Research News

Energetics of enzyme stability

Michele Vendruscolo

Advances in computational protein engineering depend on the development of an accurate treatment of the interactions among the 'elementary constituents' of enzymes. Most of the current research is focused on finding the right balance in the estimations of energy and entropy.

We are making progress in tailoring enzymatic function. There are currently two main approaches, directed evolution [1] and rational design [2]. The latter method requires an accurate knowledge of the rules of macromolecular assembly and stability. Here is where biophysics can be most helpful to biotechnology because we need to specify how different parts of a molecule come together or, in other words, to control molecular energetics.

Molecular machines

The goal is to compute the probability of observing a molecular complex in a given environment. In physical terms, we need to consider its free energy. This is because enzymes are molecular machines. Unlike macroscopic machines, of which the structure is usually rigid and the motion is predictable, the structure of enzymes randomly fluctuates under the constant effect of complicated molecular collisions. Therefore, notions such as that of a rigid structure and of deterministic motion become unsuitable at the molecular level, at least given our present computational capabilities.

Setting up the stage and recruiting the actors

Any attempt to compute (free) energies requires two preliminary choices: (1) to define which are the 'elementary parts' of the system; and (2) to define their interactions. One straightforward possibility is to define atoms as the elementary constituents of a macromolecular complex and then to use basic physical laws to deduce their interactions [3–5]. Typically, the Coulomb interaction is introduced in some form to account for electrostatics and the Pauli exclusion principle to account for the geometry (stereochemistry) of the complex. In practice, such an approach is not computationally convenient for the large-scale screening of possible single amino acid replacements. Two examples illustrate the complexity of the task. First, among the mutations that make HIV-1 protease resistant to drugs, only a fraction involve the catalytic site, whereas the others are found in distant parts of the structure [6]. Second, FNfn10 and TNfn3, two closely related fibronectin type III domains, respond very differently to mutations of corresponding residues in their hydrophobic cores [7]. Whereas TNfn3 reacts to mutations by local rearrangements, FNfn10 minimizes the impact of the mutations on its stability by a global reorganization of the structure. In other words, single amino acid replacements might result in long-range structural perturbations.

Trading energy for entropy

The calculation of the entropic consequences of a mutation requires taking into account a vast number of conformations. In principle, for each candidate mutation, all the conformations that differ from the native structure on the length scale set by the size of the rearrangements owing to the replacement must be considered. Even when relatively small conformational changes are considered, the number of such conformations can be extremely large. Thus, the accuracy achieved in computing the energy in atomistic models is usually lost in a poor estimation of the entropy. This is even more so when, in a quantum mechanical approach, the 'elementary parts' are taken to be electrons and nuclei [4].

The search for a viable approach

A long cherished goal in rational drug design is to find simple models that are both amenable to computations and able to give results accurate enough to be useful for predicting stabilities and binding affinities. Instead of atoms, amino acids can be considered as the 'elementary parts' of an enzyme. Unfortunately, this simplification comes at the price of blurring our understanding of the interactions. For example, we can define effective pairwise interactions between amino acids but it becomes very difficult to derive them from first principles. The same can be said for hydrophobic interactions, which have a major role in determining protein stability. But where does such a difficulty come from? In principle, the interaction between two amino acids is given by the sum of the interactions of their constituent atoms, plus the interactions with the atoms of the solvent in which they are immersed. The trouble appears when we want a single number to represent the energy involved in bringing two amino acids in spatial proximity, irrespective of the detailed arrangement of their atoms and of the state of their environment. In several cases it is possible to show that this is an impossible task [8]. However, other options are available. For example, one can consider environment-dependent energy parameters [9,10] in which residue–residue interactions are different for different conditions of the surroundings.

Another possibility, recently investigated by Carter *et al.* [11] is to define four-body interactions, namely to assign an energy term specific for the situation in which four residues come in spatial proximity. To carry out this plan, several decisions have to be made. First, a recipe has to be designed for defining when four objects are in spatial proximity. Carter *et al.* do this using Delaunay tessellation, which is an unambiguous way to partition the space occupied by an enzyme into tetrahedrons corresponding to quadruplets of neighbouring amino acids. Then one has to define 8855 energy parameters. This is the number of possible distinct quadruplets formed with 20 species of amino acids. By specifying a value for each one of these parameters, a point in the 8855-dimensional space of interaction parameters is assigned. Carter *et al.* single out a point in such a space by calculating the frequencies with which quadruplets appear in the protein structures deposited in the Protein Data Bank. This is a way of

http://tibtech.trends.com 0167-7799/02/$ – see front matter PII: S0167-7799(01)01871-6

determining a point in energy parameter space usually referred to as the quasi-chemical method [12]. Other approaches are also possible [7–15]. For example, given that the physical interpretation of the effective energy parameters is not straightforward, one can, at least temporarily, avoid being concerned about it and ask whether there is at all a set of energy parameters such that the experimental stability changes owing to mutations are exactly equal to the calculated ones. If such a set exists then the road is paved and one can hope to generalize the result to compute stabilities that have not yet been measured. Carter *et al.* [11] have shown that four-body energetics [16] is a promising improvement over two-body energetics and that further study on this route is likely to bear fruit.

So, can we predict the effect of a small perturbation on an enzymatic process? This question is among the most fascinating in protein science [17]. For example, altering a single group can enhance the binding affinity of an enzyme to its substrate, reduce it or sometimes even nullify it altogether [18]. Our increasing ability to accurately predict such outcome is opening the way to the systematic improvement of existing enzymes by computational protein engineering, for drug manufacturing and biocatalysis.

References

1 Kuchner, O. and Arnold F.H. (1997) Directed evolution of enzyme catalysis. *Trends Biotechnol.* 15, 523–530
2 Dahiyat, B.I. and Mayo, S.L. (1997) *De novo* protein design: fully automated sequence selection. *Science* 278, 82–87
3 Lazaridis, T. and Karplus, M. (1997) 'New view' of protein folding reconciled with the old though multiple unfolding simulations. *Science* 278, 1928–1931
4 Wang, W. *et al.* (2001) Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein–ligand, protein–protein and protein–nucleic acids noncovalent interactions. *Annu. Rev. Biop. Biomol. Struct.* 30, 211–243
5 Guerois, R. and Serrano, L. (2000) The SH3-fold family: experimental evidence and prediction of variations in the folding pathways. *J. Mol. Biol.* 304, 967–982
6 Wlodawer, A. and Vondrasek, J. (1998) Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. *Annu. Rev. Biophys. Biomol. Struct.* 27, 249–284
7 Cota, E. *et al.* (2000) Two proteins with the same structure respond very differently to mutation: the role of plasticity in protein stability. *J. Mol. Biol.* 302, 713–725
8 Vendruscolo, M. and Domany, E. (1998) Pairwise contact potentials are unsuitable for protein folding. *J. Chem. Phys.* 109, 11101–11108
9 Park, K. *et al.* (2000) Towards an energy function for the contact map representation of proteins. *Proteins* 40, 237–248
10 Banavar, J.R. and Maritan, A. (2001) Computational approach to the protein-folding problem. *Proteins* 42, 433–435
11 Carter, C.W. *et al.* (2001) Four-body potentials reveal protein-specific correlations to stability changes caused by hydrophobic core mutations. *J. Mol. Biol.* 311, 625–638
12 Lu, H. and Skolnick, J. (2001) A distance-dependent atomic knowledge-based potential for improved protein structure selection. *Proteins* 44, 223–232
13 Xia, Y. and Levitt, M. (2001) Extracting knowledge-based energy functions from protein structures by error rate minimization: comparison of methods using lattice model. *J. Chem. Phys.* 113, 9318–9330
14 Koretke, K.K. *et al.* (1998) Self-consistently optimized energy functions for protein structure prediction using molecular dynamics. *Proc. Natl. Acad. Sci. U. S. A.* 96, 2332–2337
15 Mirny, L.A. and Shakhnovich, E.I. (1996) How to derive a protein folding potential? A new approach to an old problem. *J. Mol. Biol.* 264, 1164–1169
16 Liwo, A. *et al.* (2001) Cumulant-based expression for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field. *J. Chem. Phys.* 115, 2323–2347
17 Fersht, A.R. (1999) *Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding.* W.H. Freeman & Co. New York
18 Wen, J. *et al.* (1996) Exploring the allowed sequence space of a membrane protein. *Nat. Struct. Biol.* 3, 141–148

Michele Vendruscolo
University Chemical Laboratory, University of Cambridge, Lensfield Road, Cambridge, UK CB2 1EW.
e-mail: mv245@cus.cam.ac.uk

Energetics of enzyme stability

Response from Carter, Tropsha and Edgell

We welcome the 'Research News' article about our recent paper [1] establishing a connection between four-body database-derived likelihood potentials and experimental mutational free energy changes.

'...a decisive new coherence to side-chain packing analysis.'

The Delaunay tetrahedron is the simplest three-dimensional packing motif; we noted in our article, and presented more fully elsewhere [2] that, unlike potentials of smaller dimensionality, joint consideration of four interacting side chains achieves excellent proportionality between statistical and thermodynamic four-body potentials (calculated by summing the transfer free energies, or hydrophobicities, of the four interacting side chains). Decomposing tertiary structure into elementary three-dimensional simplices therefore appears to afford a decisive new coherence to side-chain packing analysis.

Practical use of higher-dimensional potentials may also benefit from our observation that mutational $\Delta(\Delta G_{unfold})$ values for different proteins scale differently to the four-body potentials. This seemingly counter-intuitive result reflects the fact that different proteins employ hydrophobic cores with different numbers of contributors, on average, per residue. Thus, even if two hydrophobic core mutations have similar contexts in different proteins, the observed free energies are often different, owing to different proportionate changes in the overall contributions made by hydrophobic bonding to stability. In Fig. 1, this effect is illustrated graphically for two valine to alanine mutations in staphylococcal nuclease (STN; blue) and chymotrypsin inhibitor 2 (CI2; gold). Side chains surrounding the mutated residues are similar enough in the two proteins that the predicted likelihood potentials are

 0167-7799/02/$ – see front matter PII: S0167-7799(01)01872-8