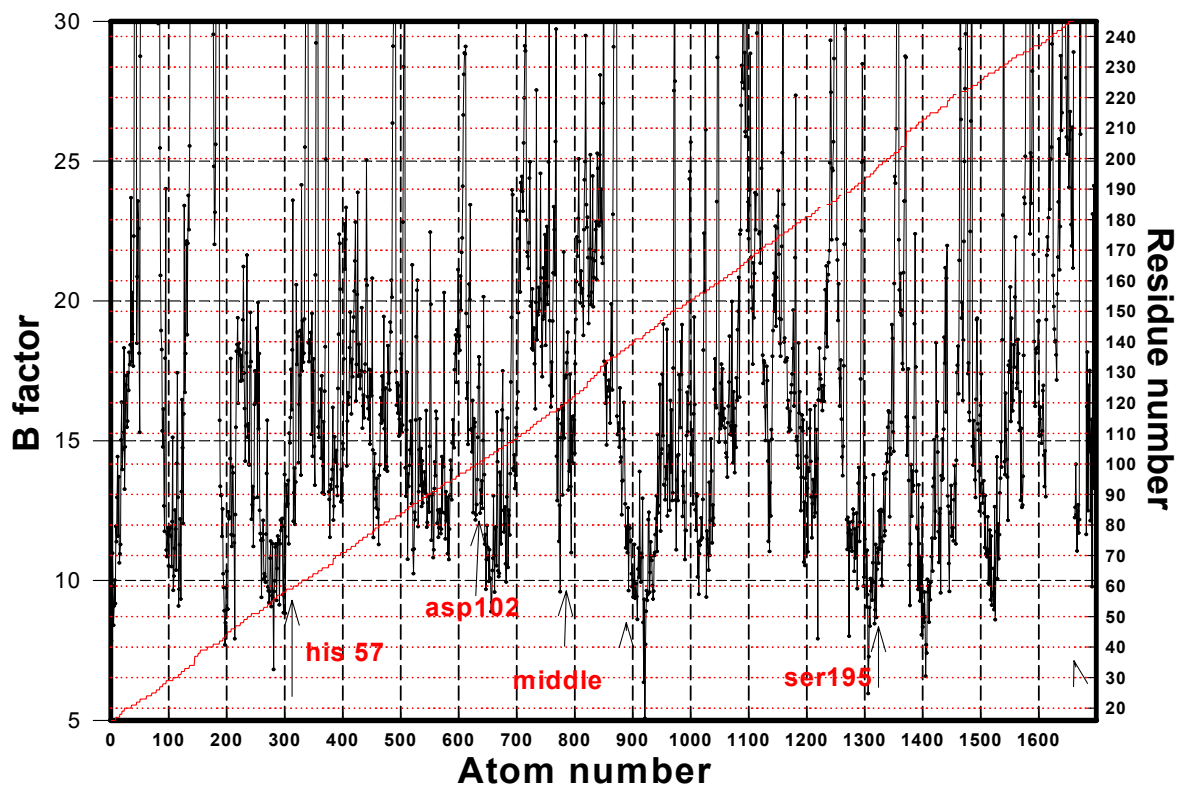


Chapter 10. Enzyme structure

A major feature of the trypsin family of proteases is the palindromic pattern taken by the B factors of the knot atoms thus producing C-2 rotation symmetry. That is so common in enzymes that we have found only two exceptions in examining a large fraction of the enzymes for which data are given in the PDB and those do not necessarily deviate from the basic construction properties implicit in the B-factor palindromes. Instead they may reveal a new level of understanding of catalytic mechanism. These are now the T-1 nuclease and the T-4 lysozyme, possibly abnormal only because of their large polymeric substrates..

Fig. 10-1. B-factor plot of trypsin family (This is salmon trypsin with some reduction in



1aoj.pdb salmon trypsin (with benzamide)

matrix values because of benzamide) Asp 102 is not essential for function but has been added in evolution for speed. It is balanced by thre139 but how and for what reason is not known.

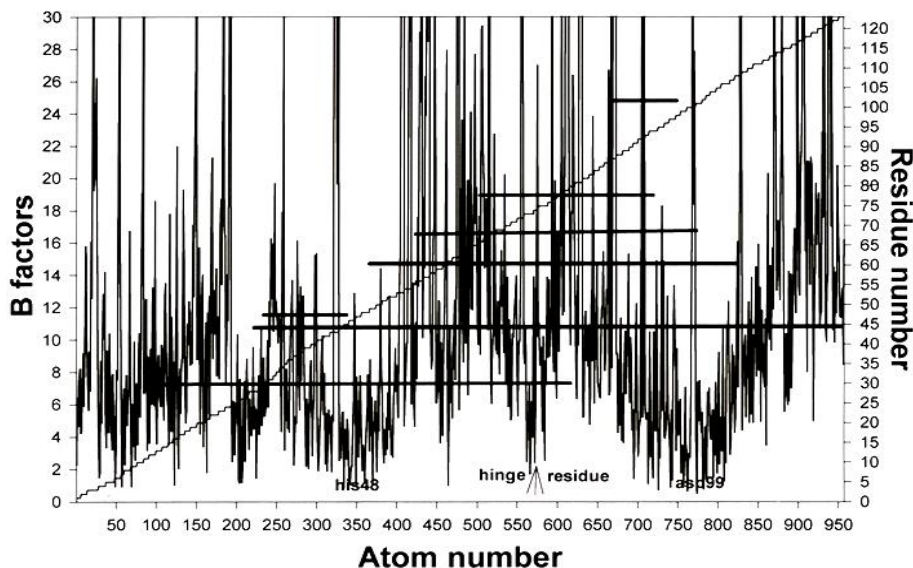


Figure 10-2. Phospholipase A2 has two C-2-matched helix cusps each broken at the middle to expose one of the functional groups (His4 or Asp99) the hinge between these functional domains consists of the three disulfide groups (horizontal bars) centered symmetrically on the cusps. Pure helix knots such as found in this enzyme are not common. Lipid substrates lie in a trough between the knots.

In the trypsin family the two knots one in each of the paired functional domains come from non-contiguous regions of the polypeptide as in BPTI and illustrate this knot class as contrasted with the class in which the knots are formed using residues in long or short contiguous series usually secondary segments, e.g., phospholipase A2 (fig. @). In either class the two functional domains required for the catalytic mechanism are of exactly equal numbers of residues and total mass. The symmetry is two-fold rotation about an axis passing from the central residue and between the two (chemically) functional groups one to each domain, well separated along the chain but very close to each other in space. There are a few other enzymes that lack two-fold symmetry although they have two functional domains. HEW lysozyme is an example of this rare class and since it is slow and inefficient in catalysis it may be of “ancient “ origin not yet fully modernized. It is abnormal also in the fact that the polypeptide chain passes between the functional domains twice. Enzymes with cofactors such as the zinc ion in carboxypeptidase A have C-2 symmetry broken only because the

cofactor is not equally divided between the catalytic functional domains. In that protein three ligands to metal ion come from one of the functional units and one from the other with a water molecule held in what appears to be the fifth ligand position. Other examples are given in Chapter. @ to illustrate the ubiquitous appearance of that symmetry.

As already noted the low-B-factor pattern defines a protein family in terms of the positions of knot atoms but without exact conversation of their residues.. Trypsin, elastase and chymotrypsin have the same low-B palindromes but also appear to have the same palindromic relationship among the highest B atom positions (Table I.). However mismatching at high and intermediate B values is more difficult to detect since the apparent palindrome precision may be related to x-ray-diffraction precision as much as to evolutionary selection. However examples from the Protein Databank suggest that in enzymes those atoms do not show rigorous palindromy suggesting that some of the matrix atoms with intermediate B values are involved in physiological functions apparently through the contacts they provide for substrates before and during domain closure.(Discussed in Chapter @.).

In that connection it is worth pointing out that the search for palindromic characteristics using B values must be done with some caution. With modern procedures for refining the x-ray-diffraction data, atom positional coordinates and B values are refined together in high-resolution studies so a simple procedure for finding the knots is to sort the B values for the N and O atoms with lowest B values. Since these form the hydrogen bonds central to knot formation, they come in pairs and matched using the midpoint residue between the catalytic domain pair. With better data the latter is easy to find. Often the functional residues of an enzyme have been determined by other experiments and although they are not precisely palindromic, their positions in the

palindrome relative to the midpoint residue are in a fixed ratio and so provide a reliability check on the choice of central residue. However this is true for trypsin and other enzymes that have only one functional group per domain. In metal enzymes such as carboxypeptidase A the zinc ion is a direct participant in catalysis and as shown in Fig. @ the liganding residues do not define the center exactly.. Thermolysin, a larger enzyme with a zinc ion in the functional system, has all the same construction features and the same ligands but the first catalytic domain has both histidines rather than one histidine and one glutamine as found in carboxypeptidase A. The atoms with lowest B values may not be in the peptide-peptide H bonds as is discussed in connection with the exceptional strengths of the knots (Chapter @) Folding may produce lower B values in backbone atoms than in their hydrogen-bonding groups. T-1 nuclease is an example and it suggests some not yet well known chemistry in knots containing secondary structure. The analogy appears to be with the fragline spider silks and is at present the only possible clue to the exceptional strength of the knots. (See Chap. @))

More detailed search methods for palindromic patterns in B-factor data are discussed elsewhere. They are useful with poor B-factor data and also in separating knot atoms from matrix atoms .Bahar and coworkers apply an algorithm to the interatomic coordinates to find the highest local vibrational frequencies and the highest packing densities in this way by-passing the use of B factors for detection of knot atoms.

Fish trypsins and bacterial trypsins have the same low-B patterns as the three enzymes in Table 2. So also do the bacterial trypsins so far reported in the PDB. Thus the critical knot peptide-peptide H bonds connect residues in the same chain positions. On the other hand α -lyctic protease and the subtilisin family, all serine proteases with the famous catalytic triad, do not have the

trypsin pattern and are not members of the same protein family as defined by its palindrome pattern. This raises the question as to how protein superfamilies should be defined.

T-1 nuclease lacks the detailed palindromy we have thus far found in other enzymes but has very close balance of the total atom free volume of its two knots so with that so far unique exception it appears that free-volume matching is more important than the exact palindrome. A number of interesting questions arise at this stage. In the trypsin family as defined by the palindrome positions residue conservation is not exact as is shown in Table 2. Although some differences are those of the like-for-like type there are some that are not. This shows that at least for the trypsin family it is knot position that is conserved in the genome rather than knot sequence. Thus like the classical secondary polypeptide structures on which major attention is focused simply because the knot-matrix alternative is not known, universal focus in protein research on residues sequences and positions may not be profitable.

The residue composition of the two branches of the palindrome show no common pattern and nothing yet suggests conservation between branches. The latter absence is consistent with current knowledge of DNA expression. Anti-sense expression might explain the positioning of the knot atoms but to explain the residues involved as an intrinsic feature of DNA expression evolutionary selection and conservation now appears to be so complicated as to be impossible. The HIV-1 protease has the same construction principle but using two identical functional domains connected by a small knot "hinge" rather than by the polypeptide chain. That construction is essentially two-dimensional and has restricted possibilities for specificity as a result. More common knot constructions such as those in trypsin probably arising from some primordial gene duplication have had to evolve to head-to-head or tail-to-tail arrangements

of the catalytic functional domains. That does not suggest parsimony in evolution but the arrangement is nevertheless almost universal among enzymes and, as shown with the streptococcus IgG binding protein, ubiquitin and many others, also very common for proteins with other physiological role. In any strictly Darwinian biosphere the evolutionary process would not be guided by cooperativity between the arms of a palindrome other than those improvements in function arising strictly by chance. There is no way for cross talk among the families so the knot palindrome for most protein families seems to be the result of separate series of accidents. However groups of families may have arising by genetic drift maintaining the same knot positions with many changes in matrix residues allowed. Thus subtilisin BPN' and proteinase K, once thought to be early on the scene and one late, have very similar palindromes differing only in the exact positions along the chain of a few knot hydrogen bonds. They have very little common residue sequence. Some changes in the knowledge of hereditary trees are likely to be required when they are explored using B factors and knots.

B values and their changes in function as measured by acyl derivatives of the trypsins and the binding of specific inhibitors such as pepstatin by rhizopepsin indicate that positional changes of matrix atoms are under 1Å angstrom. Maximal contraction of a protein diameter in the expansion-contraction process of matrices does not appear to be larger than that and thus as likely to be lost in the noise of protein crystallography.

The two-domain construction of enzymes is an example of what has been called "the pairing principle". It appears to have some of the advantages of the opposed thumb as a source of mechanical force. It was first detected in proteins by the regulation of oxygen binding and redox potential at heme iron of heme proteins by adjustments in length of the bond from heme iron to proximal

histidine and the displacement of the iron ion from the mean porphyrin plane. The stress so developed is physiologically significant but produces maximum displacement of the iron ion from the mean porphyrin plane less than 0.7 Å. Hemoglobin is an example of one kind of mechanical stress mechanism in proteins called by Eyring a “rack” for reasons that are obvious with that protein. According to the rack hypothesis for hemoglobin and other metal-dependent proteins the static and dynamic properties of metal ions in enzymes are modulated by conformationally based constraints on bond lengths, bond angles mediated by dynamical displacements to support function.

The enzymes such as the trypsins form another class of rack mechanisms not involving metal ions or other coenzymes but also apparently dependent for function on the compression produced as the two functional domains close. Closure here is much like that of a pair of scissors with C-2 symmetry about the long axis to keep the force vectors collinear. As expected with rack mechanisms, the catalytic function seems to be effected in rate and specificity by application of force on the reaction assembly of substrate and functional groups to raise their potential energy. Narrowing the transition-state free-energy barrier in this way by increasing overlap of potential wells and the resulting covalency should facilitate proton, hydride ion or hydrogen ion transfer and particularly improve proton tunneling. This idea seemed particularly appropriate for the bond-rearrangement steps in tryptic catalysis but has broad application since efficient electron and proton migrations can be effected by a small amount of force acting over a few 0.1 Å along the reaction coordinate. Work optimally generated by force on domain closure of 10 kcal generates 7 orders of magnitude in increased rate constant when applied along the reaction coordinate. Nature by the construction of large knot-matrix molecules has found a source of the force but also a vector direction for its efficient application. The pre-transition state need persist only long enough for thermal activation into the transition state to occur

longer than a picosecond. but less than some nanoseconds to balance force production against probability of thermal activation. In view of the fact that the normal fluctuations in conformation have similar periods current emphasis on equilibrium geometry is misplaced. The liquid-glass transition measured by dielectric relaxation has an average period in the range of a few ns.

Carloni and coworkers have developed a detailed picture of domain closure in the HIV-1 protease corresponding to Fig. @ and in doing so have realized that that enzyme is a “nut cracker” with a compression mechanism likely to be found common among enzymes of many types. Their terms is a considerable improvement over the ancient rack terminology of Eyring, Lumry and Spikes and we shall generally use it henceforth.