From knots to spider silk?

In explaining how the knots containing only 12% of the residues of a protein provide thermodynamic stability of the native forms the obvious place to start is the electronic variability of amide bonds and the covalence possibility of hydrogen bonds between amide groups closing ring structures in associated polypeptides. The properties of spider silks and man-made polymers like KEVLAR guarantee that that must be a fruitful undertaking even though much of the basic organic chemistry which makes it occur apparently has not been done. The possibilities are discussed here along with some comments on the fitness of the environment for life updating c. J. Hendersons's 100-year old monograph on that subject.

In chapter 12 of volume 1 of this treatise a proposal was advanced to explain the great strength of spider silks and such polymers as KEVLAR. In the last three years additional information particularly on the amyloid structures appearing in Alzheimer's disease that support that proposal have been published. Because of their properties for such things as the KEVLAR armor of much importance in the U.S. invasion of Iraq interest in explaining the properties remains high though progress toward understanding continues to be discouraging. According to our hypothesis the failure at least with the exceptionally strong dragline spider silks arises from the incorrect assumption that protein have isotropic ally properties and isotropic construction. It is a purpose of this chapter to show how the original hypothesis has become more detailed though still not proved.

The term strength is somewhat confusing. As used with spider silks it means mechanical strength in tension but an important connection between thermal stability and mechanical strength is implied consistent with the data. Free

energy and enthalpy are both used to describe quantitatively the thermodynamic stability of native species though only the free energy can be correct. Since enzymes and spider silks are both consequences of successful experiments in free-volume management of polypeptides, it would seem reasonable to use the spider silks as illustrations of just how great the stability of enzymes could be if catalytic efficiency were not the test of evolutionary success for enzymes. The dragline spider silks are stronger in tension than any metal filaments of equal diameter and at least as strong as Kevlar. Like the knots they are formed by arrangement of polypeptides but greatly exceed the knots of enzymes in strength, flexibility and electronic properties. The silks decompose rather than melt on heating and do so only at temperature near 500°C. The are hygroscopic and piezoelectric yet many kinds of spiders can reverse their construction. Thus logically the spider silks are the most promising point of attack on enzyme evolution but that strategy continues to be thwarted by the failure of the structural studies. Once though to be dominated by antiparallel association of backbone threads recently it has been shown that the arrangements are instead parallel..That new information and failure in the face of so much research interest and so much commercial interest is one justification for adding this chapter. The working assumption is that it will be necessary to understand the source of spider silk properties before one can understand how knots are able to maintain folded stability despite being subject to major stress toward unfolding from their associated matrices.

The currently popular "protein folding problem" continues to frustrate many investigators because like the spider silks, it is based on the same false assumption that the construction of proteins is isotropic. Instead as already described in the previous four chapters, they have three very different substructures. In particular the knot substructures that determine folded stability as well as genetic stability remains virtually unknown forty years after their discovery

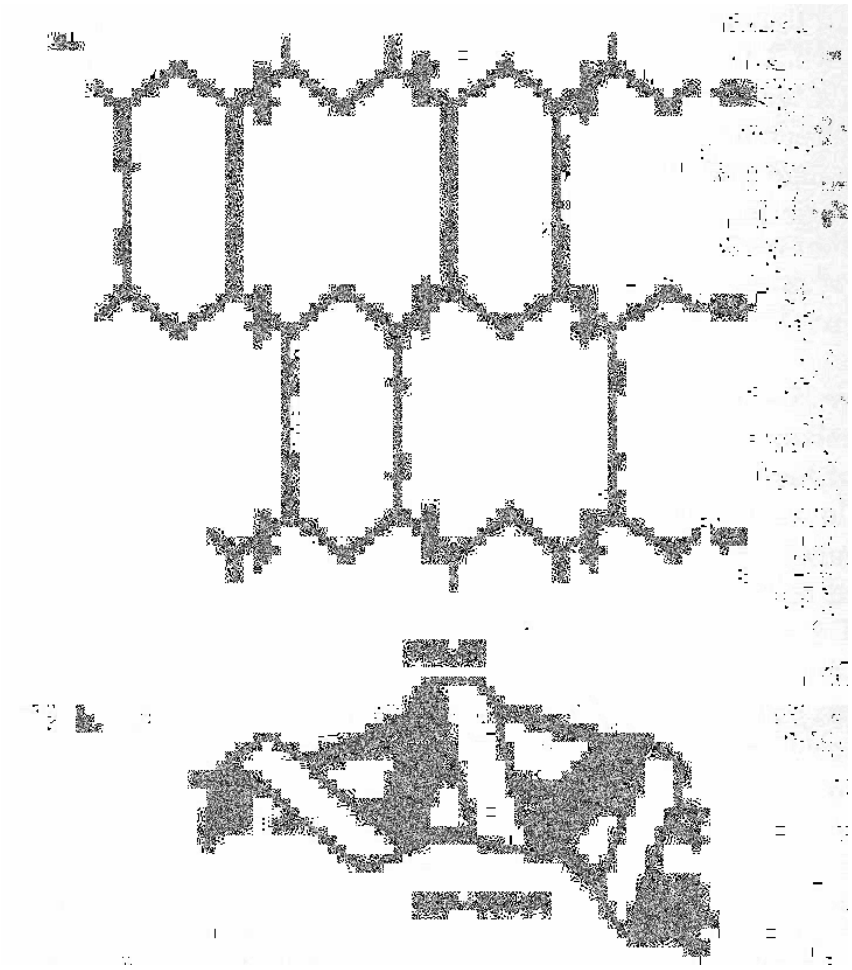
by Gregory and coworkers. In enzymes and probably quite generally in other classes of proteins twelve percent of the residues organized to establish a common knot pattern of atom free-volume arrangement not only provide necessary overall folded stability for enzymes but also the work required on folding to force matrix conformations into useful arrangements. The larger more motile (matrices) parts form amorphous bubbles on release of constraints provided by the knots. .On expansion into this common denatured state they have intrinsic minimum free energies of formation well below the values in their constrained native conformations. In the latter the knot constraints force the matrices to expand into well-organized conformations resembling ellipsoids about knot foci which oscillatre between expanded to contracted limits with periods about 1ns. The hydroxyl or hydrogen ions necessary for proton exchange enter, migrate and leave without unfolding beyond the bubble state..

Nanotubules from polypeptides?

The construction model for knots must have an important similarity to that for the dragline spider silks and as suggested in chapter 12 , volume 1, knots may derive their special properties from the in-plane resonance resembling that of graphite sheets and tubes forming the very strong carbon nanotubules. What is new are reports claiming the major silk substructures consist of parallel rather than antiparallel sheet. Also recently Eisenberg and coworkers find the strong, insoluble amyloid filaments associated with Alzheimer's disease to consist of parallel associations of polypeptide chains.. Those facts taken with the pattern by which protein domains are associated in some non-enzymic proteins can be fashioned into a unifying hypothesis in the following way.

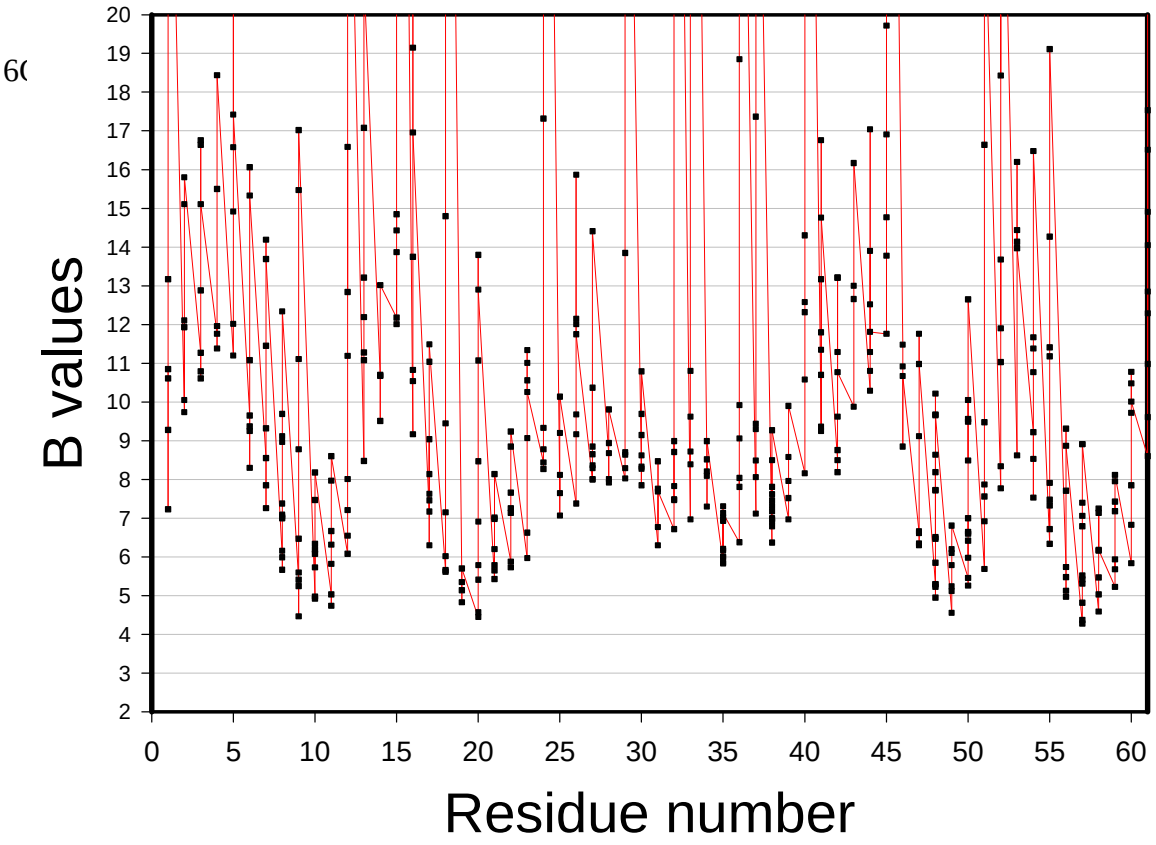
The hypothesis of resonant α-sheet macrostructures.

The proton-exchange studies of Rosenberg's group finally led to the description of two major substructures of enzymes by Gregory and Lumry in 1982. and it became apparent that the major basis of their stability is probably electrostatic arising from cooperation among a number of residues as shown here.. These have little direct relationship with the well-known chemistry of single amide groups because they are enhanced by



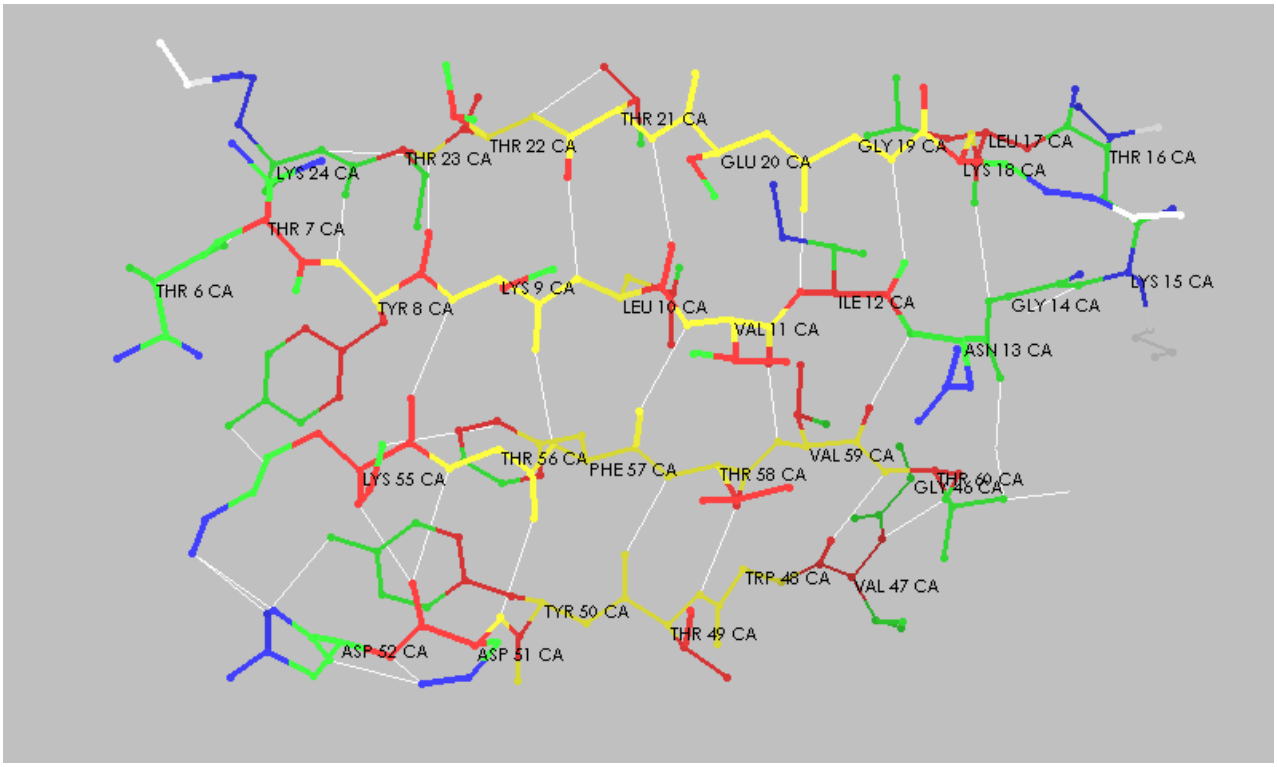
cooperativity. The beta sheet structure (top figure) often found in enzyme knots was explained by high electrostatic compression forcing partial movement of the protons in N-H-O hydrogen bonds from N to O thus adding covalency to the interaction between those atoms (arrows for dipole enhancement in upper figure) in effect producing 3-membered rings closed by those hydrogen bonds. They attributed the helix curvature shown in the lower figure to the lower dielectric constant on the side facing into the protein.

The first figure of the three given below is the B-factor plot for the trypsins which like all trypsins has the same palindrome pattern of the knot array formed by two functional domains. For this protein the knots are short distorted sections of β-sheets locking the backbone thus establishing and stabilizing the folding of the total protein The sections are the major parts of the two knots so they have C-2 pairwise symmetry. Figs. **a** and **b** are copied from Biophysical Chemistry 101(2002) 81 where they were used to diagram the major construction details of the G protein from streptococcus which is reported to attach proteins from that virus to an antibody. The protein although not an enzyme has most of the structural features found universally in enzymes.. The knots one for each domain form a single palindrome with C-2 symmetry about a no crystallographic axis shown in the companion molecular figure, figures b, given below. Pairs of corresponding atoms in the palindrome are matched to 0.05 Å, the characteristic figure of merit for protein construction. These figures illustrate the knot palindrome and the manner in which the residues in the matrices are arranged from knot to periphery to facilitate smooth contraction in enzymes but not known to be part of the function of this non-enzyme. A third figure has been added to show the functional domains to have knots formed from antiparallel association of polypeptide strands through bridging H bonds. What is different from enzymes is the strong connection between the two domains provided by parallel association and connected by a set of strong inter-domain H bonds. The bonding atoms are shown in Chapter 12 of volume 1.reproduced below **c**



Parallel elements such as this one are not found in enzymes. Other methods give somewhat better evidence for delocalization and among these the infer-red spectroscopy experiments of Zundel show that the small but very effective inhibitor of the pepsins, pepstatin, causes complete transfer from one domain to the other of the proton in the bridge between the two catalytic domains. Carey and coworkers using Raman techniques have measured a number of transfers though usually without finding complete transfer of the bridging proton. In the aspartate proteases such as the pepsins and the serine proteases typified by trypsin proton transfer across the single bridge is a measure of the degree of domain closure in the nutcracker mechanism. where they were used to illustrate some common features of proteins responsible for thermodynamic folded stability. The latter include one pair of functional domains matched by distribution of free volume with approximate C-2 symmetry about the midpoint lying ion a non-crystallographic axis, the orientation of most residues is such that the B factors increase from the attachment to the backbone out to the surface. The knots atoms are easily recognized by their very low B values

and comparison of matching pairs of residues is precise to 0.05Å with an overall resolution of 1Å and may be found to be better at higher resolution. The protein with only 61 total residues has high thermal stability as shown in Table @



used to estimate the pH dependence of the thermal-denaturation process. The thermal denaturation rate has not been reported but is likely to fall on Pohl's plot when it is. Even BPTI falls on that plot though it is a single-domain protein used only as a building block. Similarly the matrices may undergo an expansion-contraction process probably by cooling or change in suspending medium. There does not appear to be any of the functional chemical groups such as those supporting electron redistributions in catalysis in enzymes. There is no single-point hinge and the domains are connected by a row of hydrogen bonds (Fig. @ c). The pictorial array of the hydrogen bonds is shown in Fig. b. in which the C-2 axis is shown as horizontal in the plane of the page between the two domains one at the

top and one at the bottom. .Most of the H bonds between the domains are very short; the activation energy for exchange is higher than found in enzymes and the B factors are unusually low. Cornelescu using nmr methods found significant **j** coupling through those bonds indicative of partial covalency.(G.Cornelescu, d. Ramirez,. M. Frank, G. Clore, A. Gronenberg and A,. Bax, J. Am. Chem. Soc. 62 (1999) 6275).. The antiparallel association of the peptide chains as found in many enzyme knots but the array connecting the domains is formed by parallel association and quite different. The G protein has high standard enthalpy and entropy changes in thermal denaturation as also shown by Cornelescu and coworkers (computations are shown in Lumry, Biophysical Chemistry @ suggesting a possible connection between the parallel array and the free energy of thermal denaturation.

The pepstatin results of Davies and coworkers in 1963 revealed the connection between proton delocalization and domain closure but only in the B factors since the precision in length and angle coordinates was too poor (cf. Fig. @ ?). More recently Cornelescu et al showed using the j coupling across the shortest H bonds of the non-enzymic G protein of Streptococcus that they are indeed as short as those reported by Leach for bovine pancreatic trypsin inhibitor, that is, much shorter than found in H bonds between small amide groups or in most interpeptide H bonds in proteins. In particular the short H bonds now contain a significant amount of covalency. The main section of this 61-residue protein is shown in fig. @.illustrates two palindromically related functional groups so generating C-2 symmetry. However, in contrast to the standard enzymic construction the domains are fused by a chain of inter-domain hydrogen bonds.

Most of those bonds have some covalency judging from the work of Cornelescu et al who have also shown the protein to have high stability against thermal denaturation as compared with enzymes. The fusion is effected by parallel association and although helix and β-sheet secondary structures are found to be common in the knots of enzymes and many other protein families, the association of the two knots to make a strong connection between domains in non-enzymic proteins has thus far been found to depend on parallel association of the knots. The implication now tentatively drawn is that parallel association is stronger than antiparallel, the formed never found in enzymes and the latter found whenever very strong association is required. The carbon nanotubule is a promising guide to understanding what nature has found in the proteins so as to use the knowledge of cooperativity among phenyl rings to explain cooperativity in protein secondary structures. In simple terms the in plane strength of graphite is due to adding extra electrons to bonding orbitals with wide delocalization. Adding covalent character to hydrogen bonds apparently extends the familiar resonance character of the C-N backbone bond to add some covalency to the interpeptide hydrogen bond. That allows cooperativity among arrays of the hydrogen-bonded rings gives the opportunity for delocalization around the alpha carbon atoms so the ring structures take on some of the resonance cooperativity that makes graphite strong. But why are parallel sheet structures stronger than anti-parallel sheet structure?

The only suggestion so far obvious is that illustrate in Fig. @ by the G protein.

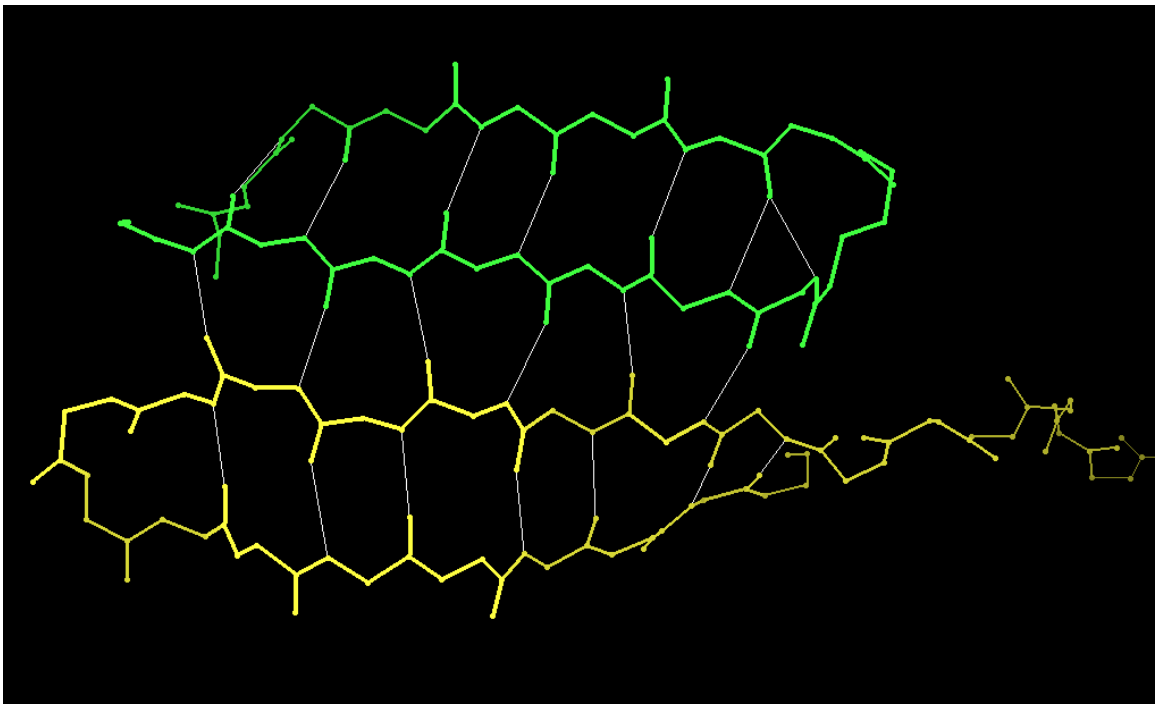
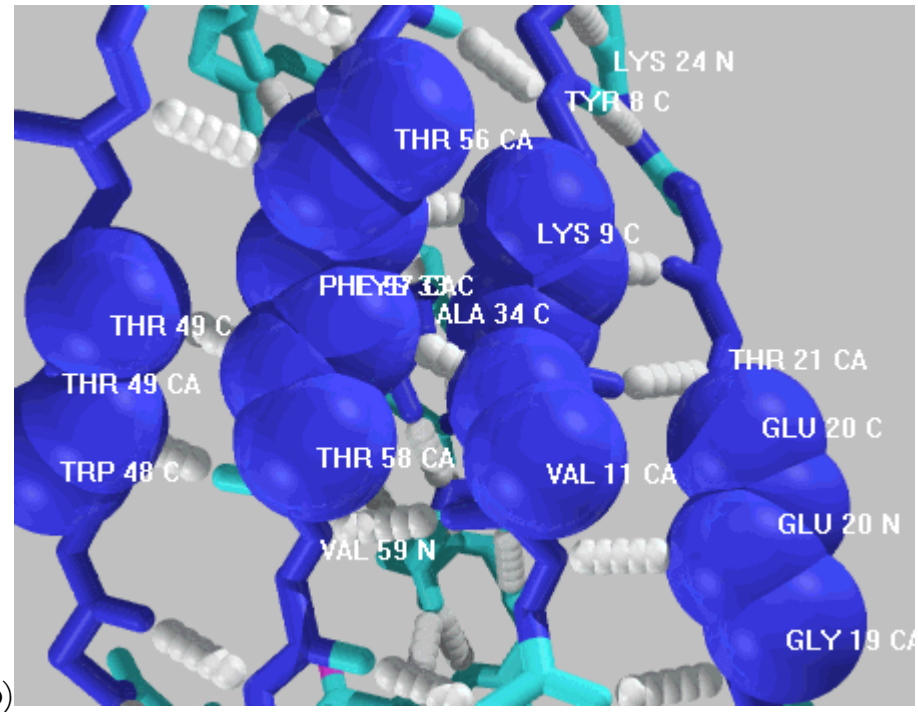


Figure @ The center section of the G protein from streptococcus showing knots stabilized by antiparallel H bonding and fused by parallel H bonding



(1igd.pdb)

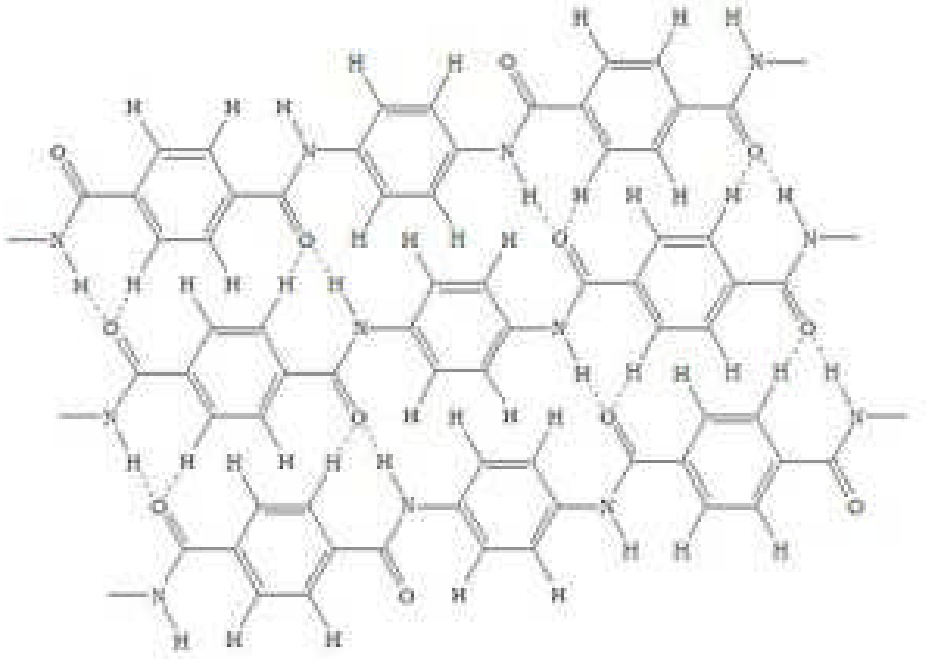
I

The use here of the G protein which is not an enzyme arises from a non-enzyme feature which is a parallel arrangement of polypeptide chains which supports a larger cooperative arrangement in a resonance assembly able to untilized the resonance possibilities in the amide bond much more than previously suspected. The key feature is the linkage through peptide –peptide hydrogen bonds along the lines suggested hy Kreevoy and coworkers. The same approach is indicated by the structures of Kevlar in which phenyl rings are coupled in pairs to eight-membered amide ring structures closes by H bond resonance. As noted above sharing of the proton in such H bonds has been demonstrated for pepsin particularly by Zundel who demonstrating total transfer by other authors. That is an expected consequence of domain closure in the aspartate and serine proteases. That variability coupled with high variability in the –N-H-O- bond between pairs of amide groups offers the possibility for resonance structures in the eight-membered rings formed by antiparallel association of polypeptide chains and in the twelve-membered rings form by parallel association. The figures for these ringe assemblies suggest that the eight-membered rings in antiparallel association are electronically isolated but the larger rings structures formed in parallel association are coupled though the hydrogen bonds shared by fusion of the larger ring structures. Further enhancement of strength my be produced by distortion of matrices and perhaps knots. That possibility has been suggested as a possibler source of enzyme efficiency and it is a promising possibility in adjustment by evolution in nutcracker operation, It is not a surprising possibility using such rearrangement conformation free volume. That distortion can favor cooperative electron displacements in cooperative structures may be obvious but such chemistry seems to have escaped research but stain produced in that way had been proposed for more than fifty years. and could be generalized as a whole-molecule device once the pairs of matched functional domains became apparent. .However

enzymes do not have high thermal stability. In contrast the G protein from streptococcus is higher though it too goes to zero at 354K. In that protein additional stability is due either to that hinge itself or to the assembly of the hinge in parallel association with the two adjacent knots which themselves have the usual antiparallel association. The figure shows two kinds of closed figures that depend for closure on hydrogen bonds within the rings. The antiparallel rings are eight-membered and are stable and strong compared with matrix construction because they are formed in extended arrays apparently for the reasons shown in the first figure. But in the latter although the interchain H bonds and electrostatic attraction due both to neutral dipoles and some charge delocalization, are undoubtedly important, the critical feature is their enhancement by cooperativity. But that cooperativity is limited to short range because the H-bonded rings are isolated though the extent is still guesswork. The possibility of greater enhancement in the structures formed by parallel association producing 12-membered rings each with three internal intermeddle hydrogen bonds is the focus of the current proposal. Parallel association in sheet structures as illustrated for proteins in the figure above is not found in enzymes and is appears to be a consequence of very different peptide chemistry arising from the electronic variability of the amide bond in cooperative arrays. The antiparallel ring aassemblies shown above apprear to have some disadvantage for enzyme action, perhaps making too strong the polypeptide association which would certainly be the case for matrices though not obviously so for knots.

Close association of strands, inter-strand compression and probably distortion produced by residue selection in evolution extends is proposed to enhance the resonance characteristics of the tweve-membered ring assemblies much like the extension over the phenyl rings in graphite. The hypothesis is that the very strong spider silks derive their special properties from the parallel

arrangement of their polypeptide chains. The special resonance properties responsible for the exceptional strength and stability appear to implicate a major covalency in the hydrogen bonds. It seems obvious that spider dragline silk and KEVLAR share an important kind of electronic cooperativity. In the KEVLAR structure here another amide-H-bond assembly is suggested. But the bifurcated H bond has not yet been established the KEVLAR family. It seems likely that that family is united by great strength in tension and then to the in-plane resonance responsible for the strength of



Dotted lines represent potential hydrogen bonding
Carbon - hydrogen and carbon nitrogen bond lengths not exact

graphite structures.

The principle with is dependence on extended cooperativity, amide resonance and mechanical distortion is suggested as the device responsible for the properties of the spider silks and the amyloid fragments responsible for Alzheimer's disease. The weakness of the proposal can be tested by successful research on the chemistry of the arrays to reveal special enhancements due to cooperativity. There is not much research in the area probably because the possibility that cooperativity might generate such unusually chemistry have not yet

attracted attention. As already mentioned, it seems likely that electron redistribution in amides is easily and perhaps dramatically altered by distortion especially in cooperative arrays The proposal depends on the relation between special array properties particularly strength in tension and the parallel versus antiparallel association of the polypeptide chains. Recent work of Eisenberg et al seems to establish the parallel association in the amyloid filaments and plucks but the organization in the dragline spider silks is uncertain both parallel and antiparallel association have been reported in recent years. One result is that the proposal will be tested by the final result of the relevant structural studies when they appear. At present it is very hard to find an alternative and one that as seems necessary, also explains the properties of the Kevlar materials and the spider silks. The dragline spider silks are often disassemble by mother. though not yet by Dupont et al. The amyloid fragments may be similar and again man has not yet been able to do so. The latter have not been disassembled or dissolved without drastic chemical destruction the solubility mirrored in polyalanine and polyglycine filaments assumed to be β-sheet but perhaps not. Notice that the alternative parallel association should assemble auto catalytically because of low solubility and that does seem to be a property of the amyloid structures thought fine details are not yet known. The parallel association in the latter has only been reported in the last few months. Strength in tension may be the only property united the supertrong polypeptide structure into a single superstring family as may or may not become clear then their structures have been more fully elucidated. There are two details. Is the formation free energy responsible for the real or apparent irreversibility and if so, is that why evolution has by passed parallel association in enzymes. It seems obvious that during the folding of a polypeptide on release from the ribosome has at least as probable change to form the parallel association as the antiparallel.and once started parallel association should win out. But I

normal protein synthesis that does not occur and an explanation is necessary. A second area related prions to autocatalysis such as conversion of other prions into the mad-cow variety. As a unimolecular process that seems unlikely almost impossible. One alternative is that growth of the amyloid filaments provides the autocatalytic feature in which case the working hypothesis is that the prions catalyze parallel folding but not antiparallel.

Prions and all that

The peculiar autocatalytic properties of prions require a special explanation and one is suggested by the features of the parallel association of polypeptide chains for several reasons. First is the fact that they are not found in protein signals something very special if not unique. Second, is the considerable strength and insolubility of the amyloid fibers. The second obviously indicates exceptional thermodynamic stability at least for polypeptide polymers. The two together can be responsible for the autocatalytic behavior in polypeptide association if not in prion self catalysis. Growth of antiparallel association is not autocatalytic and easily reversible as demonstrated by the easy melting of proteins but parallel association must be more so for these two reasons. Presumably parallel association has its highest probability of occurrence just at the chains leave the ribosome so high rates of protein synthesis, temperature and solvent factors unfavorable to the normal bubble formation in folding and genetic features not yet easily named but not an unnatural consequence of protein metabolism. The process once started in a given chain has relatively little chance of stopping. Such parallel filaments may generally require proteolysis to decompose. The same considerations can explain the high strength and stability of dragline spider silk which often can be reversed by mother spider but simply melt only a such high temperatures that they decompose first.

An obvious conjecture arising from those arguments is that prions favor the formation of parallel polypeptide arrays. Prions themselves my association to form parallel arrays of they may be very favorable to parallel rather than antiparallel association of polypeptide chains. It seems likely that both are correct but the evidence is still incomplete. As noted above, evidence for the direction of polypeptide association in amyloid filaments and the spider silks does not yet uncontroversial and complete.

"The fitness of the environment"

One hundred years ago C. J. Henderson published a book on the fitness of the environment noticing how conditions on earth favored life. His list was not very long and can be much extended by adding recent findings on aqueous mixtures for example. Henderson appreciated water as a major factor and might have been quick to understand how much proof of its two chemical species improves the story. The two chemical species of water in approximately equal proportions at 285K make large differences in the solvation of structure breakers and structure makers at just about the mean temperature of the oceans is especially noteworthy given the importance of clathrate hydration of proteins and carbohydrates. Brandts did not explain until 1964 how the two water species were coupled the two protein species responsible for denaturation, something Worley and Klotz were able to exploit as early as 1966. even though proof of the two-state model for water did not become fact before Walrafen in 1985. In fact the complete picture was not sketched until Klotz's 1999 paper (J. Phys. Chem B **1999**, 103, 5910**).**

In chapter 9 following the pathway of enzyme evolution a case has been made that the only path for evolutionary success in the biosphere was and is free-volume arrangement. That pathway appears to be the only one available unique except in the sense that a different group of aminoacids or perhaps other small amphiphiles could have been selected instead of the 21 amino acid that did win

out.. For example, it is unlikely that proteins could have been made and varied by any criterion other than free-volume arrangement and that requires proteins to be large enough to exploit free-volume arrangement by rapid an unlimited mutation once the DNA-protein or RNA-protein team got even crude beginnings.. Chemical characteristics might have served for molecules much smaller than proteins but a given type of residue in one place in a protein is quite different chemically from the same kind in another place. For free-volume management the residues are just blocks with which to build protein houses. Parsimony in enzymes-must be possible so there is a limit on complexity or otherwise how did it come about that all enzymes are constructed to have the same major features and a single common catalytic mechanism as a consequence?. We can wonder how that massive convergence in evolution occurred but without much chance that we will ever know. It is difficult to think of another way to maintain genetic conservation than by the knots thus providing an invariant definition of an enzyme family with catalytic function build in. The enzyme construction allows almost complete offset of entropy change by enthalpy change thus minimizing troublesome changes in free energy in the enzyme itself though not in the substrate reaction. But these have required minimizing heat capacity changes in the catalytic processes to produce isolated thermodynamic systems in which the protein-solvent unit is a complete system with the intrinsic reversibility property required for continuity. That is, each enzyme is an isolated thermodynamic system sensing and responding only to its environment only through the interface. Hydration of the latter is fully as important a source of quantitative properties of the enzymes as the conformation itself.

The pairing of catalytic domains defines an inter-domain axis for substrate binding that makes the L-D stereochemistry a fixed feature of the nutcracker that has almost certainly facilitated evolution particularly noteworthy in view of the

massive convergent evolution that has made a single construction principle ubiquitous.. The chemical changes catalyzed by enzymes have been selected for efficiency at least in so far as such selection is not limited by inability to work around severe complications chemically inherent in the reactants. Given the successes in solving such problems it seems reasonable to suspect that each enzyme can in time become *perfect*. Just what that means remains to be established. It may be defined by zero activation enthalpy for the turn-over rate indicating that the potential energy is sufficient to raise the pretransition state to that of the transition state or it may be tied to the quantitative comparison in substrate selectivity. For example the difference between hydrogen and deuterium atoms of substrates and regulatory small molecules is probably detectable quantitatively is not qualitatively in many specific instances but is there sometimes even greater specificity?

Ambient temperature is a major determinant of native proteins and at least for enzymes it "tunes" the nutcracker usually meaning adjustment of the expansion-contraction process of matrices if for no other pragmatic reason than the requirement for reversibility in catalytic function. Most enzymes appear to be able to adjust to temperature change but that is not certain and the adjusting device is not obvious. Recall that surface and residues change are properties of the residue composition thus change is available only through mutation Some organisms have versions of important enzymes with different residue compositions. Many bacteria quickly adjust to temperature change, even solvent change, by activating some of those standby alternatives but undoubtedly at high cost. It is not surprising that nature accepts the burden of constant temperature.