Brief descriptions of the topics included in volume 1 of the Protein Primer.

(The Protein Primer can be downloaded in whole or in part from the web site http://www.chem.umn.edu/groups/lumry.) Newer references will be found there and many figures.

I. Substructures

a. Knots, matrices and surfaces

Matrices and knots are the major substructures of mesophilic proteins (defined as being thermally stable above about 280K and below 354K). They were distinguished before 1982 by proton-exchange rates and after by the temperature B factors measured in x-ray diffraction studies of proteins. These "B factors" measure the mean square displacement of atoms from their idea lattice positions most often on the assumption that the scattering is isotropic. Their high precision has not been appreciated nor has it been fully explained. Most of the quantitatively useful information in diffraction studies can be extracted from these factors and in no other way.

In mesophiles matrices are about 85% of the residues but are not intrinsically stable and are maintained in useful conformations by the knots thus reducing the stability of the knots. This is the major explanation for the small overall standard free energy increase in thermal denaturation of proteins. Stability then is due to the 12% of the residues in the knots so called because they have to be "untied" to allow unfolding. Knots are very strong resembling spider silk and KEVLAR. The explanation for their strength has not been established nor is the important connection between proteins and the dragline spider silks generally suspected. The crippling error has been the assumption that proteins have a homogeneous construction.

The construction unit of proteins consists of a knot with its matrix and some residue collections near surfaces called "surfaces". Surfaces are not yet well characterized and may be in fact an artifact due to charge cloud and position near surfaces. Woodward and coworkers showed that folded integrity as measured by rate of proton exchange at residues near surfaces is much higher than that of in true random coils although fast relative to knots and matrices. The very slow exchange rates at knot sites were originally used to distinguish knots and matrices.

b. Knots establish genetic stability

Knots are also the basis of generic stability because they divide the polypeptide into segments of fixed length. Hence preservation of the knot pattern is the fundamental requirement for the preservation of a protein family. There are other severe requirements necessitating Darwinian selection of all the residues consistent with knot selection, gestalt functional requirements and specificity in specific function. All these characteristics are somehow encoded by residue selection but residue conservation even within a single protein family is poor.

c. Matrix contraction

Matrix contraction is the most common way to establish mechanical effects in biology and may be the only one. It depends on the manipulation of free volume but the positional changes of atoms are only a few tenths of angstroms essentially the same as the geometric changes in primary-bond chemistry. The volume reduction of the total protein in its physiological function has been estimated to be about 4% producing a half angstrom reduction in protein diameter. Relative to matrices knots are well packed and hard and undergo only very small further contraction as their matrices contract.

d. Knot-matrix cooperation

Knots establish thermodynamic, kinetic and genetic stability leaving matrix to be tailored for function. The single free energy barrier in their two-state melting process is modulated by the degree to which the knots enforce the conformation of the matrices. The difference in this respect in forming the transition state for melting between expanded and contracted matrices is not large in free energy nor in enthalpy, possibly no more than 30kcal of enthalpy in the most extreme case but by no means negligible and sufficient to explain the largest differences between homogeneous and enzymic process.

II. Temperature factors

Availability

The Debye-Waller scattering factors known as “temperature factors” in protein chemistry always computed in high-resolution X-ray-diffraction work are very useful. Coordinate errors in X-ray diffraction are large and obscure the substructures. The temperature-factor errors are much smaller. With precision of 0.05Å in modern studies they provide the practical basis with which to extract quantitative information about protein conformations from diffraction studies. NMR methods for structure have large errors in coordinate determination but they have many secondary applications of major importance especially in proton-exchange rates and conformational dynamics in general.

a. B factors diagram: the evolutionary history of a protein and the explanation of matrix contraction.

As illustrated by Figure 1, the atom B factors describe the detailed folding of a protein. It can be seen that in this example as in others, the residues are oriented in a radial fashion from the lowest B values in or near their knot end to higher B values. Matrix contraction then is to lower them in this order in uniform contraction toward contracted states increasingly resembling the knot in B values so that at maximum matrices resemble familiar glassy states of polymers although by no means identical because of the unique construction achieved in evolution.

The accuracy of the B factors in protein crystallography is not yet well explained. Recent progress in protein crystallography make it possible to compute with some reliability all three moments of inertia but the improvement is not yet useful again because the special features due to natural selection rather than ergonic considerations have yet to be factored into the analysis.

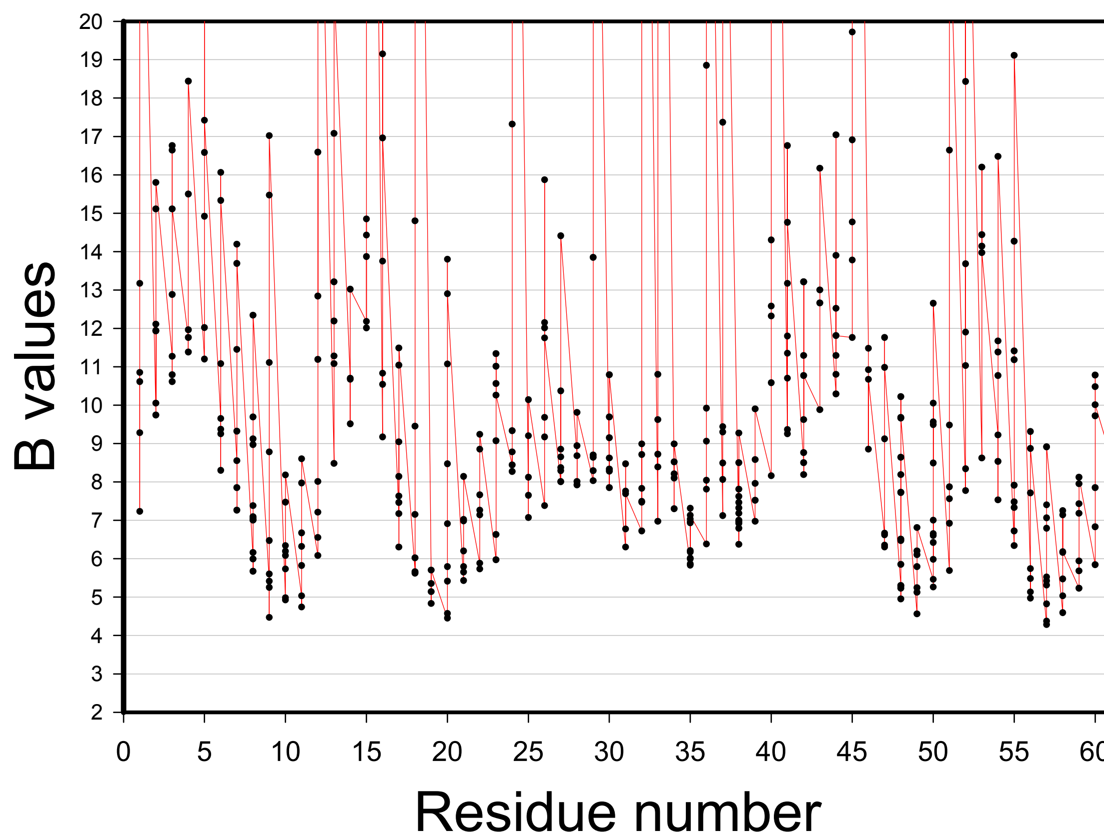

Figure 1. Structure plot of the G protein from streptococcus, IGG binding section. 61 residues. 1IGD pdb (or edited 2IGD.pdb) 1.1 angstrom resolution. The scalar couplings and H-bond lengths are from Cornilescu, Ramirez, Frank, Clore, Gronenborn and Bax, JACS 121, 6275 (1999). Figure from Lumry, Biophysical Chemistry 101-102 (2002).page 81. Some disagreement about atom numbers from the two high-resolution pdb files but residues numbers agree.

The above example of the method of plotting gives a free-volume picture of the entire protein showing quantitative agreement between two knots (clusters of lowest B values), excellent palindromy between knots and a rough two-fold symmetry in the higher B values. The knot B factors are in very good agreement with H-bond lengths and scalar couplings from the reference above. The lowest B values and thus shortest H bonds form the inter-domain hinge consistent with the largest j coupling reported by Cornilescu et al (reference above). The proton-exchange rates can be presumed to increase with increase in local B- factor. There is a stretch of fused amide bonds at the center (lowest B values) that may be a result of proton delocation across the inter-ring H bonds adding covalency to the rings The new bonding looks similar to that in

KEVLAR and quite promising as a step toward relating knot strength to the strength of spider silk. There is a long helix between diagonal extremes (not shown above). It has some short H bonds but no B values close to those of the knots. The dyad symmetry of the knot residues is centered on an axis from residue 34 through the middle of the helix. In contrast to enzymes there are several short H bonds between the knots making an unusually large and strong hinge but giving no indication of how the palindrome contributes to function in this protein.. Ubiquitin is similar and also not an enzyme since its domains are also tightly connected.

b. Substructure information from B factors.

Two of the most important features uniting proteins as a common state of matter are immediately revealed by B factors. Proton-exchange rates provided the characteristic temperatures of knots to be 354K and 450-500K for matrices. With mesophiles these parameters can be used to identify participation of one or the other. Only these substructures have such superficially simple behavior and most protein structure and function involves the two as is demonstrated by the varieties of their enthalpy-entropy compensation behavior that can be demonstrated to have slopes equal to one or the other value.

For mesophilic enzymes, the B factors provide more detail showing that after eliminating such extra sections of peptide as leader peptides and sometimes separate domains for cofactors binding, it is found that the functional domain pair always has a precisely positioned middle atom and residue separating the matched pair leaving them connected only by one of several kinds of hinge. The domains revealed in this way are appropriately called "functional domains" and have no close similarity to other domains definitions now popular. They have equal mass but little residue similarity and each contains one branch of the knot palindrome. The resulting matching in free-volume distribution appears to be consistent with dynamic rather than static matching.

The position and B values of matching knot atoms in the palindrome are identical within experimental errors although the latter often make it difficult in lower-resolution studies to describe more than the midpoint and first matching pair. Since the inter-domain interaction determines the relative positions of the chemically functional groups, those groups can also aid in finding the center of the B-factor palindrome. Matrices do not have precise palindrome and such symmetry as they show in expanded state is usually reduced or eliminated by their contraction. Aside from experimental error sometimes large a few proteins show dynamic matching in the B factors rather than the knot palindrome. Other patterns may appear but thus far have not.

III. Protein construction from substructures

a. Dyad rotation symmetry as the unifying theme in enzyme evolution.

The B-factor palindrome always indicates the knots to have C-2 rotation symmetry rather than a mirror plane. C-2 symmetry is also found in many non-enzymes, membrane transfer proteins including the visual-pigment complexes and even in some larger organelles. It is obviously of major importance in biology in such discoveries as molecular motors and active transport in membranes so it is surprising that it has attracted so little attention.

In enzymes and relative to small-molecule chemistry the large protein molecules with C-2 construction have the speed and specificity in catalysis adequate for biology. The two achievements are combined in a single mechanism supporting virtually unlimited specificity in substrate, activator and primary-bond rearrangements The C-2 pattern may be slightly misleading since T-1 nuclease for example has closely matched knots without exact palindromic construction. However, T-1 may be an exception required for its polymeric substrate rather than an indication of a more sophisticated level in normal protein evolution as is suggested by similar exceptional behavior of T-4 lysozyme and ribonuclease A. The use of the palindrome and other indications of dynamic symmetry in non-enzymic proteins is not always clear. (See figure 1 for the G protein of streptococcus, a non-enzyme with high resolution and thus a very precise palindrome)

b. Domain closure

The C-2 symmetry and the matrix contraction produce domain closure. Its role in enzyme mechanisms is suggested by the positioning of two chemically functional groups one to each domain and so placed in space that they are forced to chemical distances as the matrices contract The domains move together like a nutcracker but with a built in source of transient force. At maximum closure the two groups are briefly fused at high potential energy forcing distortion, proton and electron delocalization in proteases, ligand-field distortion at the zinc ion in carboxypeptidase A or whatever alternative required to excite the pretransition species from which the system “falls” through the transition state to products usually with a small thermal assist measured by the apparent activation enthalpy. Each of these “functional domains” is a complete protein and many kinds of proteins are only a single functional domain; e.g.: bovine pancreatic trypsin inhibitor. Dynamic matching in closure appears necessary to maximize the force generated in the process and to decouple domain closure from whole molecule. Domain-closure is at least as important as a universal device to preserve L and D stereospecificity rotation (more detail in Lumry in Gregory, Protein-solvent interactions and Lumry-Biltonen, Structure and function of biological macromolecules, Dekker 1969)

c. The precision of protein construction is far better than generally realized

The large coordinate errors hide precision in atom positioning just as they hide the expansion-contraction process of the matrices. At x-ray resolution of 1Å the palindromic knot atoms of the catalytic domain pair are matched to about

0.05Å.perhaps better at higher resolution. This is computed from the B factors as averages of the mean square deviation of an atom from its ideal lattice position and it is also a good estimate of the precision of construction of the entire protein with exceptions as in fragile proteins where multiple chain conformation possibilities obscure the refinement. The small B-factor error shows that evolution has been able to find exceptional packing for every enzyme. But successful packing for folded stability is perhaps the simplist requirement for a successful enzyme. The results do not resemble any of the conventional polymers so extensively studied in conventional chemistry. That means that in contrast to the latter, free-energy minimization is not a major guiding feature in natural selection. It is rather an obstacle. Nature has found better polymers and more efficient chemistry throughout, a general hallmark of biology.

IV. Thermodynamic changes in enzymic catalyis

a. Biology is made possible by mechanical rather than thermal activation

It has been said that the struggle for survival is fought out along the conformational compensation coordinate. Certainly most of biology manifests enthalpy-entropy compensation behavior though it may be only an artifact of the close adherence of the matrix expansion-contraction process to the two-state model of phase change. Both mean-field potential functions and linear-response behavior are such artifacts of the dependence of physiological function on that process. Proton-exchange and enzymic catalysis are characterized by the same compensation temperature the first because of matrix expansion, the second because of matrix contraction.

It seems likely that evolution has not needed to go much beyond the discovery of potential energy management through conformational adjustment. It was a big discovery quite as much as that of DNA because the one will not work without the other. Alternatives for rate enhancement such as the electronic catalytic mechanisms ubiquitous in small-molecule chemistry must have proved to be much inferior. If all modern enzymes support potential-energy management via matrix expansion and contraction, that device may be the responsible for the production of all mechanical force in all biology. That is a good hypothesis with which to search for a genome science by using protein-gene mapping now that protein B factors provide the necessary sophistication for such a method. We can suspect that the presence of C-2 symmetry implies some version of the nutcracker mechanism even in muscle, other protein motors and perhaps even in some organelles.

What are the uses of proteins like the G protein described in Fig, 1 that have very closely matched knots but fused by a system of hydrogen bonds apparently eliminating nutcracker construction? Also to be explained are the differences in construction of the hyperthermophiles and other extremophiles from archaea compared with the mesophiles? There may yet be found major differences in

mesophile construction complicating the superficially simple contraction and function principles illustrated in the Protein Primer.

b. All enzymes use the same mechanical mechanism

All enzymes so far examined using B factors have been found to have the same set of features. The C-2 symmetry supporting symmetrical domain closure obviously suggests a nutcracker and its generality supports the proposition that all enzymes use the same mechanism. That was suggested by the behavior of several conformational variables in 1960 long before x-ray-diffraction became available. The x-ray diffraction studies of chymotrypsin some years later confirmed the earlier deductions but were too crude to add any of the new information since discovered to be hidden in the B factors. Thus far HEW lysozyme is the only enzyme clearly deviating from perfect nutcracker in having unmatched domains and we can guess considering its poor catalytic efficiency and structural specialty, that it is a holdover from the era of the previous kind of enzyme.

The differences among enzyme families do not lie in the two-domain constructions but rather in the composition of the matrices and the functional pair forming the reaction bridge between the matched functional domains The earliest proposal of a mechanical protein mechanism is the "rack" mechanism of Eyring and coworkers in 1953 based on the assumption of the tension and expansion of the proximal imidazole-to-iron bond connecting mechanically the iron complex ion to the tertiary conformation of the hemoglobins. The subsequent examination of chymotrypsin suggested that compression was the correct mechanical mode so the new name first suggested by Carloni is "nutcracker". Most details of this mechanism were given in 1969 by Lumry and Biltonen in their review of the subtle change process necessary for enzyme function.

c. The "nutcracker" is a better name for the enzyme mechanism

Eyring and coworkers reasoned that the exceptional rates and specificities of enzymes know then for turnover rates as at least ten orders of magnitude greater than those in homogeneous catalysis cannot be due to thermal activation, so probably depend on mechanical activation by the free-volume management possible. The large conformations of proteins are subject to evolutionary tailoring for a wide range of specific functions and specificity in adaptation to substrates. The nutcracker formed by the two domains closes as the matrices contract forcing the functional groups together and down on the reaction assembly so as to excite the substrate which then decays through the classical transition state to products. Electron and proton migrations can be forced to take place and in such examples as the proteases, esterase and hydrolases as a class the substrate trapped in contact with the functional groups often appears to be distorted at the primary bonds about to undergo rearrangement. Similarly the ligation of metal ions such those of zinc in carboxypeptidase A appears

to be temporarily forced away from ground-state configurations to make the water ligand reactive.

As compared with classical ideas about enzymic mechanisms, the critical event is the increase in substrate free energy in a pretransition state rather than direct passage into the classical postulate of the transition state. In general this is the result of transient transfer of potential energy from matrix on contraction to raise the lowest energy level of the reacting assembly. There may be other redistributions or alternatives based on entropy increase but such examples have not yet been reported and require a reservoir for negentropy not yet described. Also in the potential-energy mechanism there is little potential energy lost as heat so little net enthalpy change but there is a major entropy loss from the matrices during contraction. That must be balanced by an entropy gain not yet located but probably hidden in the hydration contribution In any event total replacement or rerationalization of Michaelis kinetics is required.

d. How nature uses coenzymes, electron and proton migration

In acid-base reactions as in pepsin or the NAD enzymes, in which the redox end of the cofactor replaces an acidic or basic group of the reacting pair in pepsin, the reaction path up to and through the transition state can be followed by electron or more unambiguously by proton transfer as a function of the degree of domain closure. This is becoming a common approach but the results are confused by neglect of domain closure so that quite different results have been reported for pepsin as a result of the experimental conditions which effect different degrees of domain closure. In pepsin, Zundel finds that the small but very effective competitive inhibitor pepstatin forces complete proton migration across the single inter-domain hydrogen bond. The more efficiently a substrate is catalysed, the more effective it is in triggering domain closure. However, domain closure is not a result of the free energy released locally by contract interactions but rather is driven by matrix contraction. Specificity in catalysis and in small-molecule binding can be explained by the triggering of domain closure because evolution has found two matrix states that are very nearly ergonically balanced, the actual free energy change in binding pepstatin is close to that computed from the equilibrium binding constant. The enthalpy and entropy changes compensate but nevertheless contain most of the useful information about the protein mechanism. For pepstatin on pepsin the free energy change though large considering the small size of pepstatin is small compared with the large enthalpy entropy changes in matrix contraction.

Yapel using the temperature-jump method to measure the accessibility of the active histidine in chymotrypsin and thus the degree of domain closure was able to explain the effectiveness of small changes in substrate and inhibitors, He found N-acetyl-L-tryptophan a strong competitive inhibitor was bound tightly with a large overshoot along the domain-closure coordinate. N-acetyl-D- tryptophan neither inhibitor nor substrate is bound very weakly with no overshoot. Its off rate is five

orders of magnitude faster than that of its enantiomer.. Response to indole is very similar to that of the L acid and not at all like the D acid demonstrating the importance of the sidechain as trigger and domain closure as the basis of stereochemistry specificity. The true substrate in this family is N-acetyl-L-tryptophan ester or amide demonstrating on- and off- constants characteristic of domain closure but no overshoot. It is hydrolyzed at points along the contraction process well before reaching the extreme limits set by good inhibitors. The differences in behavior of this small related collection of substrate, inhibitor and passive sibling is built into chymotrypsin but not by variations in passive binding to a small collection of contract groups. Instead it is the ability to accept and utilize the potential energy released by the matrices that determines behavior. The potential-energy rearrangements are vector quantities rather than the scalars of small-molecule chemistry, a discovery probably the basis for all discoveries that followed. There is a also a time requirement. As judged by the small number of dielectric-dispersion data matrix fluctuations in the expanded state have a period near 1 ns a time estimate supported by fluorescence and EPR behavior. Good substrates must be able not only to favor major transfer but also to utilize as much as possible as quickly as possible or be trapped like pepstatin by the overshoot and slow return.. The greater the degree of domain closure, the greater the amount of potential energy transferred and the greater the mechanical distortion of the reacting assembly of substrate and the functional groups. Decay through the transition state to products must be fast with respect to the off constant but a small supporting thermal activation can occur many times in the period of a domain closure.

Workers other than Yapel using less effective ligands with chymotrypsin and pepsin detected varying degrees of domain closure with smaller proton delocalization and weaker binding. Bone et al using the related protease α-lyctic protease measured B factor averages for the boronic-acid acyl derivatives varying the sidechain in the same series as ester substrates so that the catalytic rate constants of the latter could be compared with the mean B values. The relationship was roughly linear and becomes more so after removing the knot atoms to average only the matrices The changes are large, close to the maximum possible for the best substrate, N-acetyl-L-phenylalanine. Thus the sidechains change rate and specificity by changing the degree of domain closure. Parker and coworkers measured optical rotations and ellipticities for a series of acy-intermediates for chymotrypsin the best known being di-isopropyl phosphoryl chymotrypsin. They found molar rotations of 10^{6} deg/cm ^{2}/decimal in the peptide bands near 200nm with an error estimated about 50% These remarkably large values are now understood to be a result of the cooperation of many peptide chronophers of the matrices. Like uv spectra, fluorescence, phosphorescence, proton-exchange rates; inhibitor binding, turn-over numbers for ester hydrolysis the rotations and ellipticities exhibit compensation temperatures near 450K demonstrating their origin in the matrix process,

Like the reaction of chymotrypsin with di-isopropylfluorophosphate the other acy derivatives Parker et al studied react irreversibly because hydrolysis is blocked by space limitations. Acylation by the acid groups of good ester and amide substrates would be similarity restricted were it not for the catalyzed conversion to products. That is, the catalytic cycle is completed by the chemical conversion, another major success in evolution.

A slightly more complicated use of domain closure is illustrated by superoxide dismutase in which each domain has a metal ion the pair being about 35Å apart. At one metal ion oxidation takes place and at the other reduction but probably without actual migrations of electrons. It works because domain closure transfers the free-energy change thus restoring balance in the domains. Electron migration through such large separations may also occur but relative to the nanosecond times of domain closure they are likely to have low probability. In even more complicated enzymes the use of domain closure is multiplied in matched functional domains one pair for each function making possible free energy redistribution, switching through associated enzymes, simultaneous activator, inhibitor substrate processes and so on.

V. From the native to denatured state

a. Bubble as melting product at ambient and low temperatures.

Knot formation is very cooperative so stability loss requires only small expansion of the knots to the point where cooperativity is lost. The total melting process in pure water follows the two-state model except below about 280K and the single transition state is just the conformation at which the cooperativity is lost. However the transition state is a real intermediate in this case structurally very similar to the native state and very little like the motile bubble product. The volume increase in bubble formation is small, not larger than a factor of 2 at ambient temperature increasing slightly with increasing temperature. In pure water in the absence of denaturing cosolvents there are no random-coil conformers above about 280K. It is not clear whether or not there is a so-called "molten-globule" intermediate between bubble and transition state because the bubble state is usually omitted in the assumed folding process. Furthermore most, perhaps all enzymes contain at least one matched pair of nearly independent proteins often of different stability one melting before the other.

The large surface area to volume ratios of proteins require many contacts of surface groups with water mostly as Teeter et al find with crambin, clustered as structure makers that tie up normal water in Frank-Evans icebergs. That clathrate-forming process is favored at lower temperatures as is formation of the bubble product of denaturation so that it has been difficult to tell which conformational state dominates at temperatures below 273K.

b. Melting of native states

Melting rates measure cooperative expansion to the single transition state. The melting rate has been the starting point for understanding melting quantitatively

because as is required by Pohl's compensation plot, for mesophiles the ratio of the numbers of residues in matrices to those in knots is a fixed number. The slope of the compensation plot for activation enthalpy and entropy of melting is constant at 354K for mesophiles because of that the fixed ratio so the position along the compensation line varies only with protein total size. That has provided several kinds of information applying to all members of that very large fraction of all proteins. For example, Morozov and Morozov found that the physical strength against mechanical distortion as measured by the Young's modulus disappears when the temperature reaches 354K, the characteristic value for knots.. However, single residue substitutions in a single protein family have been found that change activation enthalpy for melting by 100 kcal/M. The physiologically important species for a given family appears to be that of the wild type indicating some special role for that selection in evolution. Such observations have a major implications for DNA and taken with other features of residue selection discussed here provide for the first time a logical path toward a genome science albeit a laborious one if as we find, it depends on protein free-volume management rather than residue conservation.

c. Proteins are very different from familiar polymers

Conformational dynamics is another major source of structural information. Angell distinguished "strong" polymers with free energy surfaces characterized by a few large inter-conformer barriers so that transitions among conformers follow the Arrhenius model, and "fragile" polymers dominated instead by low barriers usually very many in soft proteins giving non-Arrhenius rate behavior. B factors make the distinction easy in any given case. The myoglobin-fold proteins are fragile proteins with complex conformational dynamics as illustrated by Frauenfelder and coworkers. That class has quite different properties from the much more common strong families and has provided a misleading model for the latter.

The domination of strong proteins by a few high free energy barriers makes the folding dynamics relatively simple once the several inter mediates in the process are correctly distinguished. At least for mesophiles evolution has found paths for melting following very closely the two-state model consistent with knot-matrix construction so the matrix expansion-contraction process has been greatly narrowed by selection to resemble very closely a two-state first-order transition.

VI. Thermodynamic considerations in catalysis

a. Activation energies in enzymic catalysis

There are major uncertainties in interpreting apparent activation energies from steady-state enzyme catalysis in part because the processes are already complicated by the two molecular species of water and more so because of Benzinger's heat error ubiquitous and uncorrectable as explained below. The rate improvement provided by enzymic catalysis appears in the reduction of the activation energy relative to some appropriate rate value for the chemical changes in the absence of a catalyst or in

catalysis in homogeneous solution often acid-base catalysis. Within errors in entropy bookkeeping the amount of enthalpy transferred from matrix to reaction site is the measure of enzymic efficiency. The latter is all potential energy since evolution has been able to minimize the associated heat changes so all the enthalpy transferred is free energy. The overall enthalpy change is nearly zero so does not appear in the apparent activation enthalpy or the activation free energy as the latter is formally computed from the rate using Eyring's theory. The activation free energy difference between the expected rate estimated usually by Wolfenden from catalysis in homogeneous solution and the enzyme rate is determined by the enthalpy borrowed from the matrices. There may be an entropy analog driven by initial expansion of the matrices but the bookkeeping to avoid irreversibility in such mechanisms is not obvious..

The upper limit for the enthalpy transfer must be something like 20 x 1.4 kcal since 20 seems to be the largest Wolfenden factor so far. Bolen and coworkers were able to trap 23kcal of the transferred enthalpy in clever experiments by reversible acylation with substrates for chymotrypsin with high ring strain in closed-ring form. The way in which the entropy is managed is complicated by contributions from hydration changes yet to be explained. Unusual features in the entropy balance were discussed by Lumry and Biltonen in relation to the conformational details of mechanism discovered by Yapel (Structure and stability of biological macromolecules 1, Dekker 1969). The latter were the first to show activation of a pretransition state as well as the overshoot on an extended conformational coordinate and slow return found in enzymes like the pepsins. Still missing is a formal reaction mechanism necessary to replace the classical steady-state mechanism of Michaelis and Menten in which transition state excitation is the central feature. The transition-state-stabilization proposal of Pauling also goes out the window though abzymes remain a misleading artifact. The "entactic effect" of Williams and Vallee now seen to arise from random nutcracker oscillations also remains potentially dangerous as an indiscriminate source of free radicals.

b. Benzinger's revision in major uses of thermodynamics

It is still not widely known that Benzinger's discovery in 1967 revealed great weaknesses in the use of enthalpy and entropy changes in isothermal processes. In most systems the experiments necessary for rigorous applications are possible only with crystalline substances in fixed phase states. He observed that both enthalpy and entropy in isothermal processes contain work components, potential energy and the capacity or degeneracy part of the entropy, and quite different heat parts. The latter must exactly balance out of the free energy change because as Carnot showed, it is not possible to obtain work from an (heat) engine when both thermal reservoirs are at the same temperature. As a result although the free-energy change in isothermal (and reversible) processes is reliable, total enthalpy and entropy changes are rarely so. Fortunately in some examples it is possible to use linear-free-energy and

compensation theory both finally explained and made useful by Benzinger's discovery. Melting of proteins is one such example so that both forward and backward rate activation thermodynamic quantities can be assigned this follows from zero activation heat capacity for melting because that requires that the heat part of the activation enthalpy be zero. Then what remains is the energy at OK plus the zero-point vibration energies in the enthalpy and that part of the entropy measuring the degree to which the original potential energy at 0K has been converted to heat. Frank called these the "motive parts" following Carnot because only these contribute to the free-energy changes. In the melting rate they are determined entirely by the expansion to transition state so the standard change can be divided between forward and back processes in melting. It is not often that a transition state can be identified with a true intermediate in an overall process.

c. Hydration puzzles hide important numbers

The thermodynamic implications of hydration are large because the number of water molecules in the hydration shell is large and non-freezing as low as 70 degrees below the normal freezing point. A major complication in the analysis of hydration data is that the expanded matrix and the contracted matrix interact with water very differently. This has been shown by Lüscher et al. comparing tosyl-chymotrypsin, fully contracted, with free chymotrypsin, expanded matrix, and is to be expected from the geometric changes in matrices in the process. Specific details have now been provided by Teeter and coworkers who recently showed with crambin that much of the water is held in clathrate structures forming Frank-Evans icebergs about exposed hydrophobic groups. The thermodynamic contributions from hydration are large and can be estimated but not with much confidence. For example, Fujita and Noda found larger free energy change with dry proteins but smaller standard enthalpy changes. Battistel and Bianchi reported similar results with ribonuclease as have others.

The thermodynamic quantities have been distributed by natural selection among conformational details, charges and hydration to support efficient catalysis with little attention to equilibrium thermodynamics. The contribution of hydration has been especially confusing being confounded by the ad hoc unsupported assumption that the polypeptide is exposed to bulk water. The small-molecules still chosen as models for water-protein interaction are grossly inaccurate even to extent of the signs of the thermodynamic changes. A further error is the historical assumption that hydration effects are attributable to melted states rather than native states. But water also plays a key role in the native state. Despite the large number of water molecules in the first hydration shell and the labializing effect of water binding on matrix conformation the activation heat capacity for the melting rate is zero within small experimental errors. The compensation pattern like the phase-like behavior of the matrix contraction process is a most unlikely success of natural selection and not closely analogous with anything in small-molecule chemistry. The other peculiarities of equilibrium melting such as the large standard heat capacity of so much concern in

Privalov's discussions of melting are attributed to the bubble product. But they are further complicated by failure to note the shift from two-state melting process to three-state process over a short temperature interval at the bottom of the semi-parabolic van't Hoff plot.

REFERENCES –Newest references for this summary document are given here but most references are in the Protein Primer and the published papers listed in the UTILITIES folder on the web site. With a few exceptions the contents of this summary have been published at least once so most references can be found at those places. Note that we have found among a large fraction of the protein included in the Protein DataBase no deviations from the structure and properties discussed in Volume 1. HEW lysozyme may be an exception perhaps one of many yet to appear. It is usually easy to check a new enzyme by plotting its atom B factors against atom number although without some practice the full knot palindrome may escape detection in lower-resolution diffraction data.

A-lyctic protease: R. Bone, D.Frank, C. Kettner and D. Agard, Biochemistry, 28 (1989) 5925,760: R Bone, A. Fujishige, C. Kettner and D. Agard, ibid 30 (1991) 10388.
Crambin hydration: M. Teeter, A. Yamamoto, B. Stec and U. Mohanty. Proc. Natl, Acad. Sci USA, 98 (2001)11242

Some topics for volume 2

(Note: Some chapters of volume 2 or their temporary replacements are also given at the URL)
Adaptation for extreme operating conditions appears to be achieved oin major part by modification of matrices. For example pepsin functions best near pH 2 because each of its many loop ends carries an aspartate or glutamate residue mimi8cing the knot pattern in a shell of these acid residues. High temperature operation requires that the temperature range of the matrix expansion-contraction process is raised well above that of the mesophiles so that at mesophile temperatures matrices are closed reducing catalytic activity but increasing thermal stability. Thus far there are too few high-resolution studies to expand the classes of extremophiles. Most are from archaea proteins for high-temperature or high-salt applications.

Enzyme reactions do not have to be intrinsically reversible as Wyman showed in his analogy to the 'turning wheel" but on the thermal-activation kinetics the deviation does not have to be large and has been ignored in the long history of Michaelis-Menten kinetics. With the nutcracker mechanism important questions about detailed balance and microscopic reversibility arise because of the cyclic mechanism and these because the step from excited pretransition state through the classical transition state

to products may be irreversible. Equilibrium, reversibility and other restrictions generally assumed in smasll-molecule chemistry may not had much influence on evolution.

To a considerable degree nature has substituted its own devices to replace small-molecule thermodynamics at least in rate processes. The close approximation of the matrix expansion-contraction process to a first-order phase transition with adjustable limit states has similar simplicities making linear-response behavior and mean-field approximation useful and probably generally reliable.. The hallmark of protein construction appears to be enthalpy-entropy compensation with somewhat lesser reliability of linear-free-energy behavior. Such behavior is found at all levels in biology due in major part to the discovery that matrix-knot construction makes entropy as useful as enthalpy, vectors replace the scalars of small-molecule chemistry..The free-energy complementarity making biology possible aries from that discovery..

Domain closure as a device has several uses. It establishes L-D stereospecificity automatically always providing the same correltly oriented axes required for Begmann's three point rule. It simjplifies requirements for matching of the two domains forming the catalytic pair in enzymes by the C-2 symmetry and good matching so that a weak side of one domain is in contact with a strong side of the second domain. It maximizes the force exterted in domain closure. It provides major but still incomplete coupling between whole-moilecule rotation and domain closure. In the latter the rotatioinal moments of inertia change which provides coupling and thus susceptibility to mechanical changes in both. Where or not this "Brownian –top" feature has major importance remains to be seen

Even in familiar proteases the nutcracker mechanism has been elaborated for greater efficiency. Thus the acyl-enzyme intermediate stage must be at least a two-stroke mechanism in the sense that the matrix expansion-contraction process must occur twice although not necessarily with the same amplitude. Pyridoxyl-phosphate enzymes appear to require at least three strokes as the cofactor gets tossed from one temporary position to another.. Successful reaction mechanisms must include both making our classical dependence on Michaelis and Menten naïve at best.

Despite major contributions by Luscher and Teeter and their respective coworkers, protein hydration is still largely understood at the pictorial level. Quantitative estimates of the contributions of hydration to domain closure and the reaction profiles of enzymes are missing.

Two-states of water explain the Hofmeister series and generally rationalize protein hydration. At one end of the series the structure-breaking solutes like hydrazine and

urea dominat; at the other are the structure-makers like sulfate ion, PEG, glycerol and other proteins. Except at concentrations well below 0.1 mole fraction in either class there is little normal water. For example cliassical expanations of denaturation by urea and guanidinium chloride in termes of direct interaction between cosolute and protein are probably less important than the competition for normal water.

.The high-wall separating genetics from physiology is the inability to predicts function or even structure for one from the other. Although its removal is likely to be long and tedious the discovery that the detailed construction of proteins is revealed by the B factors at least makes the undertaking possible. In the beginning the bases of the DNA can be mapped with the B factors of the corresponding amino-acid residues. This by-passes the current experimental limitation of DNA study to residue conservation when the both depend on the free-volume description of the protein. Residues themselves and SDM experiments can then provide a reliable foundation for experiment but the complexity which depends on the entire residue description rather point-by-point quantitation is forbidding As a starting point one looks for the basis of the knots in the DNA

Constraints on protein folding though astonishing as consequences of evolutionary discovery to not much limit the variety of possible uses. The special features and utility of the immune proteins, multi-enzyme proteins like the proteosome and larger metabolic systems such as the glycolytic cycle should become increasingly easy to u understand as the total rises. Success depends on accurate B factors and the understanding of their variance so x-ray diffraction retains major importance despite the pictorial nature of coordinate pictures.