Chapter 1, Volume 2 of the Protein Primer

Errors and misconceptions in modern protein research.

In Chapter 1 in Protein-solvent interactions, edited by R. Gregory for Dekker New York 1995 I showed how the new paradigm for protein research required by the discovery of protein substructures reveals several very serious errors in methodology and fundamental chemical understanding of protein construction and function. A reexamination of the situation in 2004 found that none of the most serious errors that have impeded research on protein structure and function had disappeared. Instead by repetition they have become so ingrained that they are rarely questioned. Some of the errors just emerging in 1995 have now emerged as major and provide the justification for this revision. The errors listed here almost guarantee failure and explain the continuous deterioration of quality in protein and enzyme research. Pictorial details mostly obtained with diffraction methods are still expected to provide sudden revelations of mechanism without serious attention to fundamental chemistry and physics. Although there has accumulated abundant evidence that the unique features of enzymic catalysis are due to the successful evolution of the special, quite unfamiliar devices packaged in proteins, it is quite possible in the current climate of error to publish papers without any basis in what is actually known about proteins. Protein research is not unlike the efforts of a man looking for his lost collar bone under the street light where the light was better than in the place he lost it but goes farther insofar as many protein investigators believe the small island of light is all there is. Thus the knot and matrix substructures are not recognized because proteins are thought to by homogeneous. Conformation changes on which physiological function depends are not detected because the errors in coordinates are larger than their changes in function. Research is focused on the classical secondary structures and residue conservation neither of

which has played a central role in evolution. By habit and default thermodynamics continues to be thought the best source of quantitative information on mechanism but Benzinger showed in 1967 that such information it provides is rarely reliable. The biosphere is not at equilibrium with respect to any variable so its behavior is not dominated by minimizing free energy but almost entirely by irreversible entropy production. Much of familiar small molecule chemistry is neither applicable nor unambiguous.

The items listed here accumulated during the preparation of volume 1 of the Protein Primer but some omitted originally have now been added... We have tried to eliminate those still not completely tested for adequacy or scientific reliability. Those that remain for the following listing are essential information for successful protein research. The supporting evidence for these deductions is to be found in the Protein Primer and in earlier discussions reprinted or referred to on that web page.

1. The famous secondary structures of Pauling, Corey and Branson. Vary in structural integrity often having no greater strength than the amorphous regions. Only the atoms with lowest B values establish folded stability. For example, in large antiparallel β-sheet arrays the B factors decrease toward the center region and may be the major source of thermodynamic stability. .

2. ***Nature conserves free-volume arrangements rather than residues and residue sequences***. SDM experiments are thus unguided and usually difficult or impossible to interpret. All residues contribute to structure and function so they all interact with each other to some extent impossible to estimate in less than wholesale exchanges quite beyond current resources.

3. Proteins have at least two common substructures and these have very different functions. The small, hard substructures determine thermodynamic, genetic and kinetic stability. These have been called knots for obvious reasons and they have minor flexibility in function. There does not appear to be any obligatory relationship between knots and secondary structures although the latter are frequently found in knots. The larger common structure is dominated by the knots but undergoes the major mutations leading to improvements in physiological function. These are called matrices as a consequence. Folded stability is due in large part to small, strong groups of residues, usually 12% in mesophiles. However, clathrate hydration structures at protein surfaces appear to be equally important. The high strength of the small substructures, known as knots is due to efficient packing favoring electrostatic interactions especially hydrogen bonds and probably to less familiar covalency changes in large peptide arrays. **Kevlar** and spider dragline silks provide possible models but polypeptides may have covalency possibilities making polypeptide structures similar to carbon nanotubules.Important parts of polypeptide chemistry are probably still unknown..

4. Precision in residue placement atom by atom in modern x-ray diffraction studies of proteins is 0.05 Å and may be found to be even lower as resolution improves. For reasons not yet fully explained the B factor precision even at resolutions of 2 Å are very precise. Conformation changes in catalytic function are less than 0.4Å in atom positions and less than 1Å overall...This

precision is essential in controlling van der Waals and hydrogen-bond potentials.

5. Experimental error in protein diffraction experiments cannot be determined because the true values of coordinate variables are not known. A reliable estimate can be deduced from the fact that the special construction features such as the knots are clearly revealed with high precision by the B factors but not by the coordinates from the same diffraction data. The latter are so much larger than the errors in the B factors that the experimental errors in lengths and angles from the diffraction data hide the conformation changes. The B factors are measures of the mean square error in atom positions from which an average displacement error can be estimated and these are of the same order as the true errors. One consequence is that the B factors and only the B factors contain the reliable quantitative information about structure. The latter are not used in this way because of assumptions made in the early history of protein structure study. They have now some minor weakness in their dependence on experimental temperature, solvent or suspension medium and operator only beginning to be studied and explained.

6. The implicit experimental error in protein x-ray-diffraction methods is large and does not diminish rapidly as resolution improves. This is true for atom separation and angle coordinates but not for B factors. X-ray-diffraction pictures have essential qualitative importance but the essential quantitative biology must be extracted using experiments designed to utilize the B factors

7. To a considerable extent the conformation changes supported enzymic function are suppressed by standard sample preparation in diffraction studies. Polyethylene glycol and ammonium sulfate are the two most used crystallization agents because they sequester more water than most others with the result that the protein surface free energy rises forcing contraction. The degree of contraction is a large fraction of the matrix contraction by which matrix potential energy is transferred to substrate and its contact groups of the protein. Although such changes are small relative to the coordinate errors in diffraction studies, they are not small compared to the errors in B factors and it is the latter which make progress in protein research possible.

8. Potential-energy functions to produce reliable results in computing should have a seventh-order dependence on atom separations. Estimates from coordinate values are usually based on classical small-molecule functions inappropriate for proteins even at the mean-field approximation. Improvements can be extracted from the B values computed from very high resolution data for all three moments of inertia of the scattering ellipsoids. That is difficult but probably achievable n the distant future... Equally large errors limit the precision and thus the accuracy with which protein structures can be determined by NMR methods such as the determination of nuclear Overhauser effects. NMR relaxation methods have great potential importance for conformational dynamics on which enzyme function depends but until the latter is generally recognized will remain unpopular.

9. In view of the low accuracy of coordinate determination it is surprising that so much reliance is currently based on their use in computations. Similar problems have always jeopardized small-molecule computations despite the better accuracy of coordinate parameters. The explanation appears to be that there is no alternative except the B factors and for reasons that appear to be entirely psychological those are universally ignored.

10. Most enzymes have at least one pair of dynamically matched functional domains with pure C-2 rotational symmetry about the interdomain cleft in which the actual local chemistry of catalysis takes place. Although the C-2 rotation symmetry this construction produces is common and thus indicative of some common feature of evolution in enzymes its function is obvious and apparently universal. It varies to provide specificity in binding and catalytic rate with substrate but also with the complexity of the catalytic process sometime achieved by a single compression cycle and sometimes requiring several such strokes. Thus the closing and opening of the crevice is sufficient for chymotrypsin and trypsin, a two-stroke mechanical motor but at least three strokes are required in the pyridoxyl phosphate enzymes and probably four such strokes in glutamate dehydrogenase according to Fisher. These enzymes depend on the mechanical excitation of a pretransition state rather than the thermal activation into the classical transition state and that provides opportunities for the evolution of very complex processes. The biosphere depends on it as it has not provided

able to depend on the "transition-state stabilization" operating in most small molecule rate processes.

10. Myoglobin is one of the few proteins so far found to have a fragile free-energy surface. Most familiar proteins including all enzymes have strong surfaces. Myogloibin has no knots, The proton-exchange rates and B factors are those of matrices. It has low thermal stability with many substates giving in low mechanical strength. Despite its historical importance along with tetrameric hemoglobins in the study of allosteric linkage ("heme-heme interactions") that linkage is weak. It is a poor model for most other proteins illustrating the danger of generalizing from a single protein.

Convergence in enzymic mechanisms appears to be total in that most and probably all share the same construction features and the same mechanism. The most fundamental feature appears to be the C2 rotation symmetry required to mobilize potential energy. The mechanism resembles a nutcracker in which the differences in function and specificity are accommodated much as a nutcracker accommodates different nuts. But the evolutionary variable is atom free volume rather than residue conservation and the same residue can contribute to the formed in different ways depending on its neighbors. The two branches of the knot palindromes are matched in free-volume distribution but there is no obvious pattern of residue palindromy. That may be misleading because the knot palindrome is not only an invariant construction feature of enzymes but also appears in most proteins. That suggests that some version of the knot palindrome always occurs in DNA expression and

perhaps in RNA encoding. Matrices of enzymes to have more limited palindromy even during nutcracker contraction. None of these features so easily detected in the B-factor patterns are easily rationalized with the currently accepted generic codes and mechanisms of DNA expression. That has become apparent with the sequencing of genomes.

12 Kauzmann's original suggestion that the poor solubility of amphiphiles in water forces folding of the polypeptide once released from the ribosome has been misapplied in the sense that the normal product of thermal denaturation is a slightly expanded but much more motile version of the native species rather than anything similar to the random coil species freshly released from the ribosome. Conversion of the latter to the bubble state is drive by poor solubility but conversion of bubble to native species is drive by the loss of electrostatic potential energy. The poor solubility of amphiphiles in pure water has been shown to be due to the weak reaction field between water and the hydrocarbon parts of the amphiphiles plus relatively poor water solubility of the peptide-bond structure, This change in point of view was required once it became established that water had two chemical species. Poor solubility is now known to be due to the enthalpic cost of converting high-density water to its low-density state. The concept of "hydrophobic bond" so well described by Kauzmann and Tanford has been turned on its head. Such strength as it has comes not from van der Waals interactions but from the poor reaction field between water and oily groups as in fact both authors noted.

13 Recognition that pure water has two chemical species removes much confusion and allows explanation of the Hofmeister series in terms of only two kinds of cosolvents, structure makers and structure breakers.

The effects of structures breakers like urea on protein stability appear to be largely an indirect response to changes in the chemical description of aqueous mixtures. Thus for example direct interaction between protein and urea is probably not of major importance. Because amphiphiles even small ones sequester a large number of water molecules in their clathrate hydration cages, competition for water establishes a general linkage among structure breakers even in macro mixtures in water.

14 Thermal denaturation of protein in highly aqueous solvents produces a small, motile bubble roughly having less than twice the volume of the native species. It has been possible for many investigators to miss the extensive literature support for this conclusion. Caught in the middle is the so-called "molten-globulin" species postulated as a distinct intermediate between an extensively unfolded denatured species and the full-folded native species. The physical characteristics used to justify its existence resemble those of a half melted native species or the bubble species itself. Both explanations are consistent with the data and superior to a molten-globule species but they do not yet completely rule out the molten-globule alternative...

15 The second major substructure in mesophile proteins undergoes a phaselike expansion contraction process known as a subtle-change because of the peculiarities of its thermodynamic changes. It has come to be called a matrix. Each enzyme consists of two separate proteins matched in mass and free-volume distribution to as move together under matrix contraction. In that process potential energy provided by matrix contraction is transferred to the reaction assembly consisting of substrate and functional sidechains one group from each matrix. In contrast to popular belief the reaction assembly is excited in this way so it then passes

though the conventional transition state with only a small supplement of thermal excitation. Mechanical activation of rate processes makes biology very different from small-molecule chemistry. This nutcracker process depends more on transition state destabilization than "transition-state stabilization"

16 In enzymes convergence in evolution appears to be total. Exceptions have not yet been found. The essential mechanistic feature seems to be the potential-energy reservoir to close the nutcracker and the structural variations necessary to establish the high order specificity unique to proteins. The rectangular-hyperbolic rate law known as the Michaelis-Menten law is not supported by the classical explanation and requires either reformulation of drastically revised description. The catalytic process is intrinsically irreversible because a cycle is completed by chemical change to effect product removal allowing the return of the protein to its initial state. A steady-state model rather than an equilibrium model is required making both detailed balance and microscopic reversibility suspect.