Chapter 3. Knots

The B factors for bovine pancreatic trypsin inhibitor (BPTI) are plotted against atom number and residue number in Fig. 3-1.

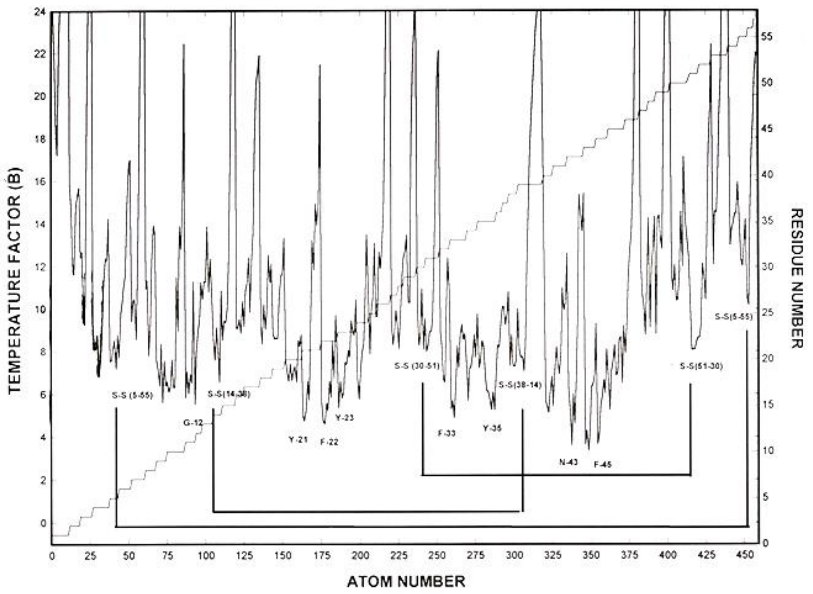

Figure 1. B-value plot for atoms and residues of the Kunitz trypsin inhibitor (BPTI from 4PTI.pdb)

The atoms with lowest B values (peak I) are shown pictorially in Figure 2 and the chain is diagrammed in Figure 4 to show the peptide-peptide hydrogen bonds and disulfide groups that divide the chain into sections and fuse the three low-B pieces. (From fig.2).

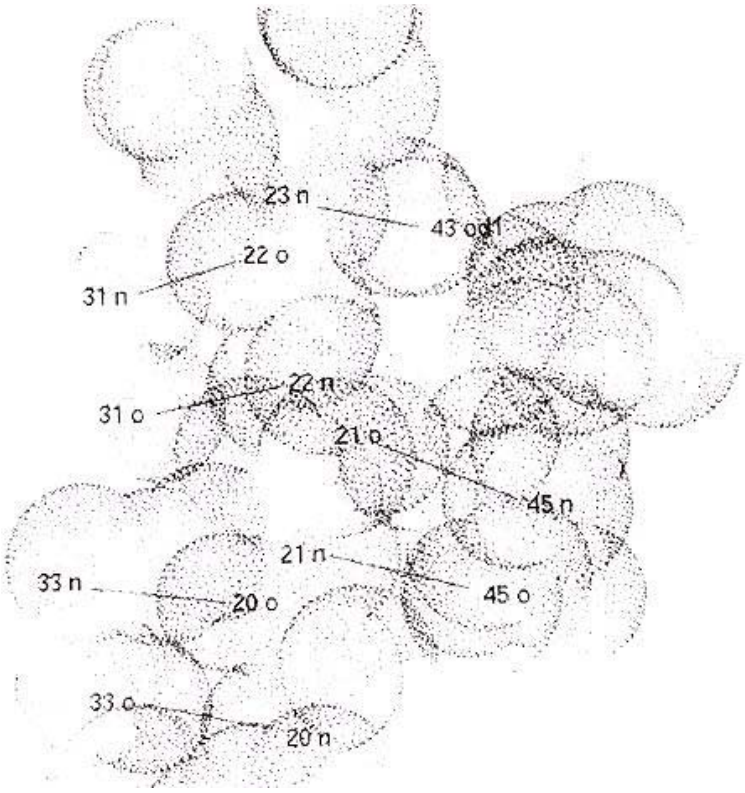

Figure 3-2. The low B atoms of BPTI with the seven hydrogen bonds.

Figure 3. BPTI exploded chain diagram

The three sections united by the hydrogen bonds form a short section of antiparallel β sheet with 7 very short hydrogen bonds in three eight-membered rings and one half-ring. (Fig. 6).

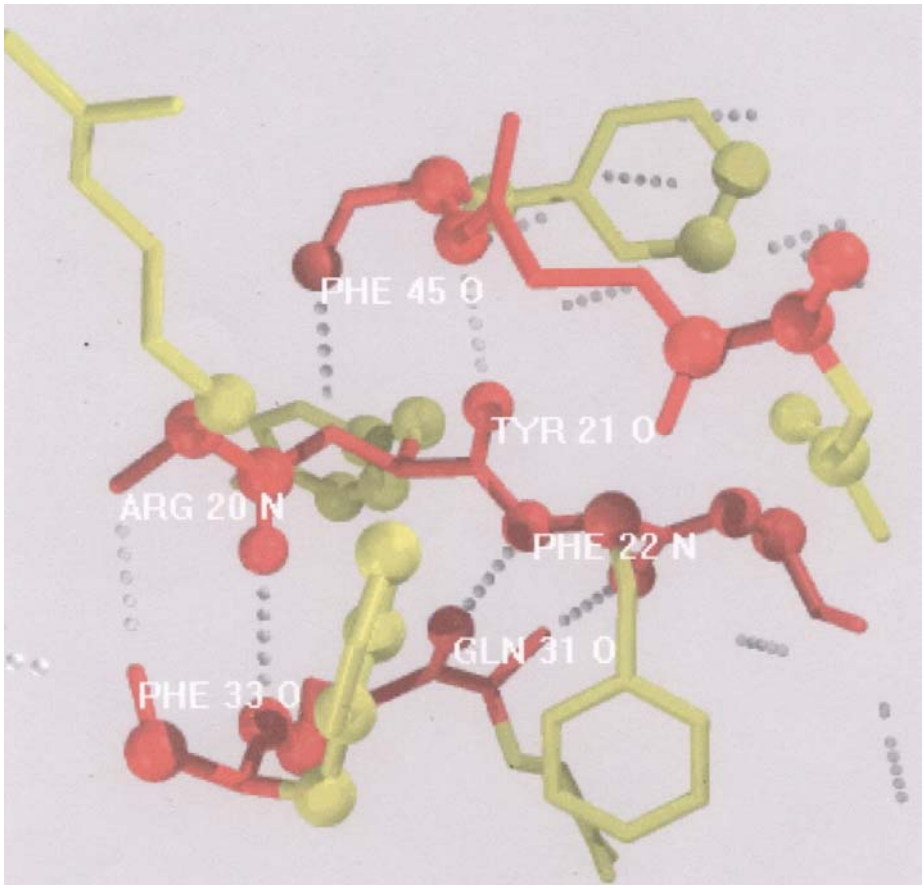

Figure 3-3. The hard, dense core of BPTI is a short section of a 3-chain antiparallel β sheet. Three of the aromatic rings are not constrained by the knot construction and demonstrate fast ring-flipping even at temperatures well below 273K.

Atoms with lowest B values are usually the N and O atoms of special peptide-peptide hydrogen bonds. In this protein there are six of the latter in three eight-membered rings plus one single H bond. These form a short section of a three-chain antiparallel β sheet packed within a shroud of their sidechains, mostly aromatic. The major effect of the sidechains appears to be the immobilization of local area by good packing. In this way the permanent dipolar fields of the H-bonding groups are attenuated almost entirely within the H bond. Whether or not the selection and packing of the knot sidechains enhances the electrostatic cooperativity in other ways remains to be determined but the mutual attenuation of the fields of the dipolar groups is probably sufficient to produce the abnormally short peptide-peptide H bonds Levitt first noticed in

BPTI. The synergism of electrostatic interactions improves with contraction until stopped by the atom-atom repulsion potentials and the tension generated by the remainder of the protein. The decrease in potential energy is large as a cooperative result achieved by great selectivity in reside selection. The domination of protein structures by small numbers of these close-packed substructures, no more than 15% of the total of non-proton atoms, shows that they are major successes in evolution.

In BPTI this highly cooperative substructure supplemented by disulfide groups is responsible for the conservation of the structure of the protein and the stability of the folded native state. Since the substructure must be "untied" to permit disruption of the native species, Gregory and Lumry called it a *knot*. No better name has been suggested in the twenty years since first discovery. That protein knots are not topological knots ceases to be confusing once their molecular description is understood. VELCRO correctly emphasizes the importance of short-range attractive potentials but the name is proprietary. Conservation of knots in single proteins and in families is a result of conservation of a few strong hydrogen bonds rather than conservation of residues. Conservation of atoms in knots rather than residues guarantees preservation of protein skeletons without compromising further evolutionary experiments. Because knots are usefully described by their atoms rather than by whole residue sidechains, knots are not identical with "slow-exchange cores" as used by Woodward and others. (vide infra) but that term has the disadvantage that it is an operational definition inadequate for a general class of substructures

Table3-1. Residue conservation in the knots of proteins related to BPTI.

PROTEIN	5	12	14	21	22	23	30	33	35	37	38	43	45	51	55
BPTI	C	G	C	Y	F	Y	C	F	Y	G	C	N	F	C	C
Bovine spleen inhibitor	C	G	C	Y	F	Y	C	F	Y	G	C	N	F	C	C
bovine serum inhibitor	C	G	C	Y	F	Y	C	F	Y	G	C	N	F	C	C
turtle egg-white inhibitor	C	G	C	Y	P	Y	C	F	Y	G	C	N	F	C	C
Snail inhibitor K	C	G	C	Y	F	Y	C	F	Y	G	C	N	F	C	C
Sea anemone inhibitor	C	G	C	Y	Y	Y	C	F	Y	G	C		F	C	C
Russel's viper inhibitor	C	G	C	I	Y	Y	C	F	Y	G	C	N	F	C	C
Ringhal's cobra inhibitor	C	G	C	F	Y	Y	C	F	Y	G	C	N	F	C	C
Cape cobra inhibitor	C	G	C	F	H	Y	C	F	Y	G	C	N	F	C	C
Black mamba toxin B	C	G	C	F	H	Y	C	F	Y	G	C	N	F	C	C
Black mamba toxin E	C	G	C	F	Y	Y	C	F	Y	G	C	N	F	C	C
Black mamba toxin I	C	G	C	F	Y	Y	C	F	W	G	C	N	F	C	C
Green mamba toxin I	C	G	C	F	Y	Y	C	F	W	G	C	N	F	C	C
Black mamba toxin K	C	G	C	F	Y	Y	C	F	Y	G	C	N	F	C	C
Green mamba toxin K	C	G	C	F	Y	Y	C	F	Y	G	C	N	F	C	C
Banded krait inhibitor	C	G	C	F	Y	Y	C	F	Y	G	C	N	F	C	C
longf-nosed viper CT inhib.	C	G	C	F	Y	Y	C	F	Y	G	C	N	F	C	C
Long-nosed viper Tryp inhib	C	G	C	F	Y	Y	C	F	Y	G	C	N	F	C	C
Silkworm chymotryp inhib	C	G	C	Y	S	Y	C	F	Y	G	C	N	F	C	C
Beta 1-bungarotoxin B chain	C	G	C	F	Y	Y	C	F	Y	G	C	N	F	C	C
Beta 2-bungarotoxin B chain	C	K	C	F	Y	Y	C	F	Y	G	C	N	F	C	C
Inter-alpha Tryp inhibitor bovine	C	G	C	Y	F	Y	C	F	Y	G	C	N	F	C	C
trypstatin	C	G	C	L	A	F	C	F	Y	G	C	N	F	C	C
Human A40751 amyloid protein precursor	C	G	C	W	Y	F	C	F	Y	G	C	N	F	C	C
Lipoprotein-associated coagulation inhib	C	G	C	F	F	F	C	F	Y	G	C	N	F	C	C
LACI(118-188)	C	G	C	Y	F	Y	C	F	Y	G	C	N	F	C	C
LACi(210-280)	C	G	C	F	F	Y	C	F	Y	G	C	N	F	C	C
Alpha3 chain of type VI collagen	C	G	C	W	Y	Y	C	F	Y	G	C	N	F	C	C

Note the tight conservation not only in the knot residues where it is 100% but also in many matrix positions. That us higher than we have found in someother proteins as in the trypsin family illustrated in the next table. However, there are some proteins such as Dennis' collection of the phospholipase A2 family where knot conservation appears to be tighter than usual. Since most genome research is based on residue conservation, a much large variety of protein families must be examined. Insofar as exceptions do prove the rule, exact knot residue conservation is not required.

The number of atoms in a knot is small. The number of very slowly exchanging proton sites was first determined directly by Hnoyewj and Reyerson using H-D exchange between protein crystals and water vapor as about 15 % of the exchangeable protons. That number is consistent with B-value data and the proton-exchange rates for familiar mezophilic proteins.

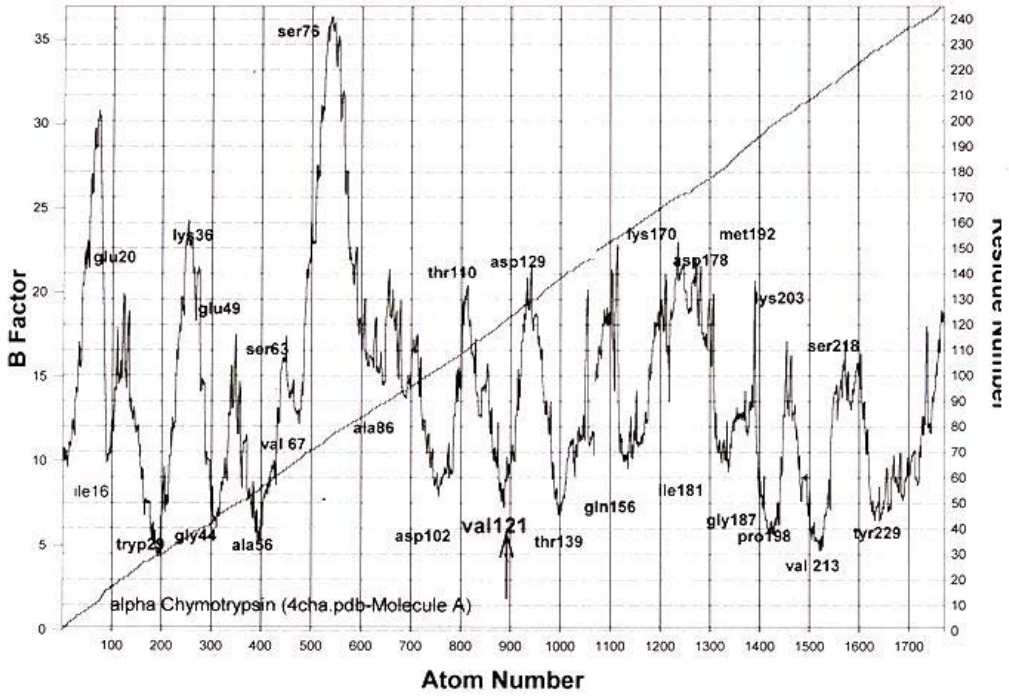

Figure3-4 Atom and residue B factors for the trypsin family of serine proteases. The typical pair of catalytic functional domains is connected at a hinge and by a hydrogen bond between the two functional groups, histidine 57 and serine 195. Otherwise they are an independent pair of complete proteins. Conservation of knot residues is not exact (see following tables).

Table 3-2. Comparison of knot residues in three members of the trypsin family

N-terminal	H-bond	partners			C-terminal			
Residue Number	Chymotrypsino	Trypsin	Elastase		Residue Number	Chymo Trypsin	Trypsin	Elastase
16	Ile	Ile	Val	to	228	Tyr	Tyr	Phe
29	Tryp	Tyr	Ser	to	213	Val	Val	Thr
44	Gly	Gly	Gly	to	198	Pro	Pro	Pro
56	Ala	Ala	Ala	to	186	Ser	Glu	Asp
67	Val	Ala	Ala	to	181	Ile	Phe	Val
85	Ile	Ala	Val	to	156	Gln	Lya	Gln
102	Asp	Asp	Asp	to	139	Thr	Ser	Thr

Table3-2. B-factor palindrome patterns for some serine proteases using lowest B factors (knots) and highest B factors (surfaces). The implicaton to be tested with additional species variations of this family and others suggests that information is not stored in a genome by residue sequence itself but rather in a complicated way which preserves the knots and presumably details of matrix construction. Clues to the latter have not yet become obvious but there is considerable evidence already unearthed showing that knots are preserved among species bur not simple sequence. Thus it is important if the information in a gene is to be revealed that the knots of protein families be compared. That involves simple comparison of B values and sequence positions. Dependence on residue sequence is likely to be as unprofitable as the assumptions of secondary structures have been as reliable guides to domains. Similar variations can be found by comparing species variants of a single enzyme, e.g. the fish trypsins.

BPTI and trypsin have knots formed from atoms widely separated along the polypeptide. Another class of knots is formed from contiguous segments frequently in secondary conformations. Protein G, the IGG binding protein from streptococcus

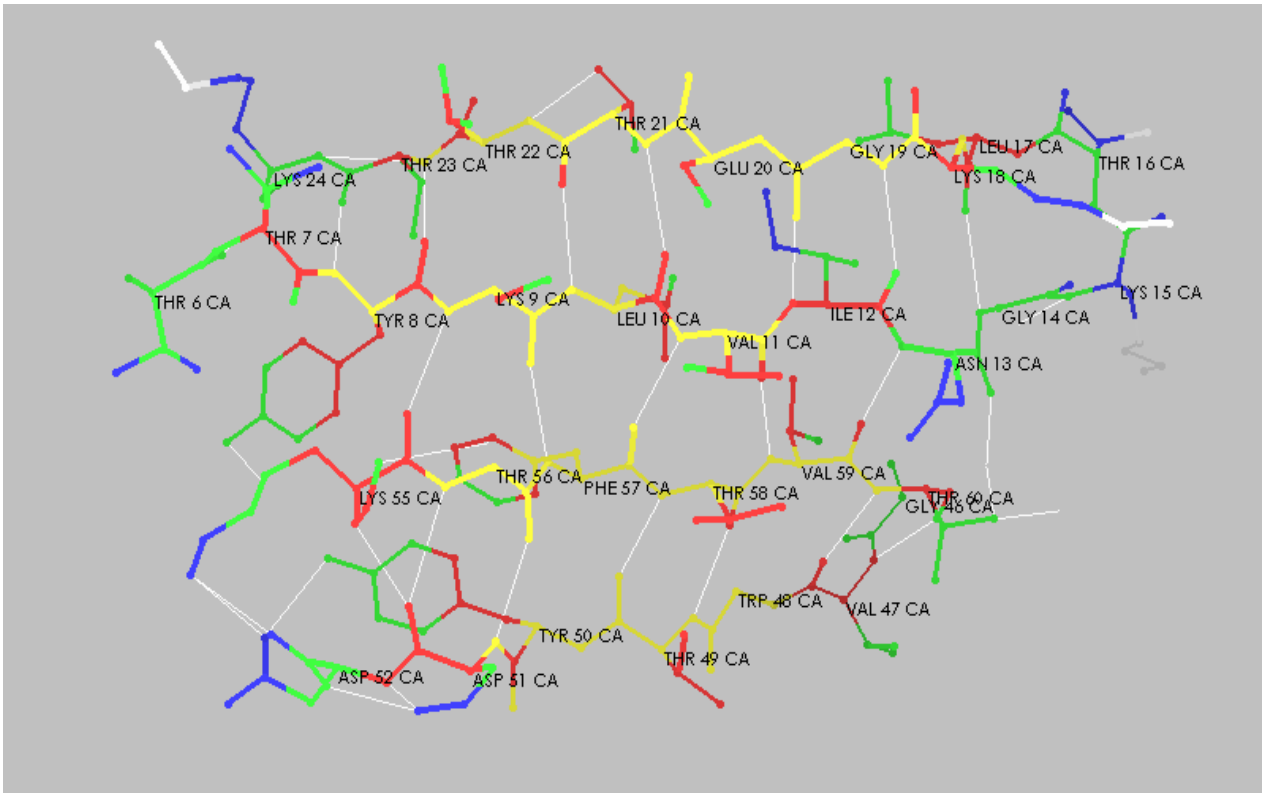

Figure 3-5 Knot-matrix picture of the G protein of staphylococcus showing the fusion of the two knots into a single β-sheet. All peptide-peptide hydrogen bonds in this structure are short with considerable through-bond coupling. Their B factors and proton-exchange rates are low.

Fig.3-5 and 3-6 describe a good example since each of its two knots contains 7 strong peptide-peptide H bonds in antiparallel β-sheet conformation. Of the 464 non-hydrogen atoms 70 are in knots. In this protein in contrast to enzymes the two knots are fused by short H bonds into a single large sheet Knot densities are high and because knot residues are so often aromatic or aliphatic number densities are often also high. The sidechains of residues in knots probably contribute to knot stability by lowering the permanent polarization but the quantitative features of this role are far from clear. Regan and collaborators converted this protein by residue substitutions into a stable helix bundle, a very different structure, by judicious replacement of residues to change the local polarization. (see Section @) In neither protein is it useful to describe the side-chain interactions as "hydrophobic bonding"nor the knots as hydrophobic cores

since neither name accurately describes their role in folded stability. Note once again that neither term correctly describes the role of knots in stability.

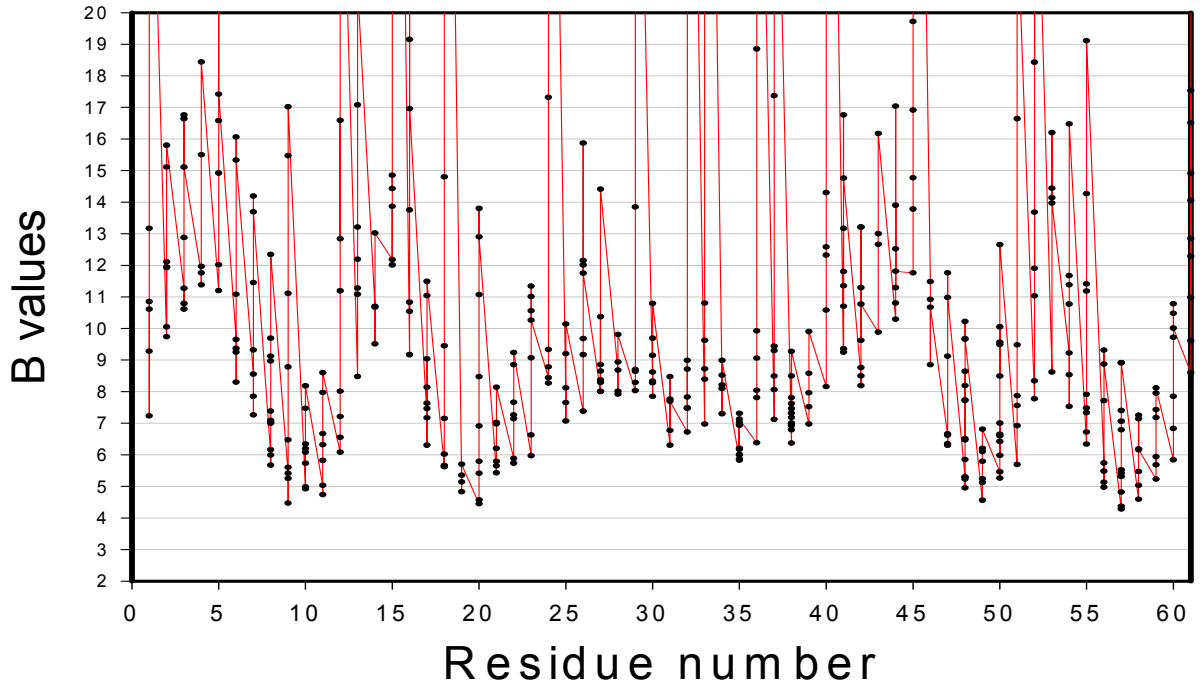

Figure 3-6 Atom B factors for the G protein. The two functional domains are very closely matched in B values and chain positions. The variation of B values within each residue is large showing that the atoms with lowest B values, all in the peptide-peptide hydrogen bonds, generate the B-factor palindrome. Whole-residue matching is closer than in enzymes but still inferior the matching of the hydrogen bonds.

Using the B-factor display in Fig.8 Atom B factors versus atom number for the staphylococcus G protein and atom B factor illustrating precision in B-factor determination and protein construction by comparison of the branches of the palindromic B-factor pattern shown for this protein. Note the wide distribution of the atom B factors in each knot residue. using a graphics program such as MoleImage that emphasize the use of B factors, the relative free-volume at each atom can be displayed so that one can construct the entire free-volume description of a protein. Then comparisons with the atom B factors for matrices in contracted forms of a protein provide descriptions of the expansion-contraction process. (Cf. Fig. Peptstatin Fig. 9)

The streptococcus G protein illustrates the cooperative contraction of peptide-peptide assemblies into large knots one for each of the two funcitional domains of the protein. There are many other examples of large antiparallel β-sheet knots in enzymes as well as of the other features of knot-matrix construction. The knot information revealed by our initial random choice of enzyme examples in the PDB can be found in any collection of high-resolution B-factor data and is indicated whenever any feature is duplicated. In particular the C-2 symmetry of knot pairs is absent in only a few enzyme examples some attributable to error in low-resolution studies but a few others discussed below that show free-volume matching between cusps without C-2 symmetry.. Knot structures are sensitive to experimental conditions but their B factors scale as a group from study to study and protein to protein. Positions and relative B values are unchanged within cusps and betweens cusps. Matrices are more complicated because their expansion-contraction process is undetectable in x-ray-diffraction coordinate values and has to be characterized with B factors and by such methods as proton-exchange rates and circular dichroism. Matrix information from diffraction coordinates complicated because a large fraction of the diffraction studies are on engineered proteins with one or more changes in residues. Such changes can produce important alterations in matrix properties often below the level of detection in x-ray coordinates and the precision level of experimental values of catalytic parameters. Again we must turn to B-factor data for further progress.

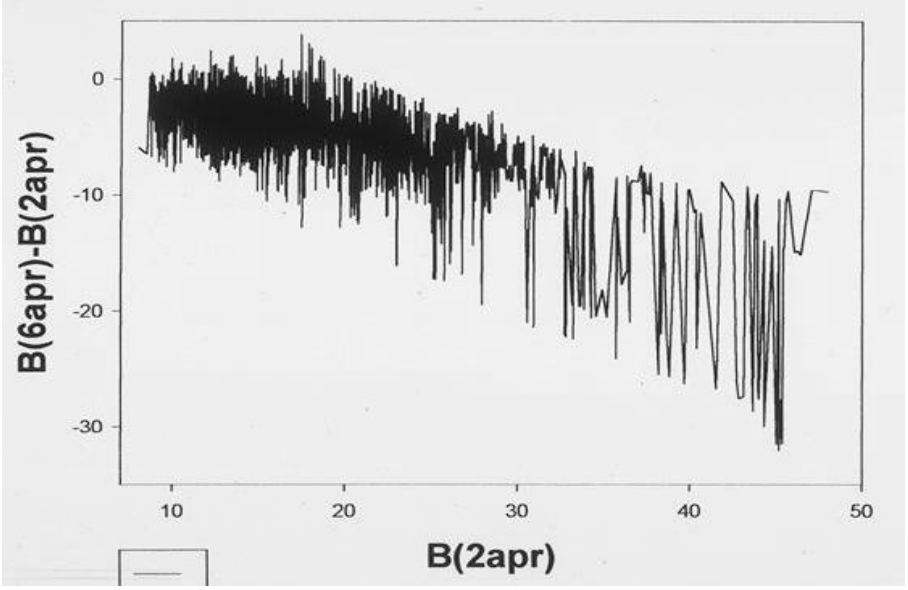

Figure3-7 Contraction of the matrices of rhizopepsin produced by binding one molecule of pepstatin. The knots are unchanged and matrix contraction at each atom is proportional to its B value in the expanded (inhibitor-free) form. The ordinate is inhibited value minus free-protein value. Contractions in B values of similar size are produced by cooling to 100K and below but the plot is linear rather than logarithmic. Free volume changes estimated from the B values are somewhat exaggerated by communal effects not yet estimated. The patterns of change in B values contain much information . For example the oscillatory pattern of change from low B to high shown roughly in this figure appears to be a common feature of the symmetry of the expansion-contraction coordinate along the radius from the knots. This figure describes the molecular changes in the process characterized by Lumry and Biltonen as a "subtle change". Contraction in this process is never more than a few 0.1Å, that is, a normal chemical distance change. Knots do not expand in normal physiological function but their expansion by about the same amount is the critical step in unfolding of the native state. That process might also be called a "subtle change" but introduces some confusion since neither process is detectable in coordinate changes in diffraction experiments.

Residue-exchange experiments are likely to be least ambiguous when designed to study only one kind of substructure at a time and knots with fewer atoms are better suited for such studies than matrices. An exceptionally informative example of such design is the study of the knot of BPTI by Kim, Fuchs and Woodward, a classic among SDM studies. With fixed matrix they varied the residues in the single knot of this protein one at a time by exchanges with alanyl residues. The knot was conserved insofar as there were no changes in the number of residues but exchange by alanine at each position modified major

properties of the knot such as thermal stability and distribution between knot and matrix groups. This comparison of knots with the "slow-exchange core" defined by residues was among the first to provide unambiguous evidence that knots play a major role in folding to native states. The consequences of the residue-by-residue exchanges were followed by proton-exchange rates at knot sites. Knots like matrices undergo random conformational fluctuations though of much lower amplitude. Exchange at knot sites depends on the amplitude of the fluctuations toward the transition state for melting. Kim et al exploited that fact by altering both native and melting transition state. Residues moved out of knots as a consequence of these changes were detected because they exchanged more rapidly often in single groups. Changes in thermal stability probed the effects of these modifications to show that overall stability is correlated with knot composition. The more work required for knot expansion, the lower the probability of exchange. This powerful method for probing knots requires only the positive identification of knots residues so it is also accessible with NMR data for exchange rates. In this way Kim et al confirmed the deduction by Gregory and Lumry that folded stability is dominated by the knots and not by the matrices or as Woodward puts it "the slow-exchange core (proton exchange) is the folding core". The validity of this conclusion lies in the fact that knots determine thermodynamic stability of proteins not the path of folding. The deduction can be established in other ways a very good one based on the relationship between the activation free energy for melting and the standard free-energy change in that process is given in Biophysical Chemistry, December 2002, (Substructures and folded stability.) That study of Kim, Fuchs and Woodward was supplmented by exchange of matrix residues by alanyl residues carried out by Kuroda and Kim. The stud is very similar to the cassette mutation studies of Goldenberg on BPTI and also on HEW lysozyme. Exchange in the matrices did not destroy stability in contrast to many of the knot muitants of Kim, Fuchs and

Woodward. Nor did elimination of one disulfide bond. It will be possible to excise or reconstruct knots to quantitative their involvement in thermodynamic stability such studies being also necessary to understand knot construction. One knot of phosophilase A2 has already been isolated and found to have high folded stability an especially interesting study because the knots are pure alpha helix. Single knots have been excised or constructed for a few other protein (e.g., @). These also have high thermal stability perhaps more than that in situ.

The order of assembly in knot formation is the order of increasing knot stability. This ordering is often called FILO (first in-last out). Residues most strongly held are generally the slowest to exchange. Thus even with only partial knot formation the strongest parts of knots tend to have slowest exchange rates. They are said to be better "protected" Confusion has arisen because of the common assumption that proteins are homogeneous and reach native states only after all parts are fully assembled. If that were true, the most protected sites would be a LIFO (last in-first out) group.

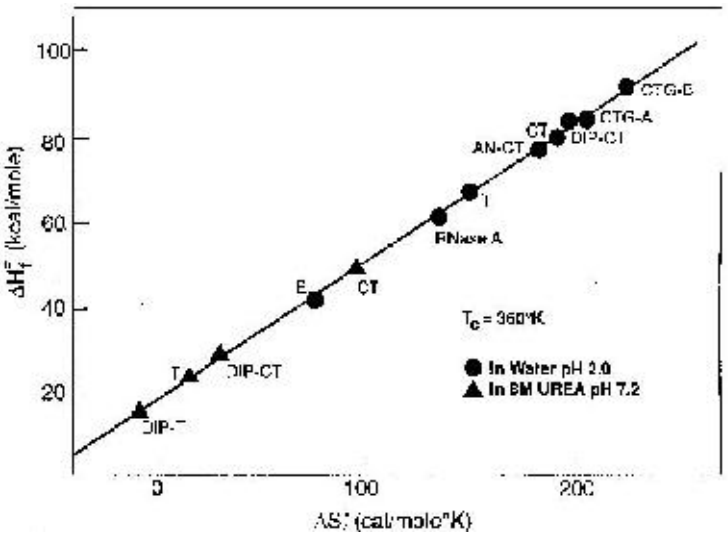

Figure 3-8. Pohl's compensation plot of activation enthalpy against activation entropy for the melting rate constant of mesophilic proteins. This plot has been the key to unraveling the thermodynamic structure of such proteins. It shows that the knots are the major source of the thermodynamic stability of the native species. This is the least-square plot of data from Pohl's habilitation.

Parsimony in protein evolution

Pohl's compensation plot is a kind of rosetta stone for protein chemistry. It has reconciled the analyses of protein melting data reported by Murphy, Privalov and Gill as a result of the fact that the activation heat capacity for melting of mesophiles is zero within small error so that for the two-step melting process of such proteins all the large standard heat capacity change applies to the refolding rate process. Biltonen first noticed this special behavior in the melting data for chymotrypsinogen obtained by Eisenberg and Schwert in 1948, a study stimulated by Henry Eyring and the first complete study of this sort to be reported. Kunitz completed the same kind of study with soybean trypsin inhibitor at about the same kind but the reversibility was not satisfactory. Statistical analysis of Pohl's data produced the unique compensation temperature for mesophile knots since found to be ubiquitous. Pohl also found a rough linear dependence of the activation enthalpy and entropy for the melting rate on total number of residues. However, he examined the relevant data for only the few

proteins of the trypsin family he had studied so this suggestion has been extended in only a very few studies. Chief among these that we have found is the melting rate data of Matusumura et al shown in Fig. 3-9. The compensation plot for their the results of their residue exchanges has a slope of 354K identifying those exchanges as in knot residues. The very large range values for the rate parameters also found in less-detailed studies such as the cassette exchange measurements of Goldenberg et al. Using the knot descriptions from diffraction studies of their protein, we find that folded stability is destroyed only in some of the substitutions in knot residues.

Activation quantities for melting rates of proteins are very rare primarily because to be useful they must be obtained in dilute buffers with no denaturing cosolvents of either structure making or structure breaking class and the activation enthalpies are very large. There is no melting up to experimentally difficult temperatures and then the rates are very fast. Pohl and others notable Biltonen used a temperature-jump apparatus following Pohl. The data of Matsumura et al shown in the last figure are thus particularly important because they test the effects of substitutions in the knots on the melting rate under these useful conditions. There are several important logical deductions stemming primarily from the zero activation heat capacity of melting:

1. The standard test for two-state behavior introduced by Lumry, Biltonen and Brandts requires a single transition state and in the special case of protein melting the two steps the first formation of the transitions state from the stable native state and the second decay from the transition state to the melted product are separated by their heat-capacity dependencies. The normal product in dilute buffers above the temperature at which the lower-temperature state of water begins to become dominant, 285K, is a motile micellelike porous polypeptide network with labile disulfide arrangement. The volume at 298K is no greater twice that of the native species but increases slightly with increasing temperature since the "bubble" is an osmometer depending on the activity of water. Despite evidence of many kinds nearly all modern

protein investigators confuse the bubble state with much greater unfolding toward random-coil conformations. The major difference is in the degree and type of exposure to bulk water.

2. The two equilibrium steps in melting are identified with the two rate steps so that the transition state very closely resembles the first equilibrium step and that step is loss of knot integrity. The stress imposed on the knots by the stress imposed on the matrices by the knots is released at the transition state with very little change in geometry of matrices or knots, about 0.3 angstroms for the thowl protein.

3. All mesophiles thus far tested fall on Pohl' compensation plot regardless of the number of residues showing that the activation enthalpy and entropy are in a fixed ratio independent of temperature their contributions from matrices are in the same ratio.

4. The position of a protein on Pohl's plot is determined by the activation enthalpy and entropy yet it is well known that a difference in sequence for a given number of residues alters these quantities. The figure for kanamycin nucleotidyl transferase from the work of Matsumura et al given above (Fig. 3-9) shows an excellent set of examples for knot substations since the compensation temperature is 354K. There are many SDM experiments showing that the point for a given protein moves along the Pohl plot with substations that do not change the total number of residues. Thus in order to show systematic dependence of activation enthalpy and entropy on total numbers of residues it is necessary to compare only the wild types of each protein. What is the basis for this most remarkable pattern?

5. The explanation if not known but it is reasonable to expect that for enzymes it is related to the catalytic parameters and thus another manifestation of the

determination of specificity by those parameters. In an experiment such as that of Kim, Fuchs and Woodward (vide supra) measurement of the catalytic parameters as a function of temperature for a series of knot mutants with a series of substrates should advance this exploration to the next step. The essential studies would measure melting rates and equilibria, catalytic rate parameters, proton-exchange rates and B factors for a short selection of specific substrates. Fragments of such studies have been reported of which the Kim, Fuchs and Woodward study is an outstanding example, but we have not found systematic catalytic parameter determination for a mutant series to give a reliable description of the their variation with residue exchange The study of the boronic-acid acyl compounds of α-lyctic protease by Stroud and @ showed the average matrix B factors to decrease systematically with the kcat of the ester substrates with the same side chain. That can be shown by separating knot and matrix substructures. It might be expected that catalytic rate parameters will demonstrate major changes with residue substitution in the matrices. So far that detail does not seem to have been met but our search of the literature for it has been somewhat perfunctory. Another important proviso is that most deductions about mesophiles made in these volumes are based on data from a relatively small number of examples. The justification for this is that the somewhat less detailed examination of a large number of enzymes for construction details has thus far found them to be all the same. Then it should be possible to generalize to the many from the few and that is the working hypothesis. For example, The 354K compensation temperature shows that the substitutions at knot positions produce wide variation in activation enthalpy and entropy values for melting rates without change in a fundamental

relationship between knots and matrices produced by these substitutions That relationship is developed and discussed in our paper entitled “Parsimony in protein evolution” (Biophysical Chemistry, December 2002).