First half of Section C of Chapter 9, volume 2 of the Protein Primer (reduced to be withini the limites for attachement to email. Oct. 2007)

Chapter 9, part C Enzyme evolution-Time for a paradigm shift.

**(note for this version on July 26, 2007- I did not intend this chapter to be a summary of Volume 2 of the Protein Primer but have been unable to find energy to do otherwise. My long battle with a malignant bacterium is still in progress. This version will have to do for another month or so. The places where editing is most necessary and marked thus: @. Nearly all of the material in this chapter has been covered in earlier papers to be found on the Protein Primer web site (http://www.chem.umn.edu/groups/lumry) the oldest material in the UTILITIES folder. The summary of volume 1 is a useful preparation for this chapter.
The extraordinary new findings on aqueous solutions due to Koga's work is described in chapters 4 and 5 of volume 2 and not here.**

Several of the major steps necessary to understand protein evolution were reported before Pauling, Corey and Branson and remain little known. For example, Pauling et al following Mirsky and Pauling on the critical role of the hydrogen bond, a concept not generally believed at that time, 1933 (J. Gen. Physiology.), also freeing up the rigid structural proposals of Astbury, Bradbury et al. A similar unfamiliar highlight was the complete thermodynamic and rate profile for melting of chymotrypsinogen reported by Eisenberg and Schwert, the first to show the critical importance of zero heat-capacity of activation for the reversible melting of proteins. The full importance of heat capacity in melting had to wait for Brandts in 1964. By 1938 Anson, Northrup, Kunitz, Mirsky, Sumner, and Stanley had reported the key features of protein construction and

denaturation but the existence of substructures were not clearly revealed until in 1951 by Linderstrøm-lang using his proton-exchange method to find that proteins are not isotropic and so must contain substructures. His proton-exchange data also established the existence and importance of conformational fluctuations but these were assumed to measure reversible melting rather than simple breathing, a major error correct for knots but not correct for matrices. Less well know are the quantitative details of that application showing that proton exchange occurs between protein crystal and water vapor, a matter also much overlooked even in recent years (Hnoyweyj and Reyerson @). They found most sites for exchange did so in short times but about 12% did not exchange in several weeks and when dry, much longer. It is only recently that the last finding has become understandable but only on the basis of a few other discoveries Among the most important of those was the demonstration by Corbett and Roche using chromatography that denaturation of proteins in dilute buffers produces very minor volume increase despite the large increase in motility, a deduction also demonstrated by other workers using other techniques. The heat capacity increase in that melting process reported by Brandts in1964 has provided a tool for better understanding especially when used by Murphy, Privalov and Gill in 1990 to separate the total melting process into two subprocesses. The first is the loss of cooperatively in native species with very minor expansion; the second

is minor expansion into a motile bubble. By comparison of the properties with those found by Pohl in measuring the rates of the first process, it was found that the thermodynamic stability is lost in the first step (Lumry, Biophysical Chemistry 101-102(2002) 81). However the structural changes required major deductions by Rosenberg and coworkers mostly using proton-exchange methodologies. In particular Gregory et al were able to establish that there are three kinds of substructures and probably no more than that. With Rosenberg and Woodward he also demonstrated that the exchange does not take place from the smaller structures of without melting and then only at temperatures above 328K But even more confusing in view of the enthusiasm of protein chemists for "the hydrophobic bond" and "hydrophobic hydration" is the critical role of hydration in all enzyme processes. Those concepts tentatively explained by Kauzmann but never established have become popular with protein crystallographers leading to the idea of a non-polar "core" as the basis of folded stability. Gregory and coworkers demonstrated that such cores are in fact the small, strong substructures of Hnoyweyj and Reyerson and are maintained by hydrogen bonds in low polarizabiity environments rather than dispersion forces. They found that their proton-exchange data generated two linear "enthalpy-entropy plots" with unique temperatures to identify them in proteins of various kinds. The value near 450K is a true protein property arising from the contraction and

expansion of the largest substructure but the 354K value is due entirely to water and not to the smaller substructures, now known as "knots", as had been thought since the time of Pohl. Instead 354K is the temperature at which the negative enthalpy change in the formation of clathrate cages of water about structure-making solutes from sulfate, ion. PEG, amphiphiles and hydrophobes such as the permanent gases becomes positive. Thus at 354K the enthalpy change goes to zero to become positive at still higher temperatures. That change makes hydrophobic hydration unfavorable; water and oil actually interact favorably below 354K explaining why protein hydration is favorable to protein folding. It might be expected then that proteins are unstable above 354K but by tuning adjustments in residues and surfaces it has been possible to find proteins stable in Yellowstone hot springs possibly up to 373K perhaps even more although temperature-stability experiments are sometimes confused by slow rates of denaturation. The lower temperature limit for stable folding is again controlled by water properties this time because the favorable hydration of structure-makes with the lower-temperature species of water becomes too large to be offset by knot strength. So between cold denaturation and hot denaturation protein evolution has had to work in a range not much over 80° wide, a considerable limit to Henderson's "fitness of the environment and one not yet apparent to NASA.

As discussed above an additional problem faced by evolution is the heat-capacity limitation discovered by Pohl for the activation heat capacity of the knot disruption rate process and by Gregory for the contraction-expansion process along the conformational coordinate in enzymic catalysis. The former is a consequence of the 354K limitation due to water. The latter is essential for matrix involvement in function because movement along the conformation coordinate must be isoergonic, i.e. reversible. If not, proteases, for example, would quickly eat themselves.

Discovery of the molecular substructures.

Pauling and Corey worked only with small amino-acid structures expecting correctly as it turned out that angle and atom separation must give correct estimates for polypeptide polymers. But of even greater importance is the idea that polypeptides can be assembled and conformationally adjusted into quite different forms. That crucial requirement for protein evolution has proved to be too novel for most protein investigators. The qualitative differences between knots and matrices are just as difficult to comprehend as the B factors. The concept nevertheless set the stage for Watson and Crick and that crucial discovery attracted enthusiasm for the diffraction method already pioneered by Hodgkin, Perutz and Kendrew. But diffraction studies were and are still deceptively convincing so that much time and effort has been wasted using only

the pictures from diffraction experiments trying to turn genome research into a real science

A major purpose of this chapter is to evaluate the relative importance of chemistry and mechanics in biological evolution. That requires an enumeration of the discoveries that describe evolution for which we can begin with material on the trypsin family of proteases given in chapter 3 of Volume 1 but also first use the elegant single example provided by the G protein from streptococcus discussed by Cornilescu, Bax et al for H-bond covalency, Alexander and coworkers for melting thermodynamics and then the integration of those findings by Lumry in Biophysical Chemistry 2002 (Copied into the Utilities folder to be used as a reference protein throughout Volume 1 of the Protein Primer.). Figure 1 recopied below from the last reference explains why the term "elegance" is appropriate
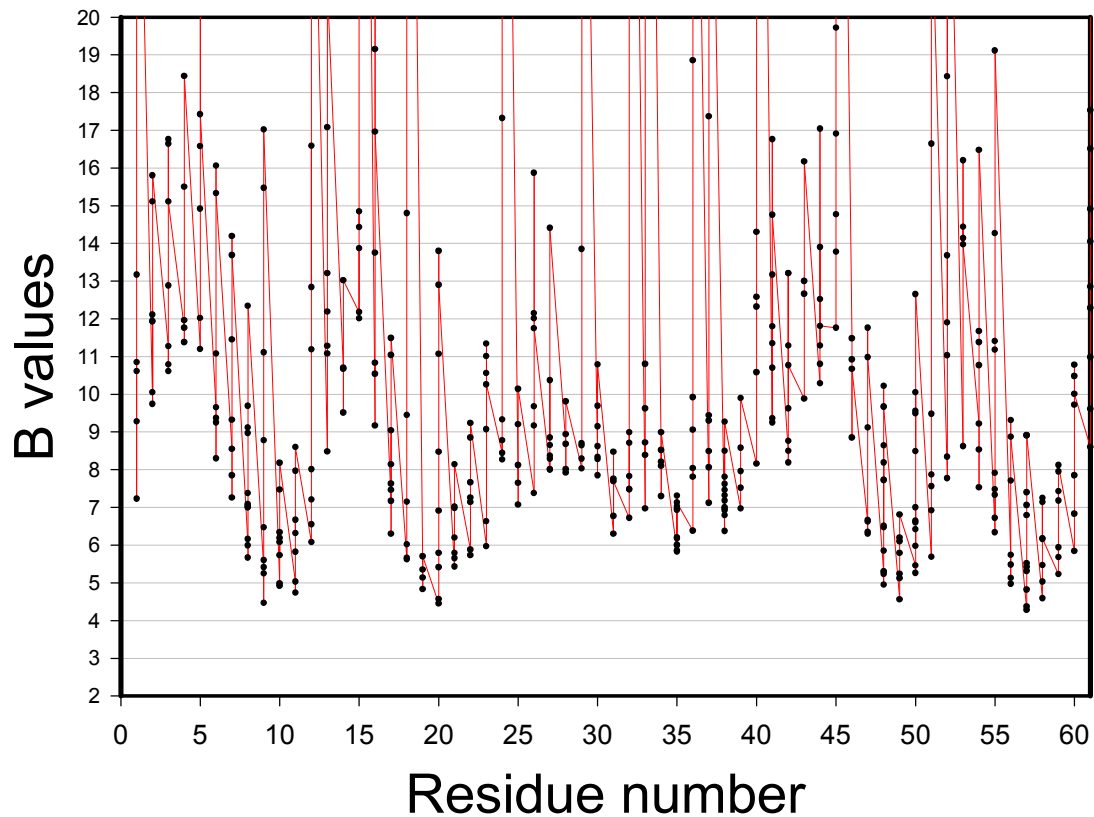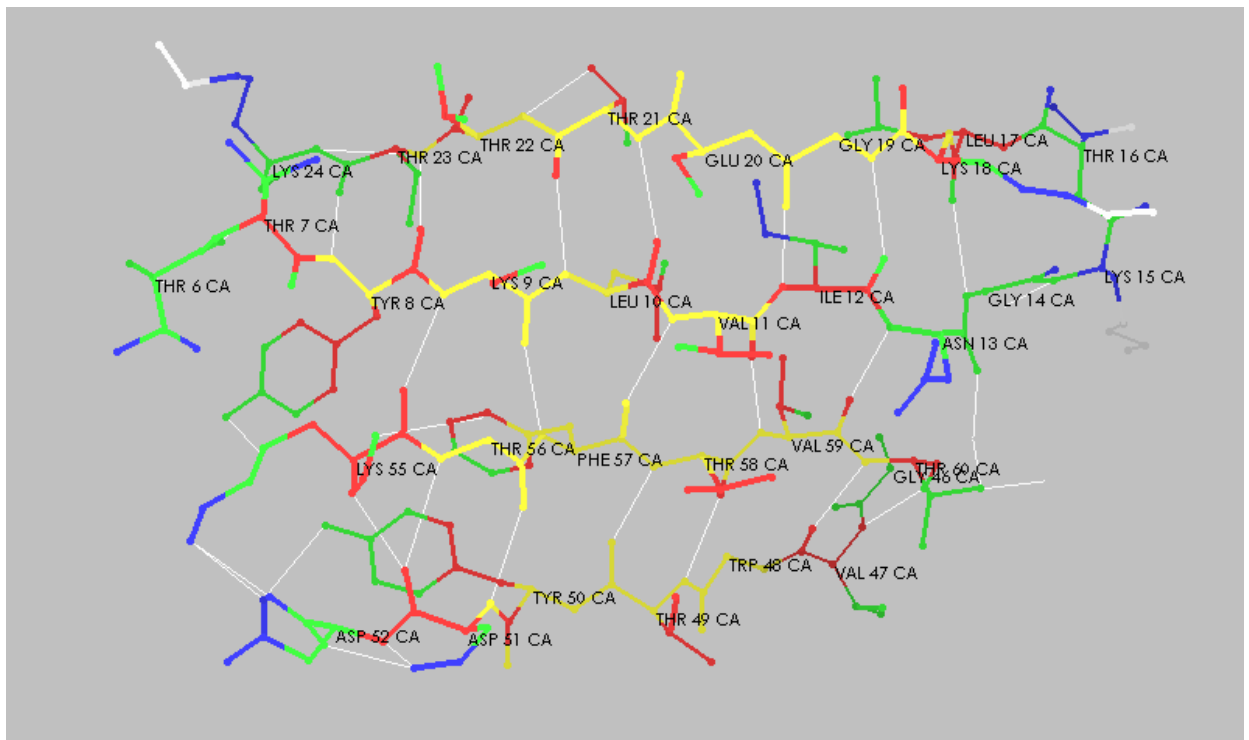
Figure 1 G protein from Streptococcus to show fan organization of the B factors and illustrate their knot palindrome pattern (lowest B factors) which is almost perfect in B factors though rarely in residue matching.. Corresponding atoms and residues from the two cusps of the B-factor palindrome have identical root mean square values exact to at least 0.05Å.The locus of lowest B values includes the α-carbon atoms from which the remainders of the residues form the fan-shaped arrangement in this figure. That arrangement is very common, almost ubiquitous in enzymes because it allows contraction of the matrices toward their knots with little obstruction from neighboring regions. Exact division of the total molecule into two functional domains arranged palindromically is usually immediately obvious as C-2 symmetry in the pictures from x-ray diffraction. The protein assemblies grow for more complex functions by adding pairs of new protein subunits

usually also matched in free volume rather than residue sequence. Aspartate transcarbamoylase is an example, fatty-acid synthetase is another but now with some trigonal symmetry. The last figure in this chapter shows how the palindrome pattern has been maintained in large organelles..

In enzymes the catalytic domains contact with preservation of knot symmetry releasing matrix potential energy to the chemically reacting assembly about the substrates. In this way the ground-state energy of the pretransition state is raised to facilitate nutcracker operation. For this protein which is not an enzymes, this redistribution of potential energy enables the protein to be a highly selective grasping device attaching the streptococcus to a specific antibody (vide infra).

Major protein characteristics:

Figure 2. G protein from streptococcus, IGG binding section. 61 residues. 1IGD pdb (or edited 2IGD.pdb) 1.1 angstrom resolution. Cornilescu Frank, Clore, Gronenborn and Bax JACS 121, 6275 (1999). Most of the hydrogen bonds between yellow atoms according to these authors have high through-bond j factors indicating some covalency. As discussed below for azurin, matrix contraction probably produces further contraction and larger covalency. According to Zundel and coworkers matrix contraction in pepsin resulting from binding of pepstatin actually drives the proton across the inter-aspartate hydrogen bond activating the chemistry.

1. The structural unit of proteins, at least the soluble ones for which there is adequate information, consists of three substructures: 12-17% forms a tightly packed structure responsible for genetic continuity and most of the thermodynamic stability of the native species, 80% with variable packing in an intrinsically unstable state maintained by the smaller substructures thus able to expand and contract as a basis for many physiological mechanisms,. and the interface between protein surroundings providing additional thermodynamic stability varied by adjustment of its composition including as well as water and associated macromolecules acid and base groups, sulfate ions, ion pairs, cysteine and cysteine, phosphate. The first is called *a knot* because it has to be untied before further unfolding can occur. The second is called *the matrix* because it is adjustable to form the moving part of a protein machine. The knots define a protein family with quantitative variations from surfaces and some accommodation to the requirements for operation of the matrices. The three-subunits form what has been called *a functional domain* not be confused with other definitions of protein domains. Particularly misleading is the popular term "hydrophobic bond" because dispersion interactions offered to justify that term are the weakest available once a polypeptide collapses into the soft, motile bubble state. Knot strength and thus folded native stability depend on hydrogen-bonds in a low polarity environment coupled with favorable interface

interactions The bubble states produced by highly cooperative knot disruption have volumes only slightly larger than that of the fully folded forms and remain that way without further unfolding in the 80-degree accessible temperature range..

2. The substructures are described by arrangements of atom free volume so that that rather than residue sequence is the vehicle for conservation and mutation by which evolution of protein families has been achieved. Bubble formation following release from the ribosome is driven by limitations on the amount ot water not held in clathrate formation, the process responsible for favorable interactions between structure-making groups of the polypeptide and water. That interaction is usually called 'hydrophobic hydration' but its dependence on the competition from clathrate formation has only become clear since the recent publications of Koga et al in the last ten years. That water itself has two distinct chemical species as finally fully established by Walrafen in 1983 remains largely unsuspected despite its critical importance. ..

Bubble formation is unspecific so most polypeptides in the absence of structure-breaking cosolvents collapse into bubbles. That step from bubble to knot formation is the critical step and the central success of many centuries of protein evolution. Proteins are often said to be misfolding when minor changes are made in knot residues without total loss of folding. However. in general because of the specificity of knots there is no-folding to native forms rather than misfolding. Real misfolding occurs in formation of the amyloid structures and apparently occurs because those are the thermodynamically stable form of associated polypeptide chains. Knots can look like familiar knots as is illustrated by the two half-hitch knots of the serine proteases, Fig. 3 and chapter 3, volume 1. That one is found in the C-terminal functional domains and is matched by the N-terminal functional domain knot the two related by C-

2 symmetry as is shown in the trypsin plot below.. Despite the absence of any conservation of residue sequence all serine protease have the same knot pair defined by free-volume arrangement without systematic residue conservation.

3. An enzyme family can be defined often somewhat inexactly as having very similar function as describe by substrate and substrate product although there are usually subtle differences resulting from specific selection. The latter are not always small as illustrated by the cavity for the charged side chain of trypsin substrates.

4. Proteins with single functional domains are relatively rare.. Bovine pancreatic trypsin inhibitor is one of the most studied. Its knots, determined originally by Rosenberg's group using the rates of exchange of its amide protons with water, were studied in1993 for the effects of residue exchanges by Kim,. Fuchs and Woodward. They replaced each knot residue one at a time with alanine and found a few of these mutants were still able to fold.. The latter can be said to be misfolded in semantic contrast to those that did not fold. Proton exchange rates below 328K do not require matrix unfolding and are thus closely correlated with the local free volumes at the exchange sites. A consequence of great utility is that because the temperature factors, called B factors by protein crystallographers, measure mean square volume with high precision and are routinely tabulated in reports of diffraction results, they are the best available source of quantitative information about proteins structure and its variation in physiological functions.

The BPTI experiments of Woodward's group should be the model for research on protein structure and function but the precision of B factors has not been appreciated thus limiting the accumulation of information on knot deceptions and matrix dynamics. Those factors are evaluated by using the Debye-Waller theory to give the mean square motion of atom positional

fluctuations in terms of the centers of the atoms. The motion is determined by the scattering of electrons in protein crystallography. Even when evaluated with the isotropic approximation for scattering, the root-mean values for knot atoms are generally precise to 0.05Å. That appears to be determined by the precision of the diffraction methods for proteins bit the absolute values of protein B factors are unreliable because of their dependence on temperature and solvent. However but scale variations can be minimized by comparisons between cusps when the knot palindrome patterns are precise as found in the G protein from streptococcus shown above @.

5. Most of the special devices supported by the evolutionary use of free volume arise because of the several ways functional domains can be combined to utilize the structural and mechanical properties possible with free volume arrangement. The bovine pancreatic protein is often utilized for structural purposes although those uses are not yet well understood. The situation is quite different with the G protein because its two functional domains are arranged about a near-perfect two-fold non-crystallographic C-2 symmetry axis. As shown by their B factors the latter is almost exact for the resulting palindrome of the knots and usually exact for the matrix atom. Not all enzymes have the knot palindrome but all have the two cusps of a palindrome with an exact division of the total atoms and residues of the entire protein into two equal parts. The precision of that division can be determined by comparing the B-factor variances of the two halves or of the two halves of the total knot (Chapter 3@,Volume 2). If there is any pattern relating the residues in the two knot cusps, it is not obvious.. By comparing atoms and residues at equivalent palindrome positions scale factors are minimized and the precision estimate is found to be 0.05Å for knots. That is an upper limit given the experimentally determine B precision and a good estimate of the figure of merit for

construction of the whole protein. This estimate has fair generality despite variation in experimental conditions. It is lower than values reported using other types of data. The estimated based on experimental precision in computed values of atom separations and angles in diffraction experiments have no reliability because there are no sets of the necessary exact values of positional coordinates. Published values obtained in that way can be ten-fold larger than the 0.05Å estimate obtained using enzymes with near perfect palindromes. Atom separation is thus not much larger than found in small molecules constructed using chemical bonds. The all too common use of additives like sulfate ion and PEG to improve diffraction patterns compresses proteins and thus lowers inter-atom distances in matrices. That artificially lowers the estimate as well as suppressing construction detail of the catalytic machine. In the catalytic cycle of the matrices the maximum compression estimated from suitably chosen free and liganded enzymes reaches a few .1Å somewhat smaller than actually occurs..

6. Mean B values for knots have large experimental variability as shown directly by Tilton and coworkers (see Chapter @ this volume) using the diffraction patterns for ribonuclease A over a wide range of temperature . Each functional domain has a single knot formed of 12-17% of the total number of atoms tightly associated usually by hydrogen bonds but assisted by dispersion interactions. It is apparent that tight packing and the low permanent-dipole contributions to the local dielectric constant provide the favorable electrostatic basis of knot stability (Gregory and Lumry, Biopolymer, 1985). There is an obviously important puzzle about the α-sheet secondary structures of Pauling et al because they are not found in enzymes. Thus far only short sections are found at all and those as strong inter-domains connecting devices as in the G protein which has no catalytic function. In the G protein there is a small

section of α-sheet structure as a hinge between the two functional domains. It may be very important as the first appearance of the α-sheet and in a non-enzyme and because its role appears to be to provide major interdomain strength. (fig. @), very recently D. Eisenberg and coworkers report a new zipper-like secondary structure in amyloid filaments, fibers and plaque (See chapter @). Knot atom-atom distances are less susceptible to error in specific function than matrices. In enzymes knots usually provide anvils to apply the force developed in the matrix contraction to the groups participating directly in the chemical changes in substrates. Those chemical groups are not parts of knots and are often found on the main chain just a few residue positions from the positions where they leave the knots.. Main-chains never pass through the functional domains more than once..

Functional changes produced by domain closure can vary with specific function as for example to produce greater distortion in serine proteases on amide substrates than on esters. The famous triad of the serine esterases, his57, ser195 and asp102 is not in knots but asp102 and its C-2 companion thr139 nevertheless have B values as low as those in the knots an illustration that evolution has not been limited to any single construction principle. It is unlikely that the ionization of asp102 is ever chemically involved in the catalytic mechanism or even takes place in the native conformation. The triad attracted attention for the same reason that there has been so much effort toward explaining protein conformations in terms of residue sequence: residue sequences have appeared to be the only possible way to rationalize conformation.

Matrices free of substrates and competitive inhibitors breath in small amplitude fluctuations sufficient to make them permeable on a transient basis to water and the water ions that catalyze the proton-exchange processes.

Breathing seems to be the real success story in evolution. not only well worth the long evolution since most devices making the biology possible depend on matrix breathing but also, as is suggested below, given the chemicals available on and in the earth, the distance to our sun and the temperature ranges there may not have been any alternative.

The attractive but ultimately unprofitable "lock and key" hypothesis of enzyme action finally gains essential detail but only by near-total revision. Originally it was assumed by most enzyme kineticists that if only a sufficiency of diffraction pictures were accumulated some passive protein feature, an advance over small-molecule catalysis would ultimately appear, an unfortunate expectation still dominating much of research on proteins. It has been anticipated to this day that something like acid-base catalysis in familiar small-molecule reactions with the protein providing an electron, an electron plus proton or some other electron or ion as found in metal oxidation-reduction reactions. .It is true that that one enzyme or another employs such a device but only as a peripheral part of mechanism and without something else these do not do the job well enough but each new suggested arising from small molecue chemistry in its turn. has failed. So it is currently most commonly supposed to be stabilization of a transition state by some device special to enzymes that stabilizes that state relative to the prior states presumably by lowering its potential energy. That is still not a bad idea and may occur on occasion but there appears to be only one enzymic catalysis mechanism and that operates by increasing the potential energy of a pretransition state, just the opposite. Furthermore the large potential-energy changes required by either choice are not likely to depend on heat, specifically on rate acceleration by heat. The inadequacies of the latter are all related through Eyring's absolute rate theory in what can be called its "heat-activation version". The common assumption that

their large total thermal energy of proteins gives them an advantage in thermal activation though still popular, ignores the fact that proteins are too soft to act like strong solids in heat fluctuations. Certainly the matrices but very likely also the knots are in independent thermal equilibrium with the supporting medium. Fortunately evolution has been able to construct proteins able to utilize potential-energy fluctuations without loss. The constraints imposed by semi-rigid construction are coordinated to provide relatively large potential-energy fluctuations in the very controlled way necessary to destabilize a pre-transition state without heat production. The model is nevertheless a strong, solid device undergoing a contractions-expansion process like the freezing and melting of a crystal; The analogy with such first-order phase transitions appears to be closer than expected and probably so because the cooperatively in that process is also a product of evolutionary selection. Briefly some major advantages in the selection of mechanical activation over thermal activation are: activation is a vector rather than scalar process, mechanical activation is a real chemical process adjustable with great delicacy by fine tuning using residue selection, the process takes place at constant free energy so enthalpy and entropy compensation allows large adjustments without large changes in free energy, in contrast to heat activation Heisenberg uncertainty limitations are small so much greater rate activation is possible. .Some useful detail is added in the next few paragraphs. (title paragraphs)

Rate activation by heat requires fluctuation in heat distribution that excites the reaction complex along the decomposition coordinate for the transition species. That is a low probability event limited by the number of contributing degrees for heat fluctuations and the Heisenberg uncertainty limit on lifetime of the transition state that is, the length of time the system remains in the transition state. The number of contributing degrees of freedom is determined by the strength of coupling to neighbor groups and is very small except in strong solids and coupling in proteins despite the small distances to neighbor atoms is strong only in H bonds.. The possab9lities for major contributions to stability from contributing degrees of freedom were examined and rejected as early at 1940 for several good reasons. More recently it has been found that coupling of any

local region of a protein to another local region to enhance thermal fluctuation is the same as coupling of the region to the environment including other parts of the protein and the solvent. Thus in general any local region is in independent thermal equilibrium with any other local region but that may not be rigorously correct when both matrices and knots are fully contacted to knot B values Proteins do not have any kind of organized structure for conserving heat energy or for channeling heat to enhance transition-state formation probability but the fan arrangement of the matrices to make contraction possible is essential for nutcracker operation.. Activation using potential energy is a vectorial process adjustable by free-volume experiment. Amounts of transient increase in the energy zero for the pretransition state vary with substrate since substrate binding triggers nutcracker operation. That lifetime is 1ns and a bit longer depending on substrate and temperature but in any case long with respect to the time required for thermal equilibration of the protein.

Heat activation into the rate range of enzymic catalysis if not possible not only because of the uncertainty limitation but also because the heat fluctuations required to produce the local hot spots would destroy the protein. Potential-energy would also be dangerous if produced by irreversible chemical change in the enzyme so nature has had to find reversible conformation changes even perhaps in rare cases using reversible chemistry to make the ustization of potential energy perfectly reversible. So long as that requirement is satisfied, there is no other obvious limit on the amount of potential energy than can migrate to enhance rates but here too is the possibility for destruction of the protein by local distortion. That is a possible limitation but it appears more likely that given matrix sizes finding the right amplitude and direction for application of the required force vector is the important limitation.

Enzymes can use force in compression or tension to accelerate substrate reactions. Hemoglobin to adjust oxygenation isotherms probably use both focused to effect distortions of the heme group through ligands to protein groups. The varieties of adjustment of the electronic properties of the iron-porphyrin group are detailed using cytochrome C (Utilities folder on Protein Primer Web site-hydration and magnetic susceptibility of cytochrome C, J. Phys. Chem. 1961) Compression probably the dominant kind of force in enzymic catalysis but may but alternatives may still be found. However, multistroke operation back and forth along the conformation coordinate can use both tension and compression as illustrated by the two-stroke operation of the serine proteases. also possible in enzymes and multistroke enzyme mechanisms may depend on tension as much as compression. The name "nutcracker mechanism" suggested by Corloni is more appropriate than the "rack mechanism", a name suggested by Eyring and applied by Takashima et al to heme proteins but both are needed so perhaps the term "mechanical activation" is a more accurate choice The direction of movement along the conformational coordinate probably correspond to the signs of the enthalpy and entropy changes but there is a major contribution to those quantities from the interface not yet understood in quantitative detail. ..

Entropy as the sum of heat entropy and the degeneracy entropy arising from the sum over states has been a source of confusion perhaps since Sadi Carnot. Benzinger first reported the error in 1967 (Described in Chapter 2 of Vol. 2). His discovery generalized Carnot's proof that heat cannot be converted to work in any reversible process at constant temperature to show that free energy and thus useful work cannot ever be generated from heat at constant temperature. Essential consequences are that heats factor in the enthalpy cancels the corresponding entropy factor in the free energy. An important but little appreciated consequence is that only part of the enthalpy change and part of the entropy change can be used to do work in any process carried out at constant temperature. Furthermore except for processes of pure crystalline solids, it is not possible to design an experiment to measure the fractions of enthalpy and entropy actually contributing to a free-energy change. The matrix contraction-expansion process produces or uses no heat so has no heat-capacity change It is thus isoergonic and relatively simple despite the complexity of its evolution.

. The transition state is defined classically as the highest point along the enthalpy coordinate describing a rate process but entropy is generally as important as enthalpy in large soft molecules such as proteins so the free-energy coordinate is the correct quantitative description. Sometimes the transition state is positioned by the lowest point of lowest entropy rather than the highest enthalpy. Protein melting rates are limited by the highest enthalpy along that reaction coordinate but the reverse process of knot formation is rate limited by the lowest entropy along that coordinate. Arrhenius did not recognize this possibility but the resulting errors for small-molecule processes are usually small. Eyring and early protein chemists like Eisenberg and Schwert did describe the problem but it is still ignored by most protein chemists. An additional complication may have appeared because an enzyme process has two activation devices, thermal fluctuations and potential-energy redistribution acting in series. but enzymes have been selected to maximize mechanical activation since evolution has found heat activation to be insufficient. Fortunately the two mechanisms can be treated by the same Absolute rate

theory so time can be inserted in the rate law using the same clever trick Eyring used to by-pass the dimensions of the box model for the transition state. Thus his pre-exponential factor, $\varkappa T/h$, can be used with the same degree of error as the use of the equilibrium approximation.
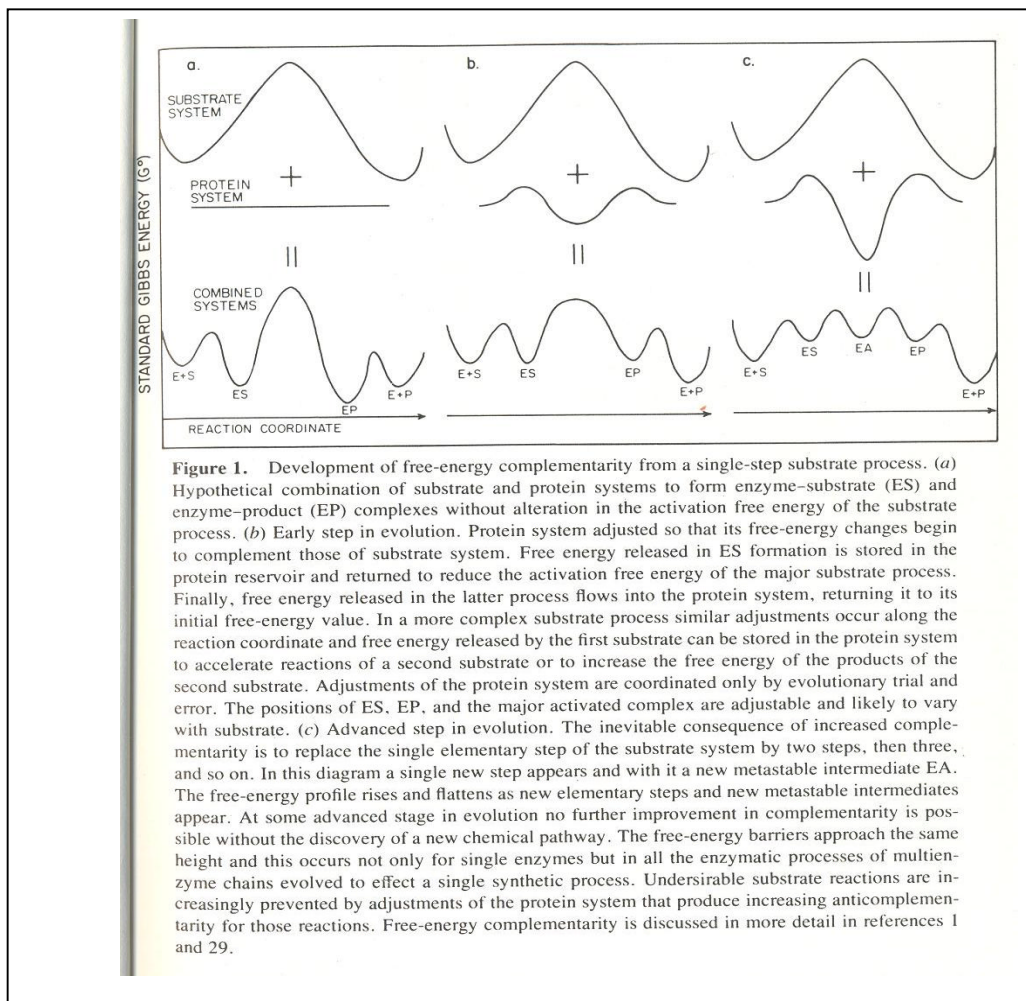
Now reexamination of enzyme kinetics and the experimental activation quantities reveals only a simple modification on the familiar thermal-activation. formalism. As shown in Chapter 7, Volume 2, the experimental activation energy for the so-called "k_{cat}" parameter in the Michaleis-Menten-Briggs-Haldane steady-state reaction scheme for enzymic catalysis measures the contribution of heat activation to the excitation of the transition state but from the pretransition state rather than the equilibrium reactants. The difference between the two is not easily estimated except by analogy with the non-enzymic mechanism. That difference is the enzyme contribution and can be formally treated as such using the thermal activation of the transition state much as it has been used in the past but now as a supplement to the formation of the transition state or the pretransition state. Only the latter is correct and must be used in interpreting thermodynamic data such as shown for the chymotryptic catalysis of the ester substrate family given in chapter 7, Vol. 1 (from Lumry in Volume 2 of Kuby "Study of enzymes"). The heat activation applies to the transition from the pretransition state to the transition state and thus includes such constraints as the uncertainty limitation...

An important source of information about the pretransition state destabilization by potential-energy transfer is provided by the B-factor variations. That is illustrated by the chymotrypsin figures in Chapter 7 and again by the data of Bone, Huber et al for a series of substrates for α-lactic protease. In the latter the substrates were boronic-acid acyl-derivatives with side chains of the ester analogs in the ester-substrate series of the chymotrypsin

illustration. Bone et al found the larger k_{cat} the lower the B value. The lower the B factor, the larger the excitation so variations in both parameters as they alter matrix and knot can be extracted. Once the knots were identified by their B factors the decreases found by Bone and coworkers were largest in the matrix atoms there being zero or small changes in the knot atoms. . Since matrix B factors are directly tied to matrix contraction, Gregory's compensation behavior for proton-exchange rates and now B factors can be related to catalytic efficiency. As already mentioned and as shown in Chapter 7, Volume 2, a more extensive comparison has been made using chymotrypsin with an entire congener family of amino-acid ester substrates. It is even more closely related to matrix compensation behavior as shown by the compensation temperatures for the parent amino-acid ester family computed from the data of Dorovka-Taran et al shown in Fig. @ in thus chapter and discussed in chapter 7). The connection here is that all steady-state parameters except that for the initial combination of enzyme and substrate reveal within small experimental errors the 450-470K compensation temperature that relates the catalytic events to the matrix process. Even the step to products from the excited pretransition state has that temperature, a result found by others in the k_{cat} behavior of chymotrypsin substrates. Thus as suggested above, the positions of pretransition states vary with substrate along the reaction coordinate showing that simple separation of thermal and mechanical activation is not entirely accurate.

The equilibrium approximation may be quantitatively correct for enzyme catalysis because nature has found constructions able to avoid the major problem with mechanical activation. These are found to be quite general for enzymes but also show up in the functions of non-enzymes such as the G protein The first is a consequence of the division into two functional domains

with the same numbers of residues. For enzymes the two domains have equal numbers of residues and equal mass in close approximation to C-2 symmetry there is no residue similarity between nor any obvious requirement for such similarity. The second process is a property of the fan-fold arrangement of the residues along the backbone shown for the G protein above but also ubiquitous in enzymes. The two functional domains in enzymes and many other proteins resemble a haircomb with flexible backbone connecting its two halves at only the middle atom. The sidechains extend from the backbone, usually the palindrome formed by the two knots, to the surface of the ellipsoidal array of the matrix atoms. The first structure is required to make the catalytic process reversible. The second provides an isoergonic process in which potential energy stored in the expanded matrix move to the reaction site That structure contracts and expands as the fan closes and expands not much impeded by the residue sidechains move toward and then away from each other. That is the discovery making mechanical activation possible in enzymology and quite possibly in most other mechanisms for rates and cooperatively in the biosphere. The matrix like a beating heart contracts and expands developing force on the substrate in contraction, and stress on expansion. The process is saltatory but totally cooperative so each protein acts as a complete machine independent of other proteins except for linkage though surface interactions with the environment. The data on polypeptide folding and the matrix contraction-expansion process suggest that most enzymes are so large that each protein can be treated at least approximately as a complete ensemble in the macroscopic mean field of its environment. There is no heat capacity change along the compensation coordinate a remarkable property but one essential for using potential energy from an external source as opposed to some tricky chemistry in the substrate. That requirement does not limit heat

Figure 1. Development of free-energy complementarity from a single-step substrate process. (*a*) Hypothetical combination of substrate and protein systems to form enzyme–substrate (ES) and enzyme–product (EP) complexes without alteration in the activation free energy of the substrate process. (*b*) Early step in evolution. Protein system adjusted so that its free-energy changes begin to complement those of substrate system. Free energy released in ES formation is stored in the protein reservoir and returned to reduce the activation free energy of the major substrate process. Finally, free energy released in the latter process flows into the protein system, returning it to its initial free-energy value. In a more complex substrate process similar adjustments occur along the reaction coordinate and free energy released by the first substrate can be stored in the protein system to accelerate reactions of a second substrate or to increase the free energy of the products of the second substrate. Adjustments of the protein system are coordinated only by evolutionary trial and error. The positions of ES, EP, and the major activated complex are adjustable and likely to vary with substrate. (*c*) Advanced step in evolution. The inevitable consequence of increased complementarity is to replace the single elementary step of the substrate system by two steps, then three, and so on. In this diagram a single new step appears and with it a new metastable intermediate EA. The free-energy profile rises and flattens as new elementary steps and new metastable intermediates appear. At some advanced stage in evolution no further improvement in complementarity is possible without the discovery of a new chemical pathway. The free-energy barriers approach the same height and this occurs not only for single enzymes but in all the enzymatic processes of multienzyme chains evolved to effect a single synthetic process. Undersirable substrate reactions are increasingly prevented by adjustments of the protein system that produce increasing anticomplementarity for those reactions. Free-energy complementarity is discussed in more detail in references 1 and 29.

production in the total catalysis process but the heat developed must be developed as the expense of the substrate and not of the enzyme. Again, that is required by reversibility which in turn is required to prevent destruction of the protein. The figure copied above from "The fluctuating enzyme" edited by Welch for John Wiley ,Chapter @), describes the probable path by which the conformation coordinate for enzymes developed. The matrix process has many important features all apparently critical for operations but not perhaps in the same degree. Nevertheless so far our extensive exploration of the Protein Databank supports the proposal that all enzymes work by this combination of mechanical and thermal activation using the same construction principles. The C-2 symmetry of the catalytic pair of functional domains is unlikely to be

perfect since it is rarely perfect in matrices but matrix conformation changes are constrained by the tethers to the enzyme's knots. The major result is coordinated domain closure which delivers force as nearly-exact vectors transferring the maximum amount of work from matrix contraction to the reaction assembly.

But perfect knot palindrome is not always required (see discussion of T-1 nuclease in the next section) A possible deduction from the mechanical features of function that have already appeared is that delivering the maximum force and potential energy for activation is the test for survival value rather than the C-2 symmetry.

The in enzyme reactions the matrix process generates what has been called "the compensation coordinate" or when known to depend on conformation change "conformation" coordinate. In enzymic catalysis both are appropriate.

The nutcracker. The name "nutcracker" suggested by Carloni is more appropriate name for mechanical activation in enzyme than Eyring's "rack".. The rack name remains suitable for hemoglobin, etc. but not so good when the potential energy is delivered via compression rather than tension. The obvious problem in the evolution of mechanical mechanisms is the *a priori* low probability of finding by the mutation of residues structures able to support domain closure. However the mutation rates have been adequate particularly judging from the age of the earth and they have been greatly supplemented by the several rapid gene diffusion processes already known and especially facile in bacteria and Achaea. At every point mechanical construction depends on the specific environment of a residue, the shape of the residue, the non-polarity of side chains for some residues the permanent polarizabiity. This implies that over time every residue position has been adjusted by random experiment for

its role in survival. Any residue change has been tested for its coupling to every other residue even the most distant since as observed above the contribution of any residue depends on its neighbors more than its own intrinsic chemical and structural features. Obvious a residue also depends on the knots in turn dominated by networks of hydrogen bonds.

Varieties of nutcracker catalytic devices.

The HIV-1 proteases-The construction and operation of the HIV-1 protease is described in the following figure opied from Methods in enzymology, 259, chapter 29.(constructed using data for 3hvp by W@ et al Newer detail in volume 3 of the Protein Primer)
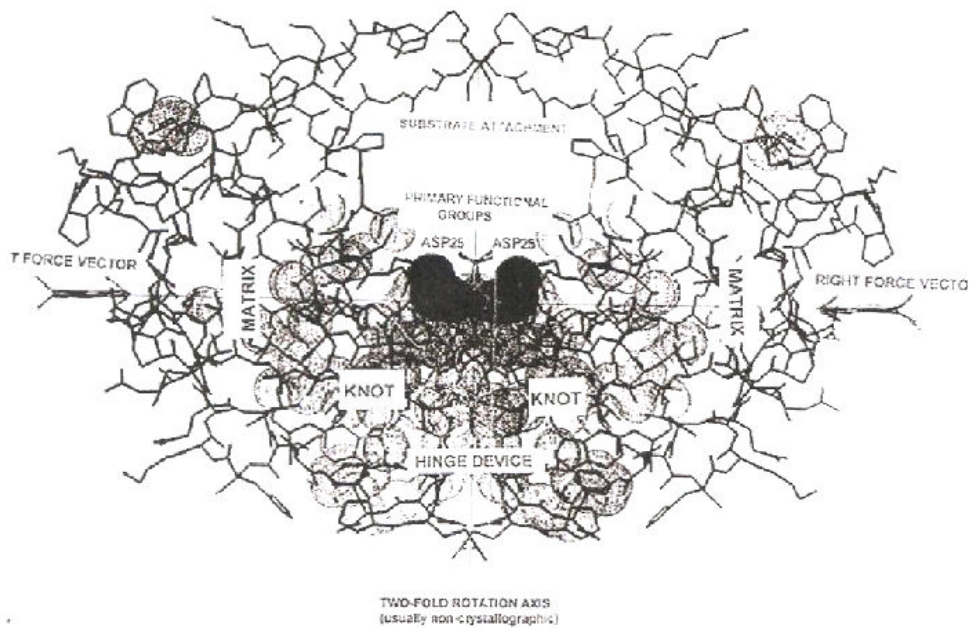

Because the molecule is assembled from two copies of the same structure, the palindrome is obvious in this case with true crystallographic axis vertical in the plane of the page. Similar C-2 symmetry structures are found in many DNA expression regulators. In this protein the hinge formed by short, overlapping β-ribbon at the bottom but a similar but unlocked arrangement at the top probably locks around the substrate and it is that region which has

attracted most attention in attempts to inhibit this protein. The matrices provide domain contracture compressing the two chemically-functional aspartate residues together and into the substrate. The domain closure driving the nutcracker though clearly revealed in diffraction studies is only now beginning to receive attention. As with most other diffraction studies of enzymes the errors in the positional coordinates are assumed considerably smaller than they are. In diffraction studies of small molecules in hard crystals both accuracy and precision are good to a few 0.001Å but protein crystals are too wet and possibly too disordered to make positional errors approach those values so without statistical tests showing the contrary, accuracy in protein studies has been much over estimated. As a consequence most any error in accuracy is as good as any other estimate so prevision estimates bases on multiple measurements on the same crystal have been substituted with the related consequence that the conformation changes in function are not seen. The resulting conundrum is to explain catalytic function in a protein that is assumed to be static and a popular solution is that the chemical groups may be held so close together that the an unusually low barrier for electron and perhaps proton makes easy migration from one group on one of the catalytic domains to the other . Such low barrier hydrogen bonds connecting the two aspartate residues shown in the picture have been proposed for the HIV-1 protease above and between the serine and imidazole groups of the serine proteases. The G protein from streptococcus shown in Fig. @ although not an enzyme is an almost perfect model for enzymes although it has a strong strap connecting the two C-2 matched domains rather than an operating hinge. The fact that that pattern is shown and shown with high precision for all but a few exceptions by the B factors but not by the positional coordinates, atom separation distance and angle, is the major support for the arguments in this

paragraph. However the construction details common to enzymes explain the dynamical basis for "low-barrier hydrogen bonds" without further speculation The rigid-conformation model is inconsistent with abundant evidence that the migrations occur because of domain closure rather than any new electron chemistry. Fortunately the temperature factors reported from protein diffractions studies provide not only an accurate estimate of precision but also reliable estimates of the ranges of conformation changes in function down to 0.05Å which though sufficient to remove the current confusion remains poorly appreciated..

The heme proteins.
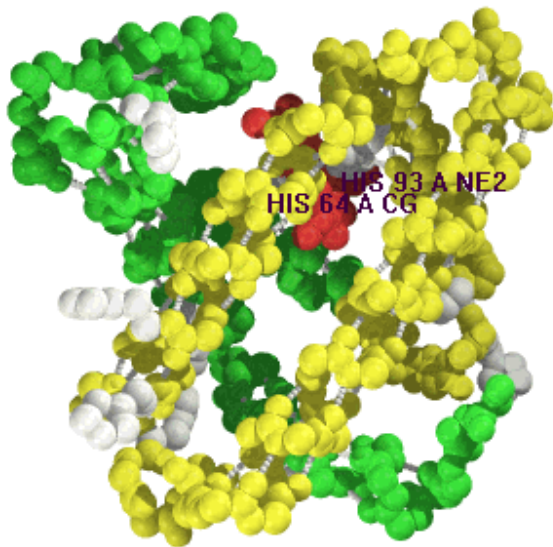
The heme group consisting of a porphyrin base for an iron complex ion embedded in a forest of α-helices forming a palindrome has many uses most of are physiologically important because they are susceptible to adjustment of the

electronic properties of the iron ion by adjustments of the ligands. This device is used in many protein families for catalysis, semi-static electronic and magnetic properties of which the best know is the oxygen binding by myoglobin and hemoglobin. The single "chains" of hemoglobin pictorially closely resemble each other and myoglobin shown below but the latter though physiologically complex is a minor cousin of its four-chain hemoglobin important for oxygenation capacity but not able to support detailed oxygenation isotherms. That is achieved in the four-chain hemoglobins by mechanical coordination of the chains producing the most famous example of Wyman's heme-heme interaction later know as "the allotter effect" The adjustments in iron ligands change the intrinsic oxygen affinity in hemoglobins by orientation of the ligands in angle and length. Those adjustments suggested by the diffraction resuolts of Kendrew and Dickerson in turn suggestion ligand adjiustment by conformation change titled by Eyring at that time as "the rack" because the adjustments of the proximal; and distal imidazole groups appeared to be depended on the variagbioity of the tensionj imposed on the complex ion by the protein. Takashima and Lumry explored this example using dielectric dispersion but the oxygen-dependend changes were not duplicated in other laboratories. Only recently has Takashima found them to be conductance changes observable only at 0.5 MHz and below and their close coordination with the degree of oxygenation apparently reflecting oxygen-control of the dimerization and perhaps higher polymerization of the protein, a different aspect of evolution. The adjustments of ligand coordinates have been detectable in B factors of these proteins but not in positional coordinates from diffraction data but recently Moffatt and coworkers with very high resolution experiments see most of the ligand adjustments without using the B factors. Some adjustments like the partial withdrawal of the ion from the porphyrin

have been accessible to conventional methods because of the large electron density of the ion. Most are too small to be detected at present. The expected effects of the several pssible adjustments have been detailed by Lumry, Solbakken, Sullivan and Reyerson (J. Phy. Chem. 1961 –see Utilities folder in Protein Primer web site). The heme device is very common used for example in cytochrome C, cytochrome C perioxidase, @ for example. In another respiratory protein, the @ of worms, the principle has been retained without the porphyrin by substituting two iron ions for the porphyrin.. Use if adjustment of metal ligand fields has proved to be very widespread in evolution.

Serine proteases
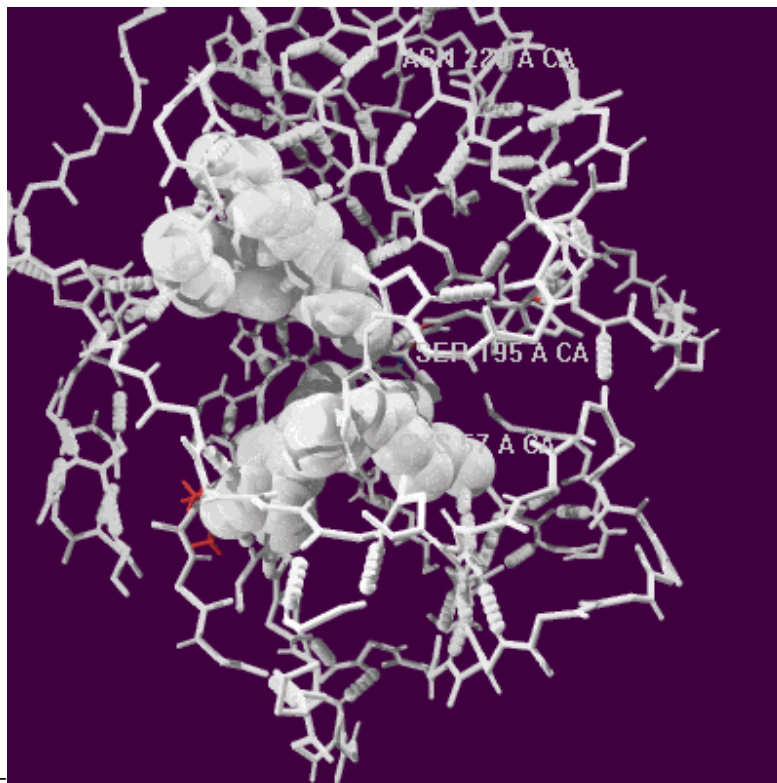
This figure shows the knots of the trypsins also found in chymotrypsin, elastase, thrombin and many other proteases. The knots are parts of β-sheets with extensive interchain H bonding. They are each two-half-hitch arrangements and are positioned relatively in the protein with C-2 symmetry. The functional groups, his57 and ser195 are positioned just where the main chain leaves the knots. Each functional domain presents one of the functional groups (either the hydroxyl groups of ser195 or the imidazole ring of his57) so that on domain closure in the nutcracker operation the two are forced together driving the ser195 proton onto the histidine group mechanically distorting the subsrate and forcing rearrangement about its α-carbon. (Those rearrangements are illustrated by Fig. @ in Chapter @,Volume 1.)