482
3 Wüthrch K. and Wagner G. (1975) NMR investigations of the dynamics of the aromatic amino acid residues in the basic pancreatic trypsin inhibitor. FEBS Lett. **50:** 265–268
4 Sternberg M. J., Grace D. E., Phillips D. C. (1979) Dynamic information from protein crystallography: an analysis of temperature factors from refinement of the hen egg-white lysozyme structure. J. Mol. Biol. **130:** 231–252
5 Brooks C. L. III, Pettitt B. M. and Karplus, M. (1988). Proteins: A Theroretical Perspective of Dynamics, Structure and Thermodynamics, vol 71 Advances in Chemical Physics, Wiley, New York
6 Artymiuk P. J., Blake C. C., Grace D. E., Oatley S. J., Phillips D. C. and Sternberg M. J. (1979) Crystallograpohic studies of the dynamic properties of lysozyme. Nature **280:** 563–568
7 Caspar D. L., Clarage J., Salunke D. M. and Clarage M. (1988) Liquid like movements in crystalline insulin. Nature **332:** 659–662
8 Bentley G., Dodson E., Dodson G., Hodgkin D. and Mercola D. (1976) Structure of insulin in 4-zinc insulin. Nature **261:** 166–168
9 Bruschweiler R. (2003) New approaches to the dynamic interpretation and prediction of NMR relaxation data from proteins. Curr. Opin. Struct. Biol. **13:** 175–183
10 Rieping W., Habeck M. and Nilges M. (2005) Inferential structure determination. Science **309:** 303–306
11 Zewail A. H. (2005) Diffraction, crystallography and microscopy beyond three dimensions: structural dynamics in space and time. Phil. Trans. A Math. Phys. Eng. Sci. **363:** 315–329
12 Schotte F., Soman J., Olson J. S., Wulff M. and Anfinrud P. A. (2004) Picosecond time-resolved X-ray crystallography: probing protein function in real time. J. Struct. Biol. **147:** 235–246
13 Weiss S. (2000) Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. Nat. Struct. Biol. **7:** 724–729
14 Alder B. J. and Wainwright, T. E. (1957) Phase transition for a hard sphere system J. Chem. Phys. **27:** 1208–1209
15 Lisfon, S. and Warshel, A. (1968) J. Chem. Phys. **49:** 5116–5129
16 Rahman, A. and Stillinger, F. H. (1971) Molecular dynamics study of liquid water. J. Chem. Phys. **55:** 3336–3359
17 McCammon, J. A., Gelin, B. and Karplus, M. (1977) Dynamics of folded protein. Nature **267:** 585–590
18 Kollman, P. A. (1996) Advances and continuing challenges in achieving realistic and predictive simulations of the properties of organic and biological molecules. Acc. Chem. Res. **29:** 461–469
19 Zoete V., Meuwly M. and Karplus M. (2004) A comparison of the dynamic behavior of monomeric and dimeric insulin shows structural rearrangements in the active monomer. J. Mol. Biol. **342:** 913–929
20 Hummer G., Schotte, F. and Anfinrud, P. A. (2004) Unveiling functional protein motions with picosecond x-ray crystallography and molecular dynamics simulations. Proc. Natl. Acad. Sci. USA **101:** 15330–15334
21 Warshel A. (2003) Computer simulations of enzyme catalysis: methods, progress and insights Ann. Rev. Biophys. Biomol. Struc. **32:** 425–443
22 Feig M. and Brooks C. L. III (2004) Recent advances in the development of and application of implicit solvent models in biomolecule simulations. Cur. Opin. Struct. Biol. **14:** 217–224
23 Roux B. and Schulten K. (2005) Computational studies of membrane channels. Structure **12:** 1343–1351

24 Garcia-Viloca M., Gao J., Karplus M. and Truhlar, D. G. (2004) How enzymes work: analysis by modern rate theory and computer simulations Science. **303:** 186–195
25 Karplus, M. and McCammon, J. A. (2002) Molecular dynamics simulations of biomolecules. Nat. Struct. Biol. **9:** 646–652
26 Karplus, M. and Kuriyan, J. (2005) Molecular dynamics and protein function. Proc. Natl. Acad. Sci. USA **96:** 103–108
27 Norberg J. and Nilsson L. (2003) Advances in biomolecular simulations: methodology and recent applications. Q. Rev. Biophys. **36:** 257–306
28 Wagner G., DeMarco A. and Wuthrich K. (1976). Dynamics of the aromatic amino acid residues in the globular conformation of the basic pancreatic trypsin inhibitor (BPTI). I. 1H NMR studies. Biophys. Struct. Mech. **2:** 139–158
29 Denisov V. P., Peters J., Horlein H. D. and Halle B. (1996) Using buried water molecules to explore the energy landscape of proteins. Nat. Struct. Biol. **3:** 505–509
30 Fischer S., Verma C. S. and Hubbard, R. E. (1998) Rotation of buried water inside a protein: calculation of the rate and vibrational entropy of activation. J. Phys. Chem. B **102:** 1797–1805
31 Verma C. S. and Fischer S. (2005) Protein stability and ligand binding: new paradigms from in-silico experiments. Biophys. Chem. **115:** 295–302
32 Dunitz, J. (1994) The entropic cost of bound water in crystals and biomolecules. Science **264:** 670
33 Fischer S. and Verma C. S. (1999) Binding of buried water increases the flexibility of proteins. Proc. Natl. Acad. Sci USA **96:** 9613–9615
34 Fischer S., Smith J. and Verma C. S. (2001) Dissecting the vibrational entropy change on protein/ligand binding: burial of a water molecule in bovine pancreatic trypsin inhibitor. J. Phys. Chem. B **105:** 8050–8055
35 Olano L. R. and Rick S. W. (2004) Hydration free energies and entropies for water in protein interiors. J. Am. Chem. Soc. **126:** 7991–8000
36 Berndt K. D., Beunink J., Schroder W. and Wuthrich K. (1993) Designed replacement of an internal hydration water molecule in BPTI: structural and functional implications of a glycine-to-serine mutation. Biochemistry **32:** 4564–4570
37 Matthews B. W., Morton A. G. and Dahlquist F. W. (1995) Use of NMR to detect water within nonpolar protein cavities. Science **267:** 1813–1817
38 Ernst J. A., Clubb R. T., Zhou H. X., Gronenborn A. M. and Clore G. M. (1995) Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR. Science. **267:** 1813–1817
39 Yu B., Blaber M., Gronenborn A. M., Clore G. M. and Caspar D. L. (1999) Disordered water within a hydrophobic protein cavity visualized by x-ray crystallography. Proc. Natl. Acad. Sci. USA **96:** 103–108
40 Auron P. E. (1998) The interleukin 1 receptor: ligand interactions and signal transduction. Cytokine Growth Factor Rev. **9:** 221–237
41 Clapperton J. A., Martin S. R., Smerdon S. J., Gamblin S. J. and Bayley P. M. (2002) Structure of the complex of calmodulin with the target sequence of calmodulin-dependent protein kinase I: studies of the kinase activation mechanism. Biochemistry **41:** 14669–14679
42 Yang C., Jas G. S. and Kuczera K. (2004) Structure, dynamics and interactions with kinase targets: computer simulations of calmodulin Biochim. Biophys. Acta **1697:** 289–300
43 Afshar M., Caves L. S., Guimard L., Hubbard R. E., Calas B., Grassy G. et al. (1994) Investigating the high affinity and low sequence specificity of calmodulin binding to its targets. J. Mol. Biol. **244:** 554–571
44 Barton, N. P., Verma, C. S. and Caves, L. S. D (2002) Inherent flexibility of calmodulin domains: a normal-mode analysis study. J. Phys. Chem. B **106:** 11036–11040
45 Prabhu N. V., Lee A. L., Wand A. J. and Sharp K. A. (2003) Dynamics and entropy of a calmodulin-peptide complex studied by NMR and molecular dynamics. Biochemistry **42:** 562–570
46 Ellis R. J. (2003) Protein folding: importance of the anfinsen cage. Curr. Biol. **13:** R881-R883
47 Karplus M., Gao Y. Q., Ma J., Vaart A. van der and Yang W. (2005) Protein structural transitions and their functional role. Phil. Trans. A Math. Phys. Eng. Sci. **363:** 331–355
48 Vijayakrishnon S., Qamra R., Verma C. S., Sen R. and Mande S. C. (2005) Cation-mediated interplay of loops in Chaperonin 10. J. Biomol. Str. Dyn. in press
49 Hunt J. F., Weaver A. J., Landry S. J., Gierasch L. and Deisenhofer J. (1996) The crystal structure of the GroES co-chaperonin at 2.8 A resolution. Nature **379:** 37–45
50 Timchenko A. A., Melnik B. S., Kihara H., Kimura K. and Semisotnov G. V. (2000) GroES co-chaperonin small-angle X-ray scattering study shows ring orifice increase in solution. FEBS Lett. **471:** 211–214
51 Taneja B. and Mande S. C. (2001) Metal ions modulate the plastic nature of *Mycobacterium tuberculosis* chaperonin-10. Protein Eng. **14:** 391–395
52 Villanueva J., Hoshino M., Katou H., Kardos J., Hasegawa K., Naiki H. et al. (2004) Increase in the conformational flexibility of beta 2-microglobulin upon copper binding: a possible role for copper in dialysis-related amyloidosis. Protein Sci. **13:** 797–809
53 Dodson G. G., Lawson D. M. and Winkler F. K. (1992) Structural and evolutionary relationships in lipase mechanism and activation. Faraday Discuss. **93:** 95–105
54 Brzozowski A. M., Derewenda U., Derewenda Z. S., Dodson G. G., Lawson D. M., Turkenburg J. P. et al. (1991) A model for interfacial activation in lipases from the structure of a fungal lipase-inhibitor complex. Nature **351:** 491–494
55 Brzozowski A. M., Savage H., Verma C. S., Turkenburg J. P., Lawson D. M., Svendsen A. et al. (2000) Structural origins of the interfacial activation in *Thermomyces (Humicola) lanuginosa* lipase. Biochemistry **39:** 15071–15082
56 Ferrer M., Cruces M. A., Bernabe M., Ballesteros, M. and Plou F. J. (1999) Lipase-catalyzed regioselective acylation of sucrose in two-solvent mixtures. Biotechnol. Bioeng. **65:** 10–15
57 Woudenberg M., Van Rantwijk F. and Sheldon R. A. (1996) Regioselective acylation of disaccharides in tert-butyl alcohol catalyzed by *C. antarctica* lipase. Biotechnol. Bioeng. **49:** 328–333
58 Fuentes G., Ballesteros A. and Verma C. S. (2004) Specificity in lipases: a computational study of transesterification of sucrose. Protein Sci. **13:** 3092–3103
59 Weissmann C. (2004) The state of the prion. Nat. Rev. Microbiol. **2:** 861–871
60 Thirumalai D, Klimov D. K. and Dima R. I. (2003) Emerging ideas on the molecular basis of protein and peptide aggregation. Curr. Opin. Struct. Biol. **13:** 146–159
61 Fernandez A. and Scheraga H. A. (2003) Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. Proc. Natl. Acad. Sci. USA **100:** 113–118
62 Tobler I., Gaus S. E., Deboer T., Achermann P., Fischer M., Rulicke T. et al. (1996) Altered circadian activity rhythms and sleep in mice devoid of prion protein. Nature **380:** 639–642
63 Shorter J. and Lindquist S. L. (2005) Prions as adaptive conduits of memory and inheritance. Nat. Rev. Genet. **6:** 435–450
64 Bussard A. E. (2005) A scientific revolution. EMBO Rep. **6:** 691–693
65 Rezaei H., Eghiaian F., Perez J., Doublet B., Choiset Y., Haertle T. et al. (2005) Sequential generation of two structurally distinct ovine prion protein soluble oligomers displaying different biochemical reactivities. J. Mol. Biol. **347:** 665–679
66 Baskakov I. V., Legname G., Gryczynski Z. and Prusiner S. B. (2004) The peculiar nature of unfolding of the human prion protein. Protein Sci.**13:** 586–595

67 Knaus K. J., Morillas M., Swietnicki W., Malone M., Surewicz W. K. and Yee V. C. (2001) Crystal structure of the human prion protein reveals a mechanism for oligomerization. Nat. Struct. Biol. **8:** 770–774
68 Haire L. F., Whyte S. M., Vasisht N., Gill A. C., Verma C., Dodson E. J. et al. (2004) The crystal structure of the globular domain of sheep prion protein. J. Mol. Biol. **336:** 1175–1183
69 Dima R. I. and Thirumalai D. (2004) Probing the instabilities in the dynamics of helical fragments from mouse PrPC. Proc. Natl. Acad. Sci. USA **101:** 15335–15340
70 Eghiaian F., Grosclaude J., Lesceu S., Debey P., Doublet B., Treguer E. et al. (2004) Insight into the PrPC→PrPSc conversion from the structures. of antibody-bound ovine prion scrapie-susceptibility variants. Proc. Natl. Acad. Sci. USA **101:** 10254–10259
71 DeMarco M. L. and Daggett V. (2004) From conversion to aggregation: protofibril formation of the prion protein. Proc. Natl. Acad. Sci. USA **101:** 2293–2298
72 Fernandez, A. (2002) Insufficient hydrogen-bond desolvation and prion-related disease. Eur. J. Biochem. **269:** 4165–4168
73 De Simone A., Dodson G. G., Verma C. S., Zagari A. and Fraternali F. (2005) Prion and water: tight and dynamical hydration sites have a key role in structural stability. Proc. Natl. Acad. Sci. USA **102:** 7535–7540
74 Tousignant A. and Pelletier J. N. (2004) Protein motions promote catalysis. Chem. Biol. **11:** 1037–1042
75 Tama F., Valle M., Frank J. and Brooks C. L. III (2003) Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. Proc. Natl. Acad. Sci. USA **100:** 9319–9323
76 Rod T. H., Radkiewicz J. L. and Brooks C. L. III (2003) Correlated motion and the effect of distal mutations in dihydrofolate reductase. Proc. Natl. Acad. Sci. USA **100:** 6980–6985
77 Miller D. W. and Agard D. A. (1999) Enzyme specificity under dynamic control: a normal mode analysis of alpha-lytic protease. J. Mol. Biol. **286:** 267–278