4-8 Deducing Function from Structure: Protein Superfamilies

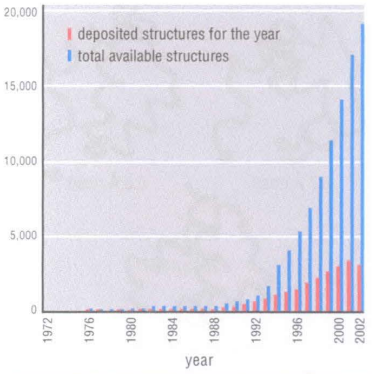


Figure 4-28 Growth in the number of structures in the protein data bank
Both yearly and cumulative growth is shown. (Taken from the Protein Data Bank website: http://www.rcsb.org/pdb/holdings.html)

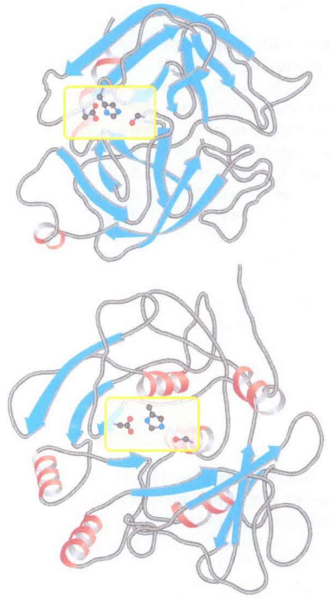


Figure 4-29 The overall folds of two members of different superfamilies of serine proteases
The enzymes are chymotrypsin (top) and subtilisin (bottom). The residues in the catalytic triad are indicated for each.

Members of a structural superfamily often have related biochemical functions

In contrast to the exponential increase in sequence information, structural information, which is still chiefly obtained by X-ray crystallography and NMR, has up to now been increasing at a much lower rate (Figure 4-28). One goal of the structural genomics initiatives that have been implemented since the release of the first complete genome sequence is to increase the rate at which experimentally derived structures of the gene products are produced. The driving force behind these initiatives is the assumption that, in addition to defining the ensemble of all possible protein folds, comprehensive structural information could provide a firmer basis than sequence for functional predictions, as three-dimensional structure changes much more slowly than sequence during evolution. A good reason for optimism that these assumptions will hold true is the existence of **superfamilies** of proteins with related structures and biochemical functions.

A superfamily is loosely defined as a set of homologous proteins with similar three-dimensional structures and related, though not necessarily identical, biochemical functions. Almost all superfamilies exhibit some functional diversity, which is generated by local sequence variation and/or domain shuffling. Within enzyme superfamilies, for example, substrate diversity is common, while parts of the reaction chemistry are highly conserved. In many enzyme superfamilies, the sequence positions of catalytic residues vary from member to member, despite the fact that they have equivalent functional roles in the proteins. These variations may make the assignment of a protein to a superfamily from sequence comparison alone problematic or impossible. Although some superfamily members may be similar in sequence, it is the structural and functional relationships that place a protein in a particular superfamily. Within each superfamily, there are **families** with more closely related functions and significant (>50%) sequence identity.

Because the total number of protein folds and the total number of biochemical functions is smaller than the total number of genes in biology, if a protein can be assigned to a superfamily from sequence or structural information, at the very least the number of its possible functions can be narrowed down, and in some instances it may be possible to assign a function precisely.

The four superfamilies of serine proteases are examples of convergent evolution

Striking examples of similarities in biochemical function but quite different biological roles come from enzymes where the chemical reactions are the same but the substrates can differ considerably. There are, for instance, many hundreds of enzymes that hydrolyze peptide bonds in protein and polypeptide substrates, but they can be grouped into a small number of classes, each with its own characteristic chemical mechanism. The most numerous class comprises the serine proteases, in which the side-chain hydroxyl group of a serine residue in the active site attacks the carbonyl carbon atom of the amide bond that is to be hydrolyzed. Two other characteristic residues, a histidine and an aspartic acid (or a glutamic acid), are involved in assisting this hydrolysis, forming a catalytic triad. Serine proteases fall into several structural superfamilies, which are recognizable from their amino-acid sequences and the particular disposition of the three catalytically important residues in the active site (Figure 4-29). Each serine protease superfamily has many members but there is no obvious relationship between the superfamilies, either in sequence or structure. The three residues of the catalytic triad are

Definitions

family: a group of homologous proteins that share a related function. Usually these will also have closely related sequences. Members of the same enzyme family catalyze the same chemical reaction on structurally similar substrates.

superfamily: proteins with the same overall fold but with usually less than 40% sequence identity. The nature of the biochemical functions performed by proteins in the same superfamily are more divergent than those within families. For instance, members of the same enzyme superfamily may not catalyze the same overall reaction, yet still retain a common mechanism for stabilizing chemically similar rate-limiting transition-states and intermediates, and will do so with similar active-site residues.

References

Gerlt, J.A. and Babbitt, P.C.: **Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies.** *Annu. Rev. Biochem.* 2001, **70**:209–246.

Krem, M.M. *et al.*: **Sequence determinants of function and evolution in serine proteases.** *Trends Cardiovasc. Med.* 2000, **10**:171–176.

Perona, J.J. and Craik, C.S.: **Evolutionary divergence of substrate specificity within the chymotrypsin-like**

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 164

found in a different order in different locations along the sequence in each superfamily: nevertheless, in the tertiary structure they come together in a similar configuration. Presumably, the existence of a similar active site is due to convergent evolution, while within each superfamily, divergent evolution has produced distinct individual proteases with very similar structures but different substrate specificity.

A given serine protease can be highly specific for a particular target amino-acid sequence—although some are relatively nonspecific—so that, in general, the substrate(s) for such a protease cannot be predicted from knowledge of the sequence or even the structure of the enzyme. However, observation of a serine protease fold combined with the right active-site residues is diagnostic for a protease. Serine proteases participate in such diverse cellular functions as blood clotting, tissue remodeling, cell-cycle control, hormone activation and protein turnover. A small number of members of some families within the serine protease superfamily have lost one or more of the catalytic residues and perform non-catalytic functions such as forming a structural matrix.

Another large enzyme superfamily with numerous different biological roles is characterized by the so-called polymerase fold, which resembles an open hand (Figure 4-30). DNA polymerases of all types from all organisms studied so far appear to share this fold, as do RNA polymerases and viral reverse transcriptases. In most cases, sequence comparisons alone do not make this functional distinction, and in some cases no sequence similarity between the families of the polymerase fold superfamily is apparent. However, structural evidence indicates that every enzyme that transcribes or replicates nucleic acid polymers has probably descended from a common ancestor.

There are many other examples of related but differing functions among members of a superfamily. It should be borne in mind, however, that in some proteins with similar function, equivalent active-site residues come from different positions in the sequence, obscuring superfamily membership until structural information is obtained.

Very closely related protein families can have completely different biochemical and biological functions

There are some well-known cases of significant differences in function at very high levels of sequence identity. One example is the crystallins, which appear to have evolved from several different enzymes. Although some crystallins retain more than 50% sequence identity to these enzymes, they function as structural proteins, not enzymes, in the eye lens.

And with increasing numbers of structures being solved for proteins of known function, the functional diversity of many other protein superfamilies has been revealed. Thornton and co-workers have assessed the functional variation within homologous enzyme superfamilies containing two or more enzymes. Combining sequence and structure information to identify relatives, the majority of superfamilies display variation in enzyme function, with 25% of the superfamilies having members of different enzyme types. For example, the α/β hydrolase superfamily has at least four different functions; the ferredoxin superfamily has at least three. For single- and multidomain enzymes, difference in biochemical function is rare above 40% sequence identity, and above 30% the overall reaction type tends to be conserved, although the identity of the substrate may not be. For more distantly related proteins, sharing less than 30% sequence identity, functional variation is significant, and below this threshold, structural data are essential for understanding the molecular basis of observed functional differences.

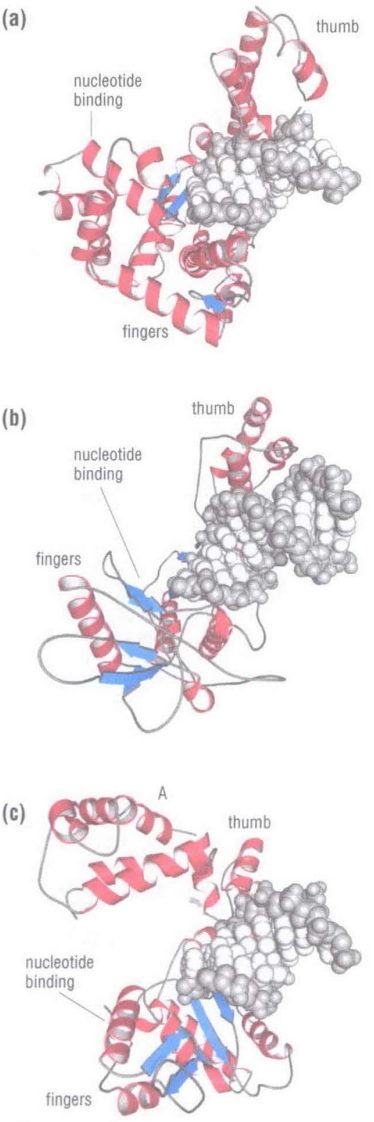


Figure 4-30 A comparison of primer–template DNA bound to three DNA polymerases
(a) Taq DNA polymerase bound to DNA. The DNA stacks against the "fingers" and is contacted across the minor groove by the "thumb" domain. **(b)** The binary complex of HIV-1 reverse transcriptase and DNA. This structure does not have a nucleotide-binding alpha helix in the fingers domain. Instead, a beta hairpin probably performs this function. **(c)** The ternary complex of rat DNA polymerase β with DNA and deoxy-ATP (not shown). Although this polymerase has an additional domain (A), the "thumb" domain similarly binds the DNA primer–template in the minor groove, while the "fingers" present a nucleotide-binding alpha helix at the primer terminus. (PDB 1tau, 2hmi and 8icp)

serine protease fold. *J. Biol. Chem.* 1997, **272**: 29987–29990.

Siezen, R.J. and Leunissen, J.A.: **Subtilases: the superfamily of subtilisin-like serine proteases.** *Protein Sci.* 1997, **6**:501–523.

Steitz, T.A.: **DNA polymerases: structural diversity and common mechanisms.** *J. Biol. Chem.* 1999, **274**:17395–17398.

Todd, A.E. *et al.*: **Evolution of function in protein superfamilies, from a structural perspective.** *J. Mol. Biol.* 2001, **307**:1113–1143.

Evolution of function in protein superfamilies, from a structural perspective:

http://www.biochem.ucl.ac.uk/bsm/FAM-EC/

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 165

4-9 Strategies for Identifying Binding Sites

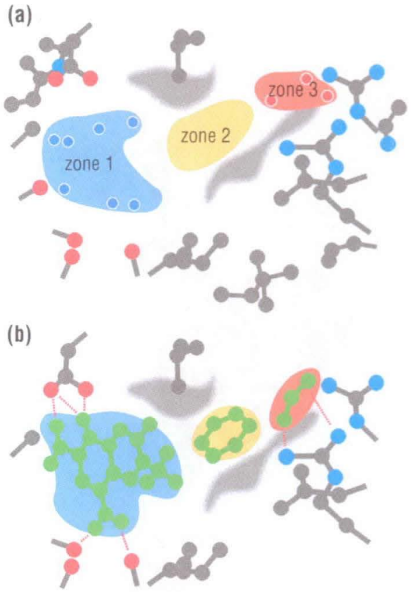


Figure 4-31 Example of the use of GRID
Three different types of probes have been used to locate binding sites for functional groups in the active site of the enzyme dihydrofolate reductase. **(a)** Zone 1 (blue) is a good site for binding electrostatically positive groups, with the energy function minima from an amino probe shown in blue dots. It was also identified with a carbon probe as being a good pocket for shape complementarity. Zone 2 (yellow) is a good site for hydrophobic interaction, as illustrated by the hydrophobic molecular surface (grey shapes) in that region. Zone 3 (red) is a good binding site for electrostatically negative groups, with minima from a carboxylate probe shown by the red dots. **(b)** Overlay of three pieces of a known inhibitor of dihydrofolate reductase onto the zones of favorable interaction energy found by GRID. Figures adapted from http://thalassa.ca.sandia.gov/~dcroe/builder.html

Binding sites can sometimes be located in three-dimensional structures by purely computational means

Whether the three-dimensional structure of a new protein is determined experimentally or computationally (see sections 4-6 and 4-7)—there will always be the problem of finding those sites on the protein surface that are involved in its biochemical and cellular functions. This problem is particularly acute when the structure reveals a polypeptide chain fold that has never been seen before, and there is no obvious cofactor or other bound ligand to identify a functional site. But it can also apply to structures of proteins with known folds, as proteins often have more than one function (see section 4-13) and even a familiar protein may have acquired additional functions and functional sites in the context of a different organism.

To some extent, the characteristics of binding sites that were discussed in section 2-4 can be used to identify regions of a protein's surface that are good candidates for functional sites. These characteristics include concavity—binding sites are usually depressions rather than protrusions or flat areas, although many exceptions are known—as well as a higher than average amount of exposed hydrophobic surface area. However, such generalizations usually only narrow down the possibilities. What is needed is a method, ideally a computational method so that it can be used with homology models as well as with experimentally derived structures, that can scan the surface of a protein structure and locate those sites that have evolved to interact with small molecules or with other macromolecules. Some success has been obtained with methods that scan the surface and look for sites of specific shape. Residue conservation analysis can also be quite revealing if there are enough homologous sequences (see Figure 4-23).

Several other computational methods have been proposed and tested on experimentally determined structures. Most use a "probe" molecule and an energy function that describes the interaction of the probe with the residues on the protein surface. Binding sites are identified as regions where the computed interaction energy between the probe and the protein is favorable for binding. Two widely used methods, GRID and MCSS (multiple conformations simultaneous search), use this strategy. In GRID, the interaction of the probe group with the protein structure is computed at sample positions on a lattice throughout and around the macromolecule, giving an array of energy values. The probes, which are usually used singly, include water, the methyl group, the amine NH_2 group, the carboxylate group and the hydroxyl group, among others. Contour surfaces at various energy levels are calculated for each probe for each point on the lattice and displayed by computer graphics together with the protein structure (Figure 4-31). Contours at negative energy levels delineate regions of attraction between probe and protein that could indicate a binding site, as such contours are found at known ligand-binding clefts.

The approach taken by MCSS is similar in principle but differs in detail and can take into account the flexibility of both the probe molecule and the protein. The resulting distribution map of regions on the protein surface where functional groups show a favorable interaction energy can be used for the analysis of protein–ligand interactions and for rational drug design.

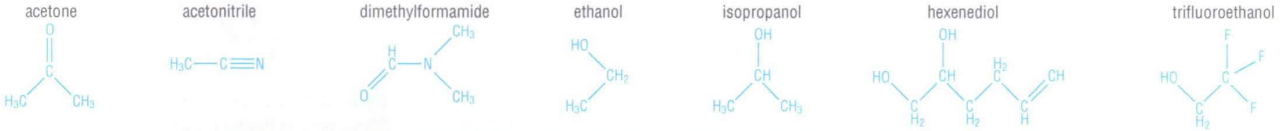


Figure 4-32 Some organic solvents used as probes for binding sites for functional groups

References

Allen, K.N. *et al.*: **An experimental approach to mapping the binding surfaces of crystalline proteins.** *J. Phys. Chem.* 1996, **100**:2605–2611.

Aloy, P. *et al.*: **Automated structure-based prediction of functional sites in proteins: applications to assessing the validity of inheriting protein function from homology in genome annotation and to protein docking.** *J. Mol. Biol.* 2001, **311**:395–408.

Bitetti-Putzer, R. *et al.*: **Functional group placement in protein binding sites: a comparison of GRID and MCSS.** *J. Comput. Aided Mol. Des.* 2001, **15**:935–960.

Byerly, D.W. *et al.*: **Mapping the surface of *Escherichia coli* peptide deformylase by NMR with organic solvents.** *Protein Sci.* 2002, **11**:1850–1853.

Dennis, S. *et al.*: **Computational mapping identifies the binding sites of organic solvents on proteins.** *Proc. Natl Acad. Sci. USA* 2002, **99**:4290–4295.

English, A.C. *et al.*: **Experimental and computational mapping of the binding surface of a crystalline protein.** *Protein Eng.* 2001, **14**:47–59.

Goodford, P.J.: **A computational procedure for determining energetically favorable binding sites on biologically important macromolecules.** *J. Med. Chem.* 1985, **28**:849–857.

Laskowski, R.A. *et al.*: **Protein clefts in molecular recognition and function.** *Protein Sci.* 1996, **5**:2438–2452.

Liang, J. *et al.*: **Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design.** *Protein Sci.*

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 166

Computational methods are, however, extremely inefficient when no information is available to limit the regions of the protein surface to be scanned. Scanning an entire protein surface with either GRID or MCSS yields hundreds of possible binding sites of roughly equivalent energy. Thus, these tools are most useful in cases where the active site is already known and one wants to determine what sorts of chemical groups might bind there.

Experimental means of locating binding sites are at present more accurate than computational methods

MSCS (multiple solvent crystal structures) is a crystallographic technique that identifies energetically favorable binding sites and orientations of small organic molecules on the surface of proteins; this experimental method can find likely functional sites on the surface of any protein that can be crystallized. The method involves soaking protein crystals in an organic solvent that mimics a functional group on a ligand: thus, ethanol will probe for hydroxymethyl-binding sites such as those that interact with a threonine side chain; dimethylformamide identifies binding sites that interact with the C=O and N–H groups of peptides, and so on (Figure 4-32). Determination of the protein structure at resolutions of the order of 2 Å in the presence of the solvent probe reveals the solvent-binding sites (Figure 4-33). If the experiment is repeated with several different probes, it is found that they cluster in only a few binding sites, regardless of their polarity. These sites are the functional sites on the surface of the protein. In contrast to the computational methods, MSCS involves direct competition between the probe and the bound water on the surface of the protein. As it is displacement of this water that drives ligand-binding events (see section 2-4), MSCS finds a much more restricted set of binding sites than do the computational methods.

High-resolution structures of crystals of the well-studied enzyme thermolysin soaked in acetone, acetonitrile, or phenol show probe molecules clustering in the main specificity pocket of the thermolysin active site and in a buried sub-site, consistent with structures of known protein–ligand complexes of thermolysin (Figure 4-34). When the experimentally determined solvent positions within the active site were compared with predictions from GRID and MCSS, both these computational methods found the same sites but gave fewer details of binding. And both GRID and MCSS predicted many other sites on the protein surface, not observed experimentally, as equally favorable for probe binding.

Related experimental methods using NMR instead of X-ray crystallography study the binding of small-molecule compounds as well as organic solvents as probes. The experimental methods are accurate but cannot be used on homology models. Computational methods are not accurate enough to discriminate among possible binding sites. What is needed is a computational analog of the experimental methods. One such new computational mapping strategy has been tested recently with promising results. Using eight different ligands for lysozyme and four for thermolysin, the computational search finds the consensus site to which all the ligands bind, whereas positions that bind only some of the ligands are ignored. The consensus sites turn out to be pockets of the enzymes' active sites, lined with partially exposed hydrophobic residues and with some polar residues toward the edge. Known substrates and inhibitors of hen egg-white lysozyme and thermolysin interact with the same side chains identified by the computational mapping, but the computational mapping did not identify the precise hydrogen bonds formed and the unique orientations of the bound substrates and inhibitors.

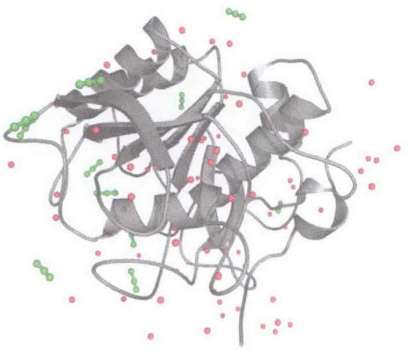


Figure 4-33 Structure of subtilisin in 100% acetonitrile Crystal structure of the serine protease subtilisin in 100% acetontitrile. The organic solvent, shown as green rods, binds at only a few sites on the protein surface, including the active site, which is approximately left of center in the figure. The red spheres are bound water molecules, which are not displaced even by this water-miscible organic solvent at 100% concentration. These bound waters should be considered an integral part of the folded structure of the protein. (PDB 1be6)

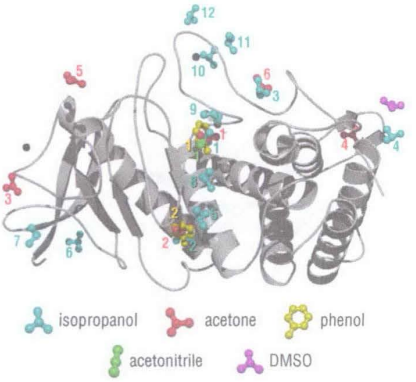


Figure 4-34 Ribbon representation showing the experimentally derived functionality map of thermolysin The binding sites for different organic solvent molecules were obtained by X-ray crystallography of crystals of thermolysin soaked in the solvents. The same probe molecules bound to different positions are numbered to identify their site of binding. The active-site zinc ion and the bound calcium ions are shown as grey and black spheres, respectively. Dimethyl sulfoxide (DMSO, purple) is present in the crystallization conditions of thermolysin; one molecule binds per molecule protein. Graphic kindly provided by Roderick E. Hubbard. (Adapted from English *et al.*: *Protein Eng.* 2001, **14**:47–59.)

1998, **7**:1884–1897.

Liepinsh, E. and Otting, G.: **Organic solvents identify specific ligand binding sites on protein surfaces.** *Nat. Biotechnol.* 1997, **15**:264–268.

Mattos, C. and Ringe, D.: **Locating and characterizing binding sites on proteins.** *Nat. Biotechnol.* 1996, **14**:595–599.

Mattos, C. and Ringe, D.: **Proteins in organic solvents.** *Curr. Opin. Struct. Biol.* 2001, **11**:761–764.

Miranker, A. and Karplus, M.: **Functionality maps of binding sites: a multiple copy simultaneous search method.** *Proteins* 1991, **11**:29–34.

Shuker, S.B. *et al.*: **Discovering high-affinity ligands for proteins: SAR by NMR.** *Science* 1996, **274**:1531–1534.

URL for GRID:
http://thalassa.ca.sandia.gov/~dcroe/builder.html

©2004 New Science Press Ltd

4-10 Strategies for Identifying Catalytic Residues

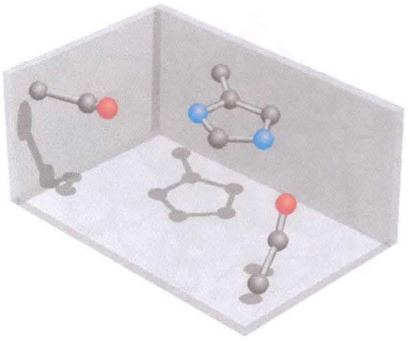


Figure 4-35 An active-site template The geometry of the catalytic triad of the serine proteases as used to locate similar sites in other proteins. Adapted from the rigid active-site geometries website: http://www.biochem.ucl.ac.uk/bsm/PROCAT/PROCAT.html.

Site-directed mutagenesis can identify residues involved in binding or catalysis

Locating binding sites, either experimentally or computationally (see section 4-9), does not automatically indicate which residues in those sites are responsible for ligand binding or, in the case of enzymes, catalysis. The standard experimental method for identifying these residues is alanine-scanning mutagenesis, in which candidate amino acids are replaced by alanine by site-directed mutagenesis of the gene. The effect of this side-chain excision on function—usually binding or catalysis—of the expressed mutant protein is then assayed. When combined with genetic assays for the *in vivo* phenotype of the mutated protein and information from the pH/rate profile of a catalytic reaction, for example, such experiments can reveal which side chains in a binding site may actually perform chemistry on a bound ligand.

Active-site residues in a structure can sometimes be recognized computationally by their geometry

But with the advent of genome-wide sequencing and protein-structure determination, a computational tool is needed that can identify such residues rapidly and automatically. This would be particularly useful in cases where the protein in question has no known function and the location of the active site is uncertain, because knowledge of the residues that can carry out chemistry might indicate what type of chemistry the protein actually performs.

The simplest of the computational methods searches the structure for geometrical arrangements of chemically reactive side chains that match those in the active sites of known enzymes. This rigid active-site approach has successfully identified the catalytic triad of a serine protease (Figure 4-35) in an enzyme of unknown function, but has not been used extensively to probe for other functions. Because it relies solely on geometry and not on position in the sequence, this method could find serine protease catalytic sites in any protein fold in which they occur. A more sophisticated variation uses a three-dimensional descriptor of the functional site of interest, termed a "fuzzy functional form", or FFF, to screen the structure. FFFs are based on the geometry, residue identity, and conformation of active sites using data from known crystal structures of members of a functional family and experimental biochemical data. The descriptors are made as general as possible ("fuzzy") while still being specific enough to identify the correct active sites in a database of known structures.

These fuzzy functional descriptors can identify active sites not only in experimentally determined structures, but also from predicted structures provided by *ab initio* folding algorithms (see sections 1-9, 4-7) or threading algorithms (see section 4-7). A disulfide oxidoreductase FFF has been successfully applied to find other disulfide oxidoreductases in a small structural database and, more recently, has been used to scan predicted protein structures derived from the entire *Bacillus subtilis* genome. A total of 21 candidate disulfide oxidoreductases were found, of which six turned out to be false positives. The method did not miss any of the known disulfide oxidoreductases and identified at least two potential new ones.

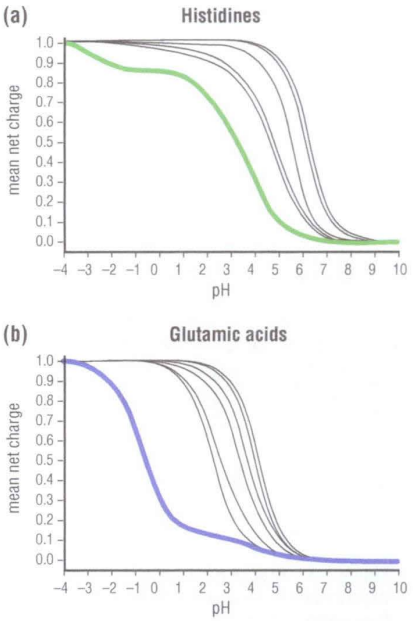


Figure 4-36 Theoretical microscopic titration curves Calculated curves for two types of ionizable residues—histidines and glutamic acids—in the structure of the glycolytic enzyme triosephosphate isomerase. In each case, most of the residues behave similarly, with a sharp change in charge as a function of pH, but one residue of each type (histidine 95 (curve in green) and glutamic acid 165 (curve in blue)) displays abnormal behavior.

References

Bartlett, G.J. *et al.*: **Analysis of catalytic residues in enzyme active sites.** *J. Mol. Biol.* 2002, **324**:105–121.

Di Gennaro, J.A. *et al.*: **Enhanced functional annotation of protein sequences via the use of structural descriptors.** *J. Struct. Biol.* 2001, **134**:232–245.

Ewing, T.J. *et al.*: **DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases.** *J. Comput. Aided Mol. Des.* 2001, **15**:411–428.

Fetrow, J.S. *et al.*: **Structure-based functional motif identifies a potential disulfide oxidoreductase active site in the serine/threonine protein phosphatase-1 subfamily.** *FASEB J.* 1999, **13**:1866–1874.

Jones, S. and Thornton, J.M.: **Prediction of protein–protein interaction sites using patch analysis.** *J. Mol. Biol.* 1997, **272**:133–143.

Laskowski, R.A. *et al.*: **Protein clefts in molecular recognition and function.** *Protein Sci.* 1996, **5**:2438–2452.

Ondrechen, M.J. *et al.*: **THEMATICS: a simple computational predictor of enzyme function from structure.** *Proc. Natl Acad. Sci. USA* 2001, **98**:12473–12478.

Reva, B. *et al.*: **Threading with chemostructural restrictions method for predicting fold and functionally significant residues: application to dipeptidylpeptidase IV (DPP-IV).** *Proteins* 2002, **47**:180–193.

Sheinerman, F.B. *et al.*: **Electrostatic aspects of protein–protein interactions.** *Curr. Opin. Struct. Biol.* 2000, **10**:153–159.

Wallace, A.C. *et al.*: **Derivation of 3D coordinate**

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 168

The limitation with all computational methods of this type is that they can only find active-site residues that conform to known active sites. A protein with a novel function will yield no result, or worse, an incorrect identification with a known site. A more general approach has been developed that does not depend on previous knowledge of active-site geometries. This employs theoretical microscopic titration curves (THEMATICS), to identify active-site residues that are potentially involved in acid-base chemistry in proteins of known structure. Location of such residues automatically determines the position of the active site, as well as providing a clue to the biochemical function of the protein.

In THEMATICS, the mean net charge of potentially ionizable groups in each residue in the protein structure is calculated as a function of pH. The resulting family of curves for each type of residue (Figure 4-36) is then analyzed for deviations from ideal behavior. A small fraction (3–7%) of all curves for all residues differ from the others in having a flat region where the residue is partially protonated over a wide pH range. Most residues with these perturbed curves occur in active sites (Figure 4-37). The method is successful for proteins with a variety of different chemistries and structures and has a low incidence of false positives. Of course, identification of acid-base residues in active sites does not necessarily establish what the overall chemical reaction must be. Most enzymatic reactions use one or more acid-base steps but catalyze other chemistries as well. Since the pK_a values of catalytic residues are likely to be perturbed by electrostatic interactions, once they are identified, computational tools such as GRASS can be used to compute and display such interactions.

Docking programs model the binding of ligands

Even when it is possible to identify active sites and to draw some conclusions about the likely chemistry they will perform, it is still necessary to determine on what substrate(s) that chemistry will operate. At present, there is no method, experimental or computational, that will enable one to find the most likely substrate for any particular active site. Some approaches involving mass spectroscopy to identify ligands pulled out of cellular extracts by the protein in question are under development, and peptide substrates for protein kinases can often be found by screening combinatorial peptide libraries, but a computational method would be most general. One promising approach is that of the program DOCK, where the shape of the binding site on a protein is represented as a set of overlapping spheres, in which the centers of the spheres become potential locations for ligand atoms. Each ligand is divided into a small set of rigid fragments that are docked separately into the binding site, allowing a degree of flexibility at the positions that join them. The fragments are rejoined later in the calculation and an energy minimum calculated for the rejoined ligand in the receptor site. The method can find binding geometries for the ligand similar to those observed crystallographically, as well as other geometries that provide good steric fit, and has been used to find possible new compounds for drug development. In such applications, each of a set of small molecules from a structural database is individually docked to the receptor in a number of geometrically permissible orientations. The orientations are evaluated for quality of fit, with the best fits being kept for examination by molecular mechanics calculations.

The method cannot take unknown conformational changes of the protein into account. In principle, it could be used to find candidate substrates for any active site, but in practice it is too computationally cumbersome, and all potential ligands are not contained in any database. The method also gives not one but many possible ligands from any database, and the energy function used to evaluate binding cannot discriminate among them.

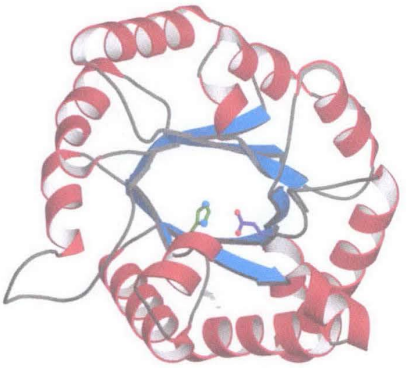


Figure 4-37 Residues that show abnormal ionization behavior with changing pH define the active site The locations of the two abnormally titrating residues in Figure 4-36 are shown on the three-dimensional structure of triosephosphate isomerase. The histidine (green) and glutamic acid (blue) that are partially protonated over a wide range of pH are both located in the active site and both are important in catalysis.

templates for searching structural databases: application to Ser-His-Asp catalytic triads in the serine proteinases and lipases. *Protein Sci.* 1996, **5**:1001–1013.

Zhang, B. *et al.*: **From fold predictions to function predictions: automation of functional site conservation analysis for functional genome predictions.** *Protein Sci.* 1999, **8**:1104–1115.

Rigid active-site geometries:
http://www.biochem.ucl.ac.uk/bsm/PROCAT/PROCAT.html

GRASS, a protein surface visualization tool:
http://trantor.bioc.columbia.edu/

DOCK:
http://www.cmpharm.ucsf.edu/kuntz/dock.html

FFF:
http://www.geneformatics.com

©2004 New Science Press Ltd

4-11 TIM Barrels: One Structure with Diverse Functions

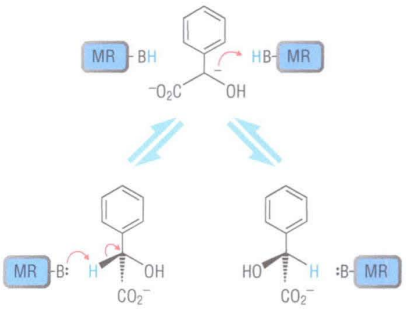


Figure 4-38 The chemical reaction catalyzed by mandelate racemase R-mandelate (left) and S-mandelate (right) can be converted into each other through the intermediate in the center. The enzyme (MR) catalyzes the reaction by removing a proton from a carbon atom adjacent to a carboxylate group and subsequently replacing it. A basic residue (B) at one side of the active site of the enzyme removes the proton and a proton is replaced by a basic residue on the other side of the active site (not shown), which can also act as an acid. The red arrows indicate the movement of electron pairs. In the reverse reaction, the two residues reverse roles.

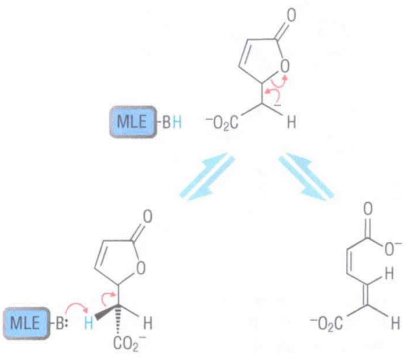


Figure 4-39 The chemical reaction catalyzed by muconate lactonizing enzyme Although the substrates are different, the core step in the catalytic mechanism of muconate lactonizing enzyme (MLE) is similar to that of mandelate racemase.

Knowledge of a protein's structure does not necessarily make it possible to predict its biochemical or cellular functions

Perhaps the most promising case for the prediction of biochemical function from structure is when two proteins show some similarity in amino-acid sequences, share the same overall tertiary structure, and have active sites with at least some residues in common. But consideration of just such an example shows that, even with favorable parameters, function cannot always be deduced from structure.

The bacterium *Pseudomonas ovalis* lives in the soil and scavenges a wide variety of organic compounds for food. One of these is mandelate, a byproduct of decaying fruit pits (Figure 4-38). Mandelate naturally exists as two mirror-image isomers, R- and S-mandelate, and *P. ovalis* can use both as a carbon source. Two enzymes essential for the bacterium to grow on either R- or S-mandelate are mandelate racemase and muconate lactonizing enzyme.

The biochemical function of mandelate racemase (MR) is to interconvert R- and S-mandelate (Figure 4-38); because only S-mandelate is a substrate for the next enzyme in the degradative pathway, this enables *P. ovalis* to metabolize all the available mandelate instead of just half. MR is a metalloenzyme. It requires a magnesium or manganese ion for catalytic activity. Muconate lactonizing enzyme (MLE), which is further along the pathway of mandelate catabolism, transforms the *cis*, *cis*-muconic acid derived from mandelate into muconolactone (Figure 4-39). This is an essential step in the overall breakdown of mandelate into acetyl-CoA, a substrate for energy production via the tricarboxylic acid cycle. MLE is also a metalloenzyme. It requires a manganese ion for activity, although magnesium can be substituted.

The substrates for MR and MLE are very different molecules, and the biochemical functions of these two proteins are also different. Their amino-acid sequences are 26% identical, which falls in the "grey area" where one cannot predict for certain that two proteins will have any domains with a similar fold (see Figure 4-19). Secondary-structure prediction is also uninformative. Nevertheless, when the three-dimensional structures of MR and MLE were determined by X-ray crystallography, they showed that the overall folds were essentially identical (Figure 4-40). Both enzymes are TIM-barrel proteins (see section 1-18) with an extra, mostly antiparallel, beta-sheet domain attached. MR and MLE also bind their catalytic metal ions in the same positions in the structures.

Examination of the active sites shows that this similarity is preserved in detail (Figure 4-41). The amino acids that bind the metal ion are conserved between MR and MLE with one exception, and that is replaced by one with similar physical-chemical properties from a different position in the sequence. Thus, the way these two proteins bind their essential metal ions is structurally and functionally conserved, even though not all the residues involved are in exactly corresponding positions in the two sequences. Both active sites contain a pair of lysine residues in identical positions: in each case, one lysine acts as a catalytic base while the other serves to reduce the pK_a of the first through the proximity of its positive charge (see section 2-12). Opposite the lysine pair, however, the active sites of MR and MLE are different. MR has a second catalytic base, histidine 297, while MLE has a lysine residue, of uncertain role in catalysis, that occupies the same spatial position as histidine 297 but comes from sequence position 273.

The striking similarity between the active sites of MR and MLE, together with their virtually identical folds, implies that these enzymes are homologous, that is, that they diverged from a common ancestor. Nevertheless, they catalyze different chemical reactions on different substrates. Nor is there any conservation of binding specificity between them. Neither R- nor S-mandelate

References

Babbitt, P.C. *et al.*: **A functionally diverse enzyme superfamily that abstracts the alpha protons of carboxylic acids.** *Science* 1995, **267**:1159–1161.

Gerlt, J.A. and Babbitt, P.C.: **Can sequence determine function?** *Genome Biol.* 2000, **1**:reviews0005.1–0005.10.

Gerlt, J.A. and Babbitt, P.C.: **Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies.** *Annu. Rev. Biochem.* 2001, **70**:209–246.

Hasson, M.S. *et al.*: **Evolution of an enzyme active site: the structure of a new crystal form of muconate lactonizing enzyme compared with mandelate racemase and enolase.** *Proc. Natl Acad. Sci. USA* 1998, **95**:10396–10401.

Nagano, N. *et al.*: **One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions.** *J. Mol. Biol.* 2002, **321**:741–765.

Neidhart, D.J. *et al.*: **Mandelate racemase and muconate lactonizing enzyme are mechanistically distinct and structurally homologous.** *Nature* 1990, **347**:692–694.

Petsko, G.A. *et al.*: **On the origin of enzymatic species.** *Trends Biochem. Sci.* 1993, **18**:372–376.

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 170

is a substrate or inhibitor of MLE, nor is *cis*, *cis*-muconate and muconolactone for MR. Evolutionary pressure for MR and MLE to become highly specific for their respective substrates led their substrate-binding pockets to become very different, and mutually incompatible, while the catalytic machinery remained similar. This argument implies an underlying commonality of catalytic mechanism between MR and MLE, as reflected in the conserved residues in their active sites. Both enzymes use a base—one of the pair of lysines—to abstract a hydrogen attached to a carbon atom in the substrate (see Figures 4-38 and 4-39), and in each case that carbon atom is adjacent to a carboxylate group that is coordinated to the metal ion in the active site (Figure 4-41). Presumably, the ancestral protein could carry out this chemical step on either mandelate or muconolactone or on some related molecule.

So, even if two gene products have similar sequences and share the same overall fold and some active-site residues, and even if the biochemical function of one of them is known, it is not always possible to predict the biochemical function of the other. Missing is knowledge of what molecules interact specifically with the active site of the protein of unknown function. Computational methods to determine which small molecules would bind to an active site of known shape and charge distribution do not yet exist (see section 4-9).

This particular example is not unique. MR and MLE belong to a large superfamily of TIM-barrel enzymes. All members of this superfamily have the additional beta-sheet domain, use a divalent metal ion, and have metal-binding residues in positions corresponding to those in MR and MLE. They catalyze chemical reactions as diverse as the dehydration of the sugar D-galactonate and the formation of phosphoenolpyruvate. For each of these enzymes a catalytic mechanism can be written involving base-catalyzed abstraction of a hydrogen from a carbon atom adjacent to a carboxylate group, but all of the substrates are different. A total of 21 different superfamilies of TIM-barrel enzymes have been identified on the basis of structural and functional relatedness. These 21 superfamilies include 76 difference sequence families.

Nor is the TIM barrel the only domain fold with this sort of versatility. The four-helix bundle has been found in hormones, growth factors, electron-transport proteins and enzymes. The zinc finger motif is usually found in DNA-binding proteins, where it makes sequence-specific contacts with bases in the double helix, but there are zinc finger domains that bind to RNA instead, and a number of proteins have zinc finger modules that mediate protein–protein interactions. Many other examples could be cited.

Nevertheless, there are a number of domain folds that are characteristic of a given biochemical or cellular function, and in some instances the presence of such a domain can be recognized from sequence information alone, allowing one to determine at least partial biochemical function from sequence and/or structural information. For instance, the kinase fold appears to be present almost exclusively in protein kinases; the SH2 domain appears to be used exclusively to bind phosphotyrosine-containing peptides; and in eukaryotes the seven-transmembrane helix fold appears to be used only in G-protein-coupled receptors. But many folds have so diverse a range of functions that sequence and structural information alone is unlikely to be sufficient to reveal their biochemical or cellular roles.

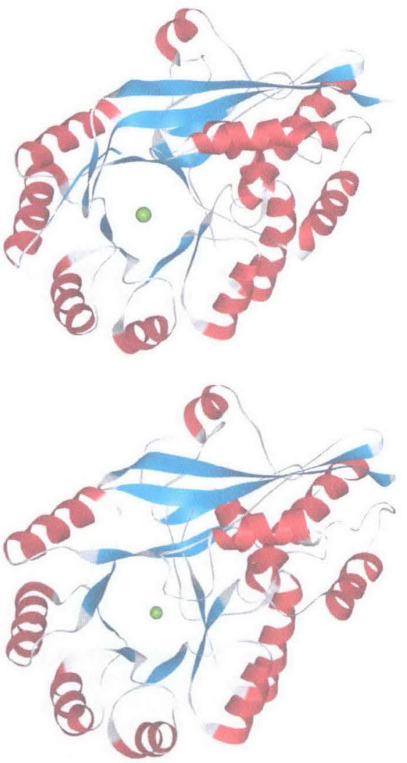


Figure 4-40 Mandelate racemase (top) and muconate lactonizing enzyme (bottom) have almost identical folds The colored spheres in the center of the structures represent the metal ions in the active sites.

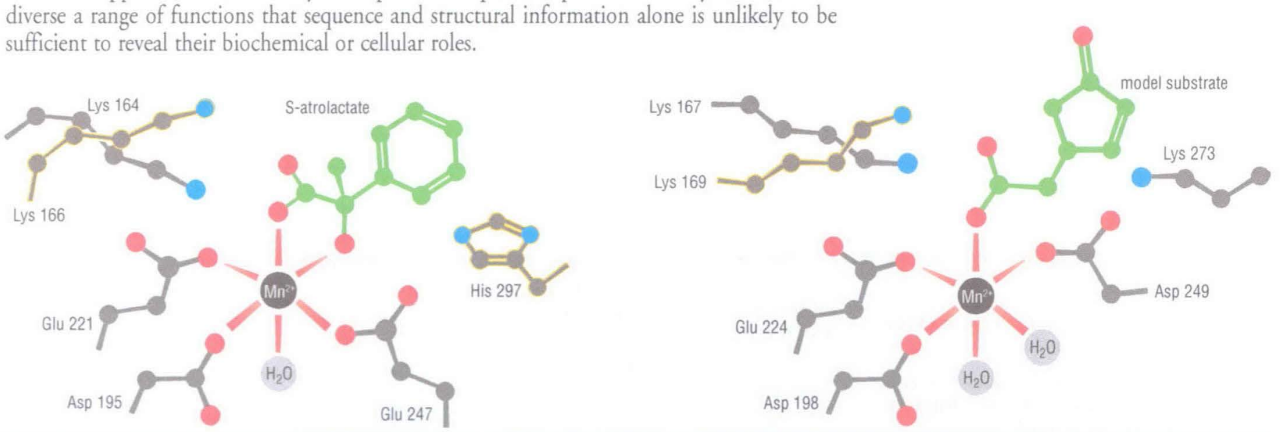


Figure 4-41 A comparison of the active sites of mandelate racemase (left) and muconate lactonizing enzyme (right) The amino acids that coordinate with the metal ion are conserved between the two enzymes, as are the catalytic residues except for histidine 297 in MR which is replaced by lysine 273 in MLE. In both cases a carboxylate group on the substrate coordinates with the metal ion. The active site of MR is shown with the inhibitor S-atrolactate bound; the MLE active site is shown with a model substrate bound. The residues shaded in yellow are the putative general acid-base catalytic residues.

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 171

4-12 PLP Enzymes: Diverse Structures with One Function

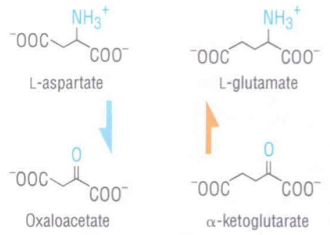


Figure 4-42 The overall reaction catalyzed by the pyridoxal phosphate-dependent enzyme L-aspartate aminotransferase

A protein's biochemical function and catalytic mechanism do not necessarily predict its three-dimensional structure

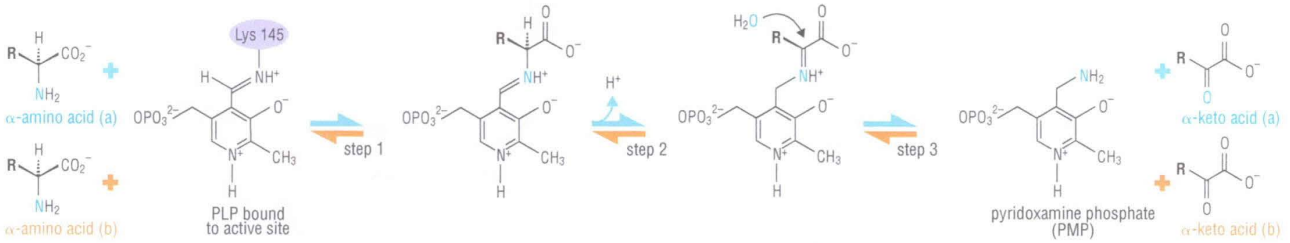
The site of biochemical function in an enzyme is characterized by a sub-site that binds the substrate and a catalytic sub-site at which the chemical reaction takes place, and these two sub-sites are usually at least partly distinct (see section 2-7). It is therefore possible to have the same arrangement of catalytic groups in combination with different arrangements of substrate-binding groups. In some cases, this produces enzymes with different specificities but which carry out the same chemistry. Such a situation is usually associated with divergent evolution from a common ancestor, in which the protein scaffold is retained but the substrate-binding sub-site is altered. It is also possible, however, for the same catalytic machinery to evolve independently on different protein scaffolds, whose substrate-binding sites may, or may not, be specific for the same substrate. This is termed convergent evolution: nature has found the same solution to the problem of catalyzing a particular reaction, but the solution has evolved in two different protein frameworks.

A prime example of convergent evolution is found among the aminotransferases. These are enzymes that "convert" one amino acid into another in a reaction known as transamination, which is central to amino-acid metabolism. In this reaction, an α-amino acid is converted to an α-keto acid, followed by conversion of a different α-keto acid to a new α-amino acid. All transaminases use the cofactor pyridoxal phosphate (PLP), derived from vitamin B_6. One of these PLP-dependent enzymes is L-aspartate aminotransferase, which converts L-aspartate to L-glutamate, via α-ketoglutarate and oxaloacetate (Figure 4-42) and is found in every living organism. Another aminotransferase, found only in bacteria, catalyzes the same reaction but is specific for the D-forms of various amino acids, including aspartate and glutamate. In bacteria, a transamination reaction involving D-glutamate, α-ketoglutarate, and pyruvate is used to produce D-alanine for synthesis of the bacterial cell wall.

Figure 4-43 The general mechanism for PLP-dependent catalysis of transamination, the interconversion of α-amino acids and α-keto acids The amino group of the amino-acid substrate displaces the side-chain amino group of the lysine residue that holds the cofactor PLP in the active site (step 1). PLP then catalyzes a rearrangement of the amino-acid substrate (step 2), followed by hydrolysis of the keto-acid portion, leaving the nitrogen of the amino acid (blue) bound to the cofactor to form the intermediate pyridoxamine phosphate (PMP) (step 3). This forward reaction is indicated by the blue arrows. To regenerate the starting form of the enzyme, a different keto acid then reverses these steps and captures the bound nitrogen, producing a new amino acid and leaving the PLP once more bound to the enzyme at the active-site lysine (orange arrows).

Both enzymes have identical catalytic mechanisms. The amino group of the amino-acid substrate displaces the side-chain amino group of the lysine residue that binds the cofactor PLP in the active site (Figure 4-43, step 1). PLP then catalyzes a rearrangement of its new bound amino acid (step 2), followed by hydrolysis of the keto-acid portion, leaving the nitrogen of the amino acid bound to the cofactor (step 3). To regenerate the starting form of the enzyme, a different keto acid then reverses these steps and captures the bound nitrogen, producing a new amino acid and leaving the PLP once more bound to the enzyme at the active-site lysine.

A number of other amino-acid side chains in the active sites of both of these aminotransferases interact specifically with the cofactor, promoting this series of transformations over the other possible chemistries that the versatile cofactor PLP can catalyze. Yet other amino acids stabilize the position of the bound substrates and confer substrate specificity. Because these two enzymes catalyze exactly the same reaction by exactly the same mechanism, one might expect their structures to be similar. On the other hand, because one of these enzymes is specific for the commonly



References

Doolittle, R.F.: **Convergent evolution: the need to be explicit.** *Trends Biochem. Sci.* 1994, **19**:15–18.

Kirsch, J.F. *et al.*: **Mechanism of action of aspartate aminotransferase proposed on the basis of its spatial structure.** *J. Mol. Biol.* 1984, **174**:497–525.

Smith, D.L. *et al.*: **2.8-Å-resolution crystal structure of an active-site mutant of aspartate aminotransferase from *Escherichia coli*.** *Biochemistry* 1989, **28**:8161–8167.

Sugio, S. *et al.*: **Crystal structure of a D-amino acid aminotransferase: how the protein controls stereoselectivity.** *Biochemistry* 1995, **34**:9661–9669.

Yennawar, N. *et al.*: **The structure of human mitochondrial branched-chain aminotransferase.** *Acta Crystallogr. D. Biol. Crystallogr.* 2001, **57**:506–515.

Yoshimura, T. *et al.*: **Stereospecificity for the hydrogen transfer and molecular evolution of pyridoxal enzymes.** *Biosci. Biotechnol. Biochem.* 1996, **60**:181–187.

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 172

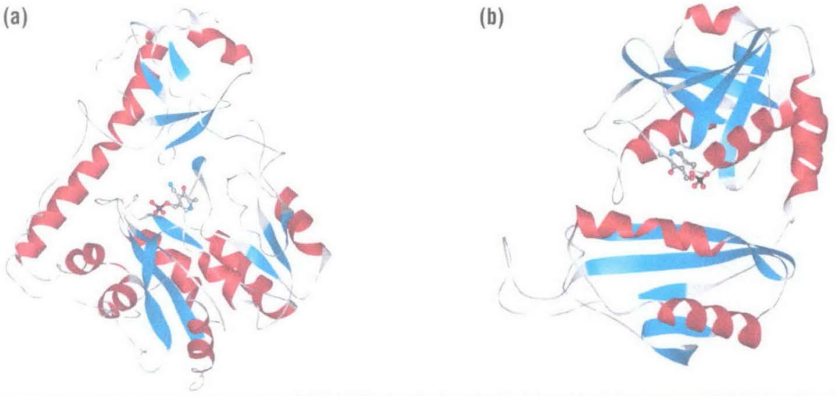


Figure 4-44 The three-dimensional structures of L-aspartate aminotransferase (left) and D-amino acid aminotransferase (right) The two proteins have completely different architectures. Not only are they different in size, they differ in their amino-acid sequence and in the folds of the protein domains. In the L-aspartate aminotransferase structure, the cofactor intermediate PMP (in a ball-and-stick representation) is shown at the active site. In the D-amino acid aminotransferase, the cofactor PLP is shown. (PDB 2aat, 1daa).

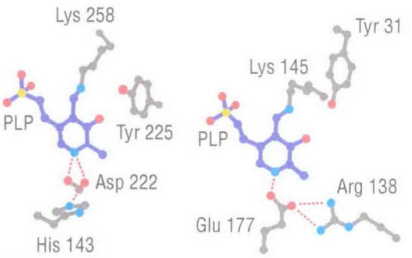


Figure 4-45 Comparison of the active sites of L-aspartate aminotransferase (left) and D-amino acid aminotransferase (right) Despite the different protein folds of these two enzymes, the active sites have converged to strikingly similar arrangements of the residues that interact with the cofactor and promote catalysis. The residues that determine which amino acid is used as a substrate and whether it will be the D- or the L-form are arranged differently in the two enzymes (not shown).

occurring L-forms of the amino acids while the other exclusively uses the rarer D-enantiomers, one might also expect that their active sites would look quite different in terms of the arrangement of side chains around the cofactor. Both these expectations are, however, far from the case.

Comparison of the amino-acid sequences of the two enzymes reveals absolutely no identity; however, as we have seen, the absence of detectable sequence identity does not necessarily mean that the protein fold will be different (see section 1-16). But in this case, comparison of the three-dimensional structures of the two enzymes shows their polypeptide-chain folds to be totally different (Figure 4-44). They clearly did not evolve from a common ancestor. When one finds two sequences and two structures that are completely different, one might on the face of it expect that they represent different mechanisms for solving the problem of catalyzing the same chemical transformation. When these two structures are examined in detail, however, the active sites are found to be strikingly similar, both in the nature of the amino acids interacting with the cofactor and their positions in space (Figure 4-45). Moreover, detailed analysis of genomic sequence data suggests that all known aminotransferases possess one or other of these polypeptide-chain folds and this same active-site configuration. It appears that this constellation of catalytic groups, in combination with the intrinsic chemistry of the PLP cofactor, is especially suited to promoting transamination, and nature has independently discovered this twice, using two different protein frameworks.

One possible explanation for the differences in three-dimensional framework is the difference in the "handedness" (or chirality) of the substrate: perhaps one fold is only suited to recognizing D-amino acids. This cannot be so, however, because there is another aminotransferase that only recognizes L-amino acids but whose sequence and polypeptide chain fold are similar to those of D-amino acid aminotransferase (Figure 4-46). These two enzymes clearly represent divergent evolution from a common ancestor. Apparently, modification of substrate specificity within the context of a given protein fold, even to the extent of reversing the handedness of the substrate, is easier than evolving a completely new catalytic mechanism.

A number of biochemical functions are carried out by enzymes that differ in their protein fold but have remarkably similar active sites. Convergent evolution to a common chemical mechanism has been observed among the serine proteases, the aminopeptidases, the NAD-dependent dehydrogenases and the sugar isomerases, to name just a few. Consequently, even if you know the biochemical function of a newly discovered protein, you cannot necessarily predict the protein fold that will carry it out. The catalytic function of enzymes can, however, sometimes be predicted by genomic analysis aimed at identification of patterns of active-site residues, and we discuss this in section 4-2.

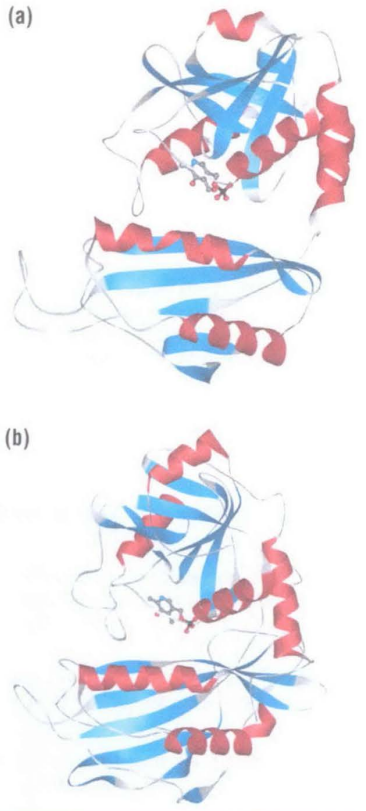


Figure 4-46 The three-dimensional structures of bacterial D-amino acid aminotransferase (top) and human mitochondrial branched-chain L-amino acid aminotransferase (bottom) These two enzymes are similar in amino-acid sequence, overall fold and active site even though their substrates are of opposite handedness. They have diverged from a common ancestor. (PDB 1daa, 1ekp).

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 173

4-13 Moonlighting: Proteins with More than One Function

In multicellular organisms, multifunctional proteins help expand the number of protein functions that can be derived from relatively small genomes

The genomes of multicellular organisms are remarkably small, in terms of number of genes (see Figure 4-10), considering the enormous increase in complexity of the organisms themselves compared with bacteria or the single-celled eukaryotes. One explanation is that the actual number of different proteins derived from a given gene can be expanded by mechanisms such as alternative splicing. An additional explanation is that, in multicellular organisms especially, a given protein may have more than one distinct biochemical and/or cellular function. The biochemical functions may include catalysis, binding, participating as a structural molecule in an assembly, or operating as a molecular switch (see sections 1-0 and 2-0). The extent of this functional diversity for any one protein is only just beginning to be appreciated.

Phosphoglucose isomerase (glucose-6-phosphate isomerase or PGI) is the second enzyme in the glycolytic pathway, a core metabolic pathway that converts glucose to pyruvate, and in which PGI converts glucose 6-phosphate to fructose 6-phosphate. The gene for PGI is thus a housekeeping gene found in nearly all organisms. Sequences of PGI from numerous organisms are well conserved, indicating that this intracellular biochemical function is catalyzed by a single type of polypeptide fold. But if one takes the protein sequence of PGI from a rabbit, say, and looks for homologous sequences in databases of other mammalian protein sequences, one finds, in addition to PGI, an identical sequence labeled neuroleukin. The protein neuroleukin was discovered as a cytokine secreted by T cells that promotes the survival of some embryonic spinal neurons and sensory nerves. It also causes B cells to mature into antibody-secreting cells. One also finds two other named protein activities with sequences identical to PGI: autocrine motility factor (AMF) and differentiation and maturation mediator (DMM). Like neuroleukin, these proteins are also secreted cytokines. AMF is produced by tumor cells and stimulates cancer-cell migration; it may be involved in cancer metastasis. DMM is isolated from culture medium in which T cells have been grown and has been shown to cause differentiation of human myeloid leukemia cells *in vitro*. Purified rabbit PGI will cause the increase in cell

Multifunctional Proteins

Protein	Function	Additional functions
Phosphoglucose isomerase	Glycolytic enzyme	Cytokine
EF-1	Elongation factor in translation	Actin-bundling protein
Cyclophilin	Peptidyl-prolyl *cis–trans* isomerase	Regulator of calcineurin
Macrophage inhibitory factor (MIF)	Activator of macrophages and T cells	Phenylpyruvate tautomerase
PutA	Proline dehydrogenase	Transcriptional repressor
Aconitase	TCA-cycle enzyme	Iron-responsive-element binding protein
Thioredoxin	Maintains SH groups in reduced state	Subunit of phage T7 DNA polymerase
Thrombin	Protease in blood clotting	Ligand for cell-surface receptor
Thymidylate synthase	Enzyme in DNA synthesis	Inhibitor of translation
FtsH	Chaperone protein in bacteria	Metalloprotease
LON	Mitochondrial protease	Chaperone protein
Methionine aminopeptidase 2	Peptidase	Protects eIF2 from phosphorylation

Figure 4-47 Some examples of multifunctional proteins with their various functions The first function column lists the biochemical function that was first identified. In most cases, this is an enzymatic activity because such activities are easily assayed. The additional functions usually depend on binding to a specific partner.

References

Cutforth, T. and Gaul, U.: **A methionine aminopeptidase and putative regulator of translation initiation is required for cell growth and patterning in *Drosophila*.** *Mech. Dev.* 1999, **82**:23–28.

Datta, B.: **MAPs and POEP of the roads from prokaryotic to eukaryotic kingdoms.** *Biochimie* 2000, **82**:95–107.

Griffith, E.C. *et al.*: **Methionine aminopeptidase (type 2) is the common target for angiogenesis inhibitors AGM-1470 and ovalicin.** *Chem. Biol.* 1997, **4**:461–471.

Haga, A. *et al.*: **Phosphohexose isomerase/autocrine motility factor/neuroleukin/maturation factor is a multifunctional phosphoprotein.** *Biochim. Biophys. Acta* 2000, **1480**:235–244.

Jeffery, C.J.: **Moonlighting proteins.** *Trends Biochem. Sci.* 1999, **24**:8–11.

Jeffery, C.J. *et al.*: **Crystal structure of rabbit phosphoglucose isomerase, a glycolytic enzyme that moonlights as neuroleukin, autocrine motility factor, and differentiation mediator.** *Biochemistry* 2000, **39**:955–964.

Liu, S. *et al.*: **Structure of human methionine aminopeptidase-2 complexed with fumagillin.** *Science* 1998, **282**:1324–1327.

Lubetsky, J.B. *et al.*: **The tautomerase active site of macrophage migration inhibitory factor is a potential target for discovery of novel anti-inflammatory agents.** *J. Biol. Chem.* 2002, **277**:24976–24982.

Sun, Y.J. *et al.*: **The crystal structure of a multifunctional**

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 174

motility seen with AMF and the dosage-dependent differentiation of human leukemia cells seen with DMM; conversely, both AMF and DMM have PGI activity. PGI, neuroleukin, AMF and DMM are the same protein, encoded by the same gene.

To carry out its cytokine and growth-factor functions it is likely that PGI/neuroleukin/AMF/DMM binds to at least one type of cell-surface receptor on a variety of target cells, and a receptor corresponding to AMF activity has been cloned from fibrosarcoma cells. Inhibitors of the PGI reaction block some, but not all, of the cytokine functions of the protein, indicating that the sites on the protein surface responsible for these different activities are at least partly distinct. Remarkably, PGI from a bacterium has been reported to have activity in the AMF assay.

The PGI reaction is extremely similar to that catalyzed by the glycolytic enzyme triosephosphate isomerase, and the catalytic mechanisms of the two reactions are identical. Nevertheless, the three-dimensional structure of PGI is completely different from that of TIM. Nothing in PGI's sequence or structure, however, suggests its additional cytokine functions; there are no obvious domains with structural similarity to any known cytokine and no sequence motifs suggestive of known growth factors or signal transduction molecules. One concludes that the extracellular cytokine functions of this protein in higher eukaryotes are at least partially independent of each other, and have evolved without gross modification of the ancestral fold.

Now that numerous genome sequences are available, each annotated according to the literature of previous studies of that organism, many other examples of multiple functions for the same protein are being discovered (Figure 4-47). One is methionine aminopeptidase type 2 (MetAP2), which catalyzes the removal of the amino-terminal methionine from the growing polypeptide chain of many proteins in eukaryotes. MetAP2 is the target for the anti-angiogenesis drugs ovalicin and fumagillin, which act by inhibiting this catalytic activity. DNA sequence analysis of MetAP2 genes reveals, as expected, sequence homologies with MetAP2 genes from other organisms but also with various eukaryotic homologs of a rat protein known originally as p67. This intracellular protein protects the alpha-subunit of eukaryotic initiation factor 2 (eIF2) from phosphorylation by its kinases. This activity of p67 is observed in different stress-related situations such as heme deficiency in reticulocytes (immature red blood cells), serum starvation and heat shock in mammalian cells, vaccinia virus infection of mammalian cells, baculovirus infection of insect cells, mitosis, apoptosis, and even possibly during normal cell growth. MetAP2 and p67 are identical proteins encoded by the same gene. Inhibitors of MetAP2 activity do not inhibit the translational cofactor activity of MetAP2, suggesting that the two functions are independent. Some mutations in the MetAP2 gene in *Drosophila* result in loss of ventral tissue in the compound eye as well as extra wing veins, whereas others impair tissue growth. However, it is not clear whether these phenotypes are due to loss of MetAP2's catalytic activity, translational cofactor activity, or both. Another example is the cytokine macrophage inhibitory factor, MIF (Figure 4-48), which also has enzymatic activity.

The term "moonlighting" has been coined to describe the performance of more than one job by the same protein. From the point of view of any one experiment, each job may appear to be the main activity and the other(s) to be the sideline. Consequently, in multicellular organisms especially, knowledge of one function of a gene product does not necessarily mean that all its functions have been determined. This fact has profound consequences for gene knockout experiments (by, for example, antisense RNA, RNA interference (RNAi), or gene disruption) as a means of determining function. The phenotype of a knockout animal or cell may be the result of the loss of all the different functions that a protein can carry out, or may differ in different tissues or under different conditions in which various functions are dominant.

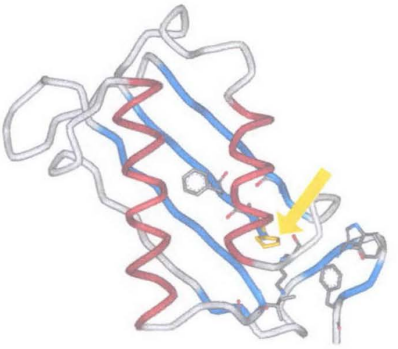


Figure 4-48 The three-dimensional structure of the monomer of macrophage inhibitory factor, MIF The protein is an important proinflammatory cytokine that activates T cells and macrophages. It is also an enzyme that catalyzes the tautomerization of phenylpyruvic acid. The residues involved in substrate binding and catalysis are shown. The proline associated with the active site is indicated by the yellow arrow. Because the active site overlaps with the binding site for receptors of the cytokine function of MIF, inhibitors of its enzymatic activity are also potential antiinflammatory drugs. (PDB 1ljt)

protein: phosphoglucose isomerase/ autocrine motility factor/neuroleukin. *Proc. Natl Acad. Sci. USA* 1999, **96**:5412–5417.

Swope, M.D. and Lolis, E.: **Macrophage migration inhibitory factor: cytokine, hormone, or enzyme?** *Rev. Physiol. Biochem. Pharmacol.* 1999, **139**:1–32.

Watanabe, H. *et al.*: **Tumor cell autocrine motility factor is the neuroleukin/phosphohexose isomerase polypeptide.** *Cancer Res.* 1996, **56**:2960–2963.

Xu, W. *et al.*: **The differentiation and maturation mediator for human myeloid leukemia cells shares homology with neuroleukin or phosphoglucose isomerase.** *Blood* 1996, **87**:4502–4506.

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 175

4-14 Chameleon Sequences: One Sequence with More than One Fold

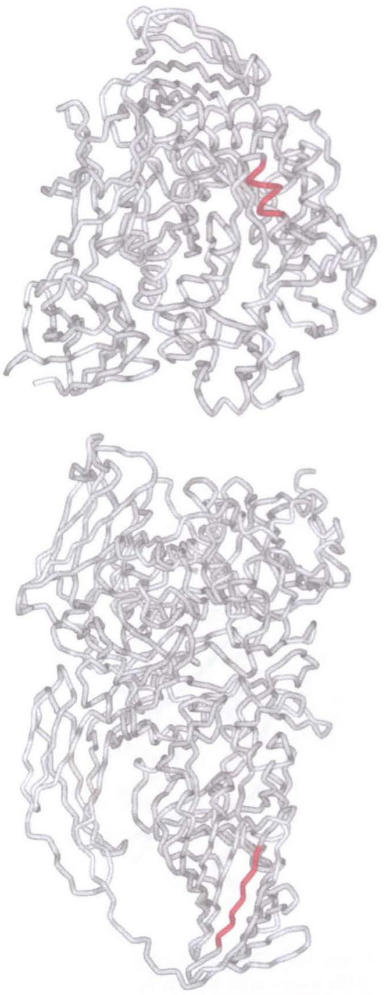


Figure 4-49 Chameleon sequences The protein backbones of the enzymes cyclodextrin glycosyltransferase (PDB 1cgu) (top) and beta-galactosidase (PDB 1bgl) (bottom), each of which contains the chameleon sequence LITTAHA (shown in red), corresponding to residues 121–127 in the sequence of cyclodextrin glycosyltransferase and residues 835–841 in beta-galactosidase. In the former structure, the sequence forms two turns of alpha helix; in the latter, it is a beta strand.

Some amino-acid sequences can assume different secondary structures in different structural contexts

The concept that the secondary structure of a protein is essentially determined locally by the amino-acid sequence is at the heart of most methods of secondary structure prediction; it also underlies some of the computational approaches to predicting tertiary structure directly from sequence. Although this concept appears to be valid for many sequences, as the database of protein structures has grown, a number of exceptions have been found. Some stretches of sequence up to seven residues in length have been identified that adopt an alpha-helical conformation in the context of one protein fold but form a beta strand when embedded in the sequence of a protein with a different overall fold. These sequences have been dubbed **chameleon sequences** for their tendency to change their appearance with their surroundings. One survey of all known protein structures up to 1997 found three such sequences seven residues long (Figure 4-49), 38 such sequences six residues long, and 940 chameleon sequences five residues long. Some were buried and some were on the surface; their sequences varied considerably but there tended to be a preponderance of alanines, leucines and valines and a dearth of charged and aromatic residues.

We have already seen that some segments in certain proteins can change their conformation from, for example, an alpha helix to a loop in response to the binding of a small molecule or another protein or to a change in pH. For example, when elongation factor Tu switches from its GTP-bound form to its GDP-bound form, a portion of the switch helix unravels, breaking an interaction between two domains (see section 3-9).

The ability of amino-acid sequences to convert from an alpha-helical to a beta-strand conformation has received extensive attention recently, as this structural change may induce many proteins to self-assemble into so-called amyloid fibrils and cause fatal diseases (see section 4-15). A number of sequences that are not natural chameleons can become such by a single point mutation, suggesting a possible mechanism whereby such diseases may be initiated. One example is the bacterial protein Fis, a DNA-binding protein that is implicated in the regulation of DNA replication and recombination as well as in transcriptional regulation. A peptide segment in Fis can be converted from a beta strand to an alpha helix by a single-site mutation, proline 26 to alanine. Proline 26 in Fis occurs at the point where a flexible extended beta-hairpin arm leaves the core structure (Figure 4-50a). Thus it can be classified as a "hinge proline" located at the carboxy-terminal end of one beta strand and the amino-terminal cap of the following alpha helix. The replacement of proline 26 with alanine extends the alpha helix for two additional turns in one of the dimeric subunits of Fis; therefore, the structure of the peptide from residues 22 to 26 is converted from a beta strand to an alpha helix by this one mutation (Figure 4-50b). Interestingly, this peptide in the second monomer subunit retains its beta-strand conformation in the crystal structure of Fis, suggesting that the alpha-helical and beta-sheet conformation are very similar in energy for this sequence and that only small local changes in environment are needed to cause it to flip from one form to the other.

While the conversion of a beta strand to an alpha helix in Fis is caused by a mutation, and has no implications for normal function, some proteins contain natural chameleon sequences that may be important to their function. One example is a DNA-binding transcriptional regulator from yeast, the MATα2 protein, which helps determine two differentiated cell types (mating types) in growing yeast cells by repressing genes whose expression is required for one of the two types. MATα2 binds to DNA in association with a second protein, MCM1, so that one copy of MATα2 binds on each side of MCM1 (Figure 4-51a). In the crystal structure of this complex

Definitions

chameleon sequence: a sequence that exists in different conformations in different environments.

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 176

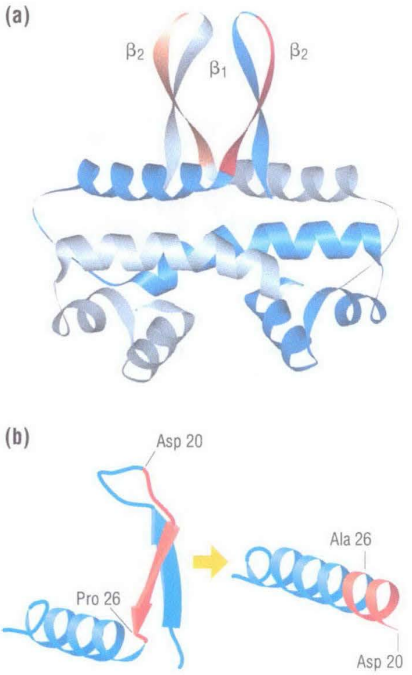


Figure 4-50 Chameleon sequence in the DNA-binding protein Fis (a) The structure of the dimer of the sequence-specific DNA-binding protein Fis shows a predominantly alpha-helical fold with two strands of antiparallel beta sheet, β_1 and β_2 (red), at the amino terminus. (PDB 1f36) **(b)** The replacement of proline 26 at the end of the second beta strand with an alanine converts this beta strand into two additional turns of the alpha helix that follows.

bound to DNA, an eight-amino-acid sequence adopts an alpha-helical conformation in one of two copies of the MATα2 monomer and a beta-strand conformation in the other (Figure 4-51b). Although there is no direct evidence that both forms exist in biology, such an alternative fold could have functional consequences. In most sites the sequences recognized by the MATα2 monomers are identical. However, there are separations of two to three base pairs between the MCM1- and MATα2-binding sites in the natural yeast promoters to which this transcription factor binds, and the different conformations may permit such variations to be tolerated. MATα2 can also form a complex with another transcriptional modulator, MATa1, and in this context, the change in conformation may again allow MATα2 to accommodate differences in the spacing of the sites on DNA. The ability of parts of MATα2 to change conformation in different contexts could help this protein to bind to a number of sites on the genome.

To probe the context dependence of the structures of short polypeptide sequences, an 11-amino-acid chameleon sequence has been designed that folds as an alpha helix when in one position but as a beta sheet when in another position of the primary sequence of the immunoglobulin-binding domain of protein G. This protein from *Staphylococcus aureus* binds to the Fc region of IgG antibodies and is thought to protect the bacteria from these antibodies by blocking their interactions with complement and Fc receptors. Both proteins, chameleon-alpha and chameleon-beta, are folded into structures similar to native protein G except for the small region of the chameleon sequence.

These examples illustrate the general principle that the secondary structures of short peptide segments can often depend more on the tertiary structural context in which they are placed than on their intrinsic secondary structure propensities. The balance between inherent tendency and the effect of environment will be different for different sequences. If the free energies of a peptide in its alpha-helical and beta-sheet conformations are similar, then the energies of interaction between the peptide and the environment could be enough to tip the balance in favor of one or the other.

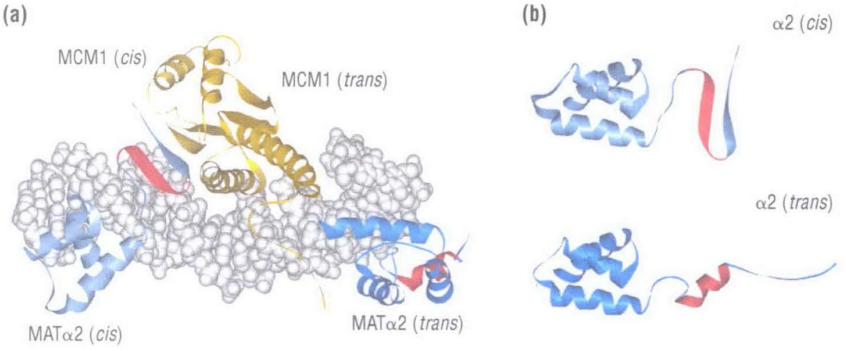


Figure 4-51 Chameleon sequence in the DNA-binding protein MATα2 from yeast (a) The structure of the complex of MATα2 (blue and red) with its transcriptional co-regulator MCM1 (yellow) bound to a target site in DNA. At such a site, two monomers of MATα2 bind to two (usually identical) DNA sequences on either side of two monomers of MCM1. **(b)** An eight-amino-acid sequence (red) adopts a beta-strand conformation in one MATα2 molecule (the *cis* monomer; light blue) and an alpha-helical conformation in the other (the *trans* monomer; dark blue). (PDB 1mnm)

References

Mezei, M.: **Chameleon sequences in the PDB.** *Protein Eng.* 1998, **11**:411–414.

Minor, D.L. Jr. and Kim, P.S.: **Context-dependent secondary structure formation of a designed protein sequence.** *Nature* 1996, **380**:730–734.

Smith, C.A. *et al.*: **An RNA-binding chameleon.** *Mol. Cell* 2000, **6**:1067–1076.

Sudarsanam, S.: **Structural diversity of sequentially identical subsequences of proteins: identical octapeptides can have different conformations.** *Proteins* 1998, **30**:228–231.

Tan, S. and Richmond T.J.: **Crystal structure of the yeast MATalpha2/MCM1/DNA ternary complex.** *Nature* 1998, **391**:660–666.

Yang, W.Z. *et al.*: **Conversion of a beta-strand to an alpha-helix induced by a single-site mutation observed in the crystal structure of Fis mutant Pro26Ala.** *Protein Sci.* 1998, **7**:1875–1883.

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 177

4-15 Prions, Amyloids and Serpins: Metastable Protein Folds

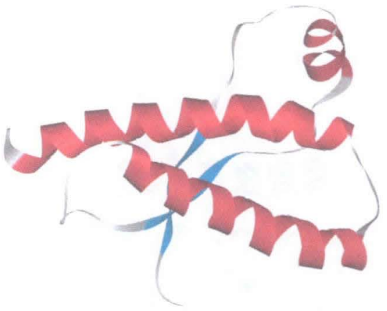


Figure 4-52 The prion protein This figure shows the soluble form of Syrian hamster prion protein PrPSc, which was generated by removing the amino terminus. This protein consists of residues 90–231, incorporating the region thought to be involved in the structural change. The protein was generated by refolding to produce the structure of the cellular form of the protein. The precise structure of the disease-causing form is not yet known, but is known to have much more beta sheet than the cellular form.

A single sequence can adopt more than one stable structure

The existence of chameleon sequences may reflect a general principle: that not all sequences fold into one unique structure. Some structures may be **metastable**—able to change into one or more different stable structures. But are complete protein sequences of such plasticity found in nature? It appears that the answer is yes although rarely, and that such cases are often, at least so far, associated with severe mammalian disease. The best characterized of these changeable structures is the prion associated with Creutzfeldt-Jakob disease (CJD) in humans and scrapie and bovine spongiform encephalopathy (BSE) in sheep and cattle (Figure 4-52). Prions are infectious proteins whose misfolded form is identical in sequence to the normal cellular form of the same protein. The two forms have, however, quite different conformations and physical properties. The infectious form has a propensity to form aggregrates, possesses secondary structure content that is rich in beta sheets, is partially resistant to proteolysis, and is insoluble in nonionic detergents. In contrast, the cellular form contains little beta structure, is sensitive to protease digestion, and is soluble in nonionic detergent. Contact with the infectious form causes the cellular counterpart to undergo pronounced conformational changes that lead, ultimately, to the formation of cytotoxic protein aggregates consisting almost entirely of the infectious conformation. Although the molecular events that lead to this profound conformational change are poorly understood, there is substantial evidence that the infectious form acts as a template directing the structural rearrangement of the normal form into the infectious one. Studies of synthetic peptides derived from the prion sequence indicate that a stretch of up to 55 residues in the middle of the protein has the propensity to adopt both alpha-helical and beta-sheet conformations. Presumably, the infectious form arises spontaneously in a small number of molecules as a result of this inherent plasticity.

A similar mechanism may underlie protein aggregate formation in a group of about 20 diseases called amyloidoses, which include Alzheimer's, Parkinson's, and type II diabetes. Each disease is associated with a particular protein, and extracellular aggregates of these proteins are thought to be the direct or indirect origin of the pathological conditions associated with the disease. Strikingly, the so-called amyloid fibrils characteristic of these diseases arise from well known proteins, including lysozyme and transthyretin, that have well defined, stable, non-identical folds but produce fibrous protein aggregates of identical, largely beta-sheet, structure (Figure 4-53).

Recent studies suggest that the ability to undergo a refolding leading to amyloid formation is not unique to these proteins, but can be observed in many other proteins under laboratory conditions such as low pH. One conclusion from such findings is that prions and other proteins that cause disease by this mechanism may differ from the vast array of "normal" cellular proteins only in having sequences that can undergo such refolding spontaneously under physiological conditions. Clearly, if this is the case, it is possible that a single point mutation may convert a harmless protein into one that can refold spontaneously; such mutations have been found associated with some of the amyloidoses.

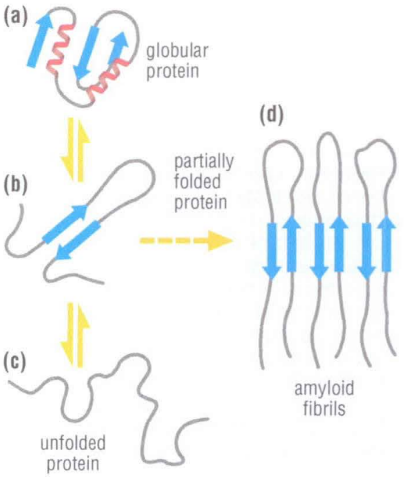


Figure 4-53 A possible mechanism for the formation of amyloid fibrils by a globular protein The correctly folded protein **(a)** is secreted from the cell. Under certain conditions, or because it contains a mutation, the protein unfolds partially **(b)** or completely **(c)**; the unfolded forms can also refold partially or completely. The partially unfolded form is prone to aggregation, which results in the formation of fibrils **(d)** and other aggregates that accumulate in the extracellular space.

Definitions

metastable: only partially stable under the given conditions. In the case of protein structures, a metastable fold exists in equilibrium with other conformations or with the unfolded state.

References

Cohen, B.I. *et al.*: **Origins of structural diversity within sequentially identical hexapeptides.** *Protein Sci.* 1993, **2**:2134–2145.

Dalal, S. and Regan, L.: **Understanding the sequence determinants of conformational switching using protein design.** *Protein Sci.* 2000, **9**:1651–1659.

Dobson, C.M.: **Protein misfolding, evolution and disease.** *Trends Biochem. Sci.* 1999, **24**:329–332.

Engh, R.A. *et al.*: **Divining the serpin inhibition mechanism: a suicide substrate "springe"?** *Trends Biotechnol.* 1995, **13**:503–510.

Kabsch, W. and Sander, C.: **On the use of sequence homologies to predict protein structure: identical pentapeptides can have completely different conformations.** *Proc. Natl Acad. Sci. USA* 1984, **81**:1075–1078.

Ko, Y.H. and Pedersen, P.L.: **Cystic fibrosis: a brief look at some highlights of a decade of research focused on elucidating and correcting the molecular basis of**

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 178

Structural plasticity can also be part of the normal function of a protein. Often, when this is the case, ligand binding or specific proteolytic modification is needed to drive one folded form into the other. Large, ligand-induced domain rearrangements such as that found in elongation factor Tu (see section 3-9) can be considered examples of different overall protein folds induced by the state of assembly. Perhaps the best example of a large structural rearrangement caused by limited proteolysis is found in the family of protein protease inhibitors called the serpins. Some protein protease inhibitors function as rigid substrate mimics; the serpins differ fundamentally in that the loop that recognizes and initially binds the protease active site is flexible and is cleaved by the protease. The serpin remains bound to the enzyme but the cleavage triggers a refolding of the cleaved structure that makes it more stable: one segment of the cleaved loop becomes the central strand of an existing beta sheet in the center of the serpin, converting it from a mixed beta sheet to a more stable antiparallel form (Figure 4-54); in at least one case, the downstream segment also becomes the edge strand of another beta sheet. If the cleaved serpin is released it cannot reassociate with the protease because this refolding has made the recognition strand unavailable for binding.

Mutations in serpins leading to misfolding and aggregation have also been found in some human diseases. For example, the so-called Z-variant of the serpin alpha$_1$-antitrypsin (in which glutamic acid 342 is mutated to lysine) is retained within hepatocytes as inclusion bodies; this is associated with neonatal hepatitis and cirrhosis. The inclusion bodies form because the mutation perturbs the conformation of the protein, facilitating a sequential interaction between the recognition loop of one molecule and beta-sheet A of a second; this could be thought of as a pathological case of domain swapping (see section 2-4).

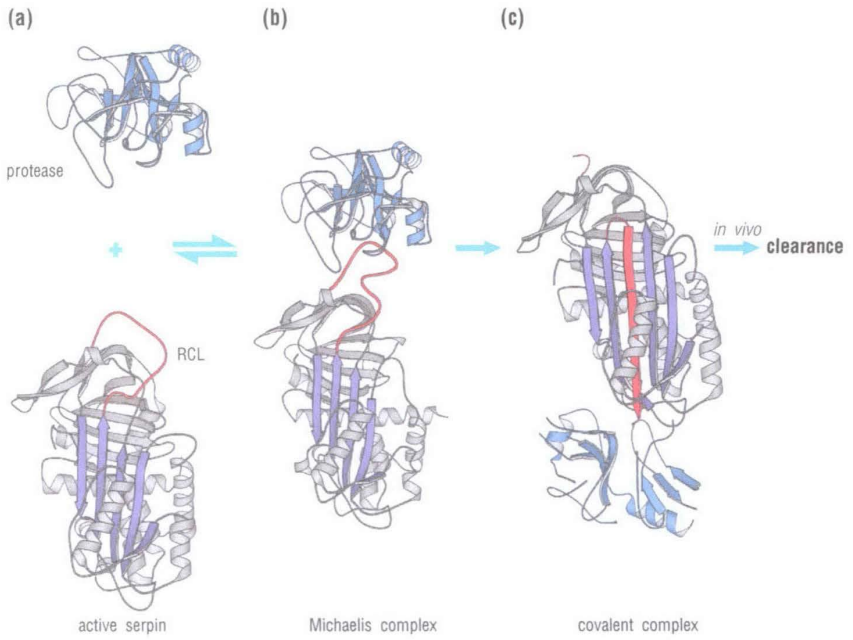


Figure 4-54 Structural transformation in a serine protease inhibitor on binding protease (a) The exposed reactive center loop (RCL; red) of the serpin alpha$_1$-antitrypsin (grey) is shown binding to the target protease (blue) to form an enzyme–substrate complex, the Michaelis complex **(b)**. The RCL of the serpin is then cleaved by the target protease, leading to the insertion of the unconstrained RCL into the serpin beta sheet and the formation of a covalent complex that is trapped by release of the newly formed amino terminus of the serpin **(c)**. This complex is then targeted for clearance. Adapted from Ye, S. and Goldsmith, E.J.: *Curr. Opin. Struct. Biol.* 2001, **11**:740–745.

the disease. *J. Bioenerg. Biomembr.* 2001, **33**:513–521.

Leclerc, E. *et al.*: **Immobilized prion protein undergoes spontaneous rearrangement to a conformation having features in common with the infectious form.** *EMBO J.* 2001, **20**:1547–1554.

Mahadeva, R. *et al.*: **6-mer peptide selectively anneals to a pathogenic serpin conformation and blocks polymerization. Implications for the prevention of Z alpha(1)-antitrypsin-related cirrhosis.** *J. Biol. Chem.* 2002, **277**:6771–6774.

Ye, S. and Goldsmith, E.J.: **Serpins and other covalent protease inhibitors.** *Curr. Opin. Struct. Biol.* 2001, **11**:740–745.

Zahn, R. *et al.*: **NMR solution structure of the human prion protein.** *Proc. Natl Acad. Sci. USA* 2000, **97**:145–150.

©2004 New Science Press Ltd

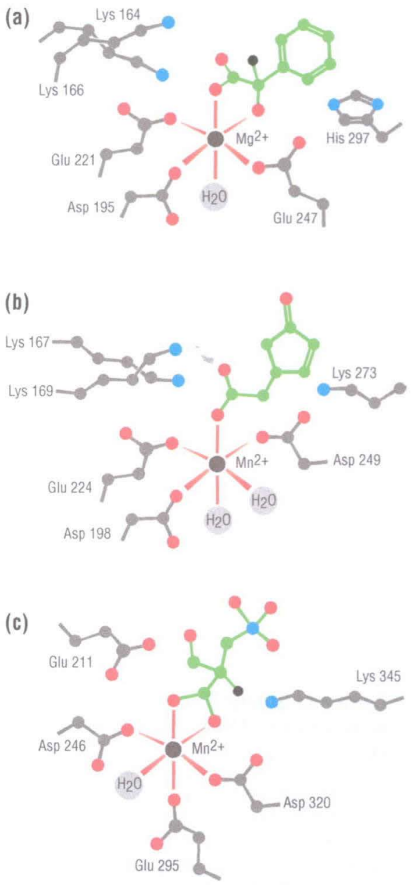


Figure 4-55 Active sites of MR, MLE, and enolase Schematic diagrams of the arrangements of the active-site residues of **(a)** mandelate racemase (MR), **(b)** muconate lactonizing enzyme (MLE) and **(c)** enolase. The types of amino acid that coordinate the divalent metal ion are conserved between the three enzymes. The other catalytic residues, however, are conserved neither in exact position nor in chemical type; nevertheless, these various residues can carry out similar chemistry. In each reaction, the carboxy group of the substrate forms a ligand to the metal ion in the active site, facilitating abstraction of a proton. The enolic intermediate resulting from this common core step is stabilized by interactions with other electrophilic groups in the active site. These groups differ among the three enzymes. The rest of the substrate-binding pocket differs considerably among the three enzymes. (PDB 1mns, 1muc and 1one)

Determining biochemical function from sequence and structure becomes more accurate as more family members are identified

The identification of the function of the *E. coli* protein f587, known originally only as an uncharacterized open reading frame in the *E. coli* genome sequence, is an illustration of the importance of lateral thinking. In this case, the sequence and structural information was eventually interpreted in the light of the known genetics and physiology of the bacterium.

The story starts with the related enzymes mandelate racemase (MR) and muconate lactonizing enzyme (MLE) (see section 4-11). The enzyme enolase, which catalyzes the conversion of 2-phospho-D-glycerate to phosphoenolpyruvate, was subsequently found to have a similar degree of sequence identity to MR and MLE (26%) and a similar distribution of conserved residues. It has the same polypeptide chain fold and many of the conserved residues map to the active-site region of the fold. Together, these three enzymes form part of the so-called enolase superfamily.

With three related sequences and structures in hand, it became apparent that not every part of the active site was preserved (Figure 4-55). All three enzymes require at least one divalent metal ion for activity, and the carboxylate ligands to these metal ions are present in all three proteins; in enolase however, glutamic acid and aspartic acid are substituted for one another. The remainder of the catalytic machinery is even more divergent in enolase. Yet, the residues are conserved in terms of their broad chemical role although not in terms of their identity. Glutamic acid 211 in enolase could in principle act as a general acid-base group, just like lysine 166 in MR and lysine 169 in MLE. On the other side of the substrate-binding pocket, lysine 345 in enolase occupies the same position as lysine 273 in MLE; the corresponding position in the structure of MR is occupied by histidine 297, which might have a different role. All three enzymes use their bound metal ions in the same way—to activate a C–H bond adjacent to a carboxylate group for abstraction of the hydrogen by a base on the enzyme.

MR has two substrates with the C–H bond in different positions (see Figure 4-38) and its function is to interconvert them; thus, two different acid-base groups are needed. In the case of enolase and MLE, only one proton needs to be abstracted and so only a single base is required. Thus, both lysine 166 and histidine 297 in MR function as acid-base groups. In MLE and enolase, however, the abstraction of a single proton is carried out only by lysine 169 and lysine 345, respectively. This conserved base-catalyzed, metal-promoted proton-transfer step is the common function linking all three enzymes.

Alignments based on conservation of residues that carry out the same active-site chemistry can identify more family members than sequence comparisons alone

If chemistry is the conserved feature, rather than the absolute identity and position in the sequence of the groups that carry out the chemistry, then a more sophisticated approach to finding homologous sequences would be to search for patterns of residues that can perform the same chemistry regardless of their specific amino-acid identities. Such a search, using a specialized computer program, identified dozens of potential members of the enolase/MR/MLE superfamily, most of which could not have been detected by conventional sequence comparison.

One of these predicted homologs, open reading frame f587 in the *E. coli* genome, coded for a protein of unknown function. Alignment of the f587 sequence with those of the other three proteins on the basis of the conservation of active-site chemical function showed that f587 contains the requisite metal-ion ligands and conservation of an active-site histidine—histidine 285—which aligns with histidine 297 of MR. From the position of this base, the prediction would be that the substrate for f587, whatever it might be, would be a carboxylate-containing molecule with a proton on an adjacent carbon that has the R-configuration, as does R-mandelate.

In well studied model organisms, information from genetics and cell biology can help identify the substrate of an "unknown" enzyme and the actual reaction catalyzed

The remaining problems—what is the substrate and overall reaction of f587—are insoluble from sequence and structural information alone. For f587, however, additional information was available. *E. coli*, like most bacteria, tends to organize its genes into operons encoding a set

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 180

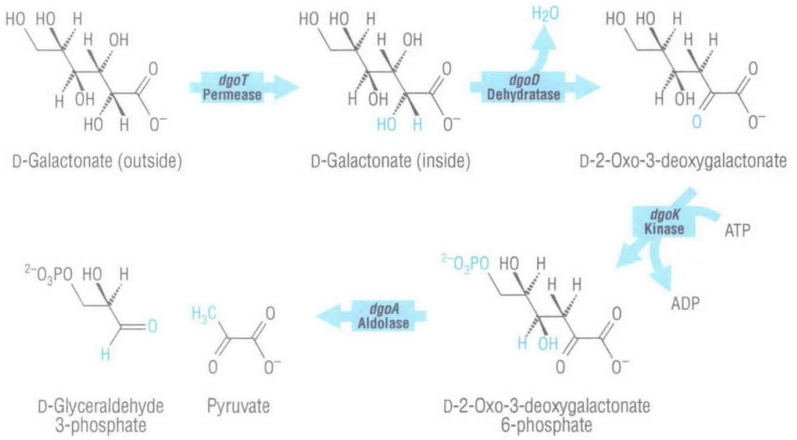


Figure 4-56 The pathway for the utilization of galactonate in *E. coli* D-galactonate is transported into the cell, dehydrated, phosphorylated, and then cleaved to produce glyceraldehyde 3-phosphate and pyruvate. This pathway allows the bacterium to grow on galactonate as a sole carbon source. The names of the genes and the enzyme activity they represent are given in the blue arrows. F587 has now been identified as the gene *dgoD*, encoding galactonate dehydratase.

of proteins that act in the same pathway. The genome sequence suggested that f587 is the fourth in a series of five open reading frames that could constitute a single operon. The other open reading frames were known to encode proteins involved in galactonate metabolism. They appear to be part of the pathway in which galactonate is imported into the cell by galactonate permease and then degraded stepwise into glyceraldehyde 3-phosphate and pyruvate. Given the proposed functions of the other proteins, the only function needed to complete this pathway was galactonate dehydratase (GalD) (Figure 4-56).

Dehydration of galactonate could be catalyzed by abstracting the 2-R proton via a catalytic base with assistance of a metal ion to activate the C–H bond. From its sequence, f587 appeared to contain all the necessary elements in the right stereochemical configuration to catalyze just such a reaction. To test the hypothesis that f587 encodes GalD, a cell-free extract of *E. coli* transformed with a plasmid overexpressing f587 was assayed for GalD activity. An alpha-keto acid was produced from galactonate, but not from other similar sugars. Subsequent purification of the f587 gene product confirmed it as GalD. Finally, the crystal structure of GalD was determined with an analog of galactonate bound. The overall polypeptide chain fold (Figure 4-57) is the same as that of MR, MLE, and enolase and the active site also resembles those of the other family members (Figure 4-58).

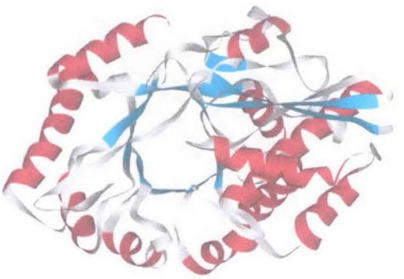


Figure 4-57 Structure of galactonate dehydratase The fold is the same as those of MR, MLE, and enolase (see Figure 4-40 for MR and MLE).

The case of galactonate dehydratase shows that sequence comparisons can identify overall protein structure class and locate the active-site residues, even in cases of very low sequence identity. They can also provide specific information about bound ligands—in this case, a divalent metal ion—and about the residues that bind them. Comparison of a protein of unknown function with just one other homologous sequence will usually not identify the other active-site residues or establish any commonality of function unless the overall sequence identity is very high. Comparisons of multiple sequences are much more informative, however, and will often be able to detect the functional groups that carry out the common core chemical step. When the active-site structures of at least some of the proteins being compared are known, the arrangement of these groups may also reveal the stereochemistry of the reaction being catalyzed. But what sequence and structure usually cannot do alone is to identify the substrate(s) of the reaction and the overall chemistry. Identification of chemistry is more reliable than identification of specificity. Proceeding from sequence to consequence in cases of very low sequence identity requires other sources of information, as illustrated by this case study.

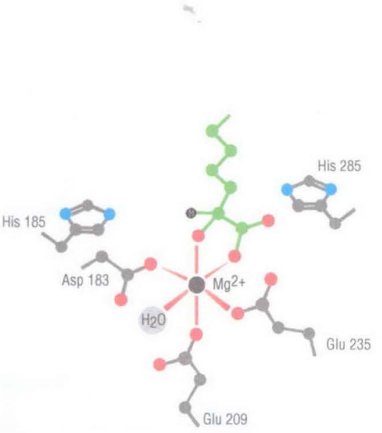


Figure 4-58 Schematic diagram of a model of the active site of galactonate dehydratase with substrate bound The metal-ion coordination and the disposition of the catalytic base histidine 285 and the electrophilic groups that interact with the substrate are similar to those found in MR, MLE and enolase, even though the overall reactions they catalyze are completely different.

References

Babbitt, P.C. and Gerlt, J.A.: **Understanding enzyme superfamilies. Chemistry as the fundamental determinant in the evolution of new catalytic activities.** *J. Biol. Chem.* 1997, **272**:30591–30594.

Babbitt, P.C. *et al.*: **The enolase superfamily: a general strategy for enzyme-catalyzed abstraction of the alpha-protons of carboxylic acids.** *Biochemistry* 1996, **35**:16489–16501.

Babbitt, P.C. *et al.*: **A functionally diverse enzyme superfamily that abstracts the alpha protons of carboxylic acids.** *Science* 1995, **267**:1159–1161.

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 181

4-17 Starting from Scratch: A Gene Product of Unknown Function

Function cannot always be determined from sequence, even with the aid of structural information and chemical intuition

About 30% of the 6,282 genes in the genome of the budding yeast *Saccharomyces cerevisiae* code for proteins whose function is completely unknown. One of these is gene *YBL036c*, whose sequence indicates that it encodes a protein of 257 amino acids. A comparison of this sequence against all genomic DNA sequences in the databases as of 1 June 2002 indicates about 200 other proteins whose amino-acid sequences show a greater-than-chance similarity to that of YBL036c. The putative homologs come from every kingdom of life—and none has a known function. At the time of writing, this is still true: although there are now clues to the function of the gene in yeast, the last chapter in this story is still to be written. What follows is an account of the avenues explored and where they lead.

As homologs of YBL036c are ubiquitous, the protein is more likely to function in some fundamental cellular process than to be involved in, for example, some aspect of cell–cell communication that would be confined to multicellular organisms. Thus yeast, a single-celled eukaryote whose complete genome sequence is known and whose metabolic processes can easily be studied by genetic methods, should be an ideal model organism in which to uncover the functions of this family of proteins.

YBL036c was selected as one of the first gene products to be studied in a project aimed at determining structures for yeast proteins which, because of the absence of clear sequence similarity to other proteins, seemed likely to have novel folds. In fact when the three-dimensional structure was determined by X-ray crystallography it proved to have a familiar fold: the triosephosphate isomerase alpha/beta barrel (Figure 4-59). This is yet another clear illustration of the fact that a protein fold can be encoded by very divergent sequences. Structural comparison between this three-dimensional fold and all other folds in the structural database shows the greatest similarity with the large domain of the bacterial enzyme alanine racemase. Like alanine racemase, the structure of YBL036c also revealed a covalently bound pyridoxal phosphate cofactor, which accounts for the yellow color of the purified protein.

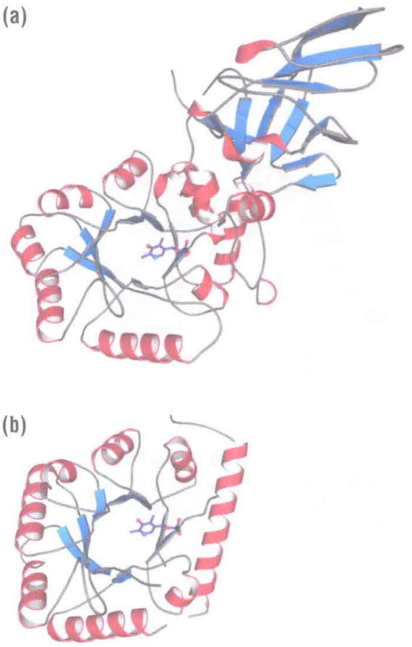


Figure 4-59 The three-dimensional structures of bacterial alanine racemase and yeast YBL036c The structure of the large domain of alanine racemase **(a)** is similar to the overall structure of YBL036c **(b)**. The yeast protein lacks the largely antiparallel beta-sheet domain of the racemase; however, the active sites, indicated by the presence of the bound pyridoxal phosphate cofactor (shown in ball-and-stick form), are located in the same place in both proteins. (PDB 1sft and 1ct5)

Comparison of the active sites of alanine racemase and YBL036c revealed both similarities and important differences. The essential lysine residue required of all pyridoxal-phosphate-dependent enzymes is present, covalently linked to the cofactor (Figure 4-60). A second interaction between the protein and the cofactor that is diagnostic for the chemical function of bacterial alanine racemase, an arginine interacting with the pyridine nitrogen of the cofactor, is also present in YBL036c. However, there were several significant differences. Alanine racemase is an obligatory dimer: residues from both subunits contribute to each other's active sites. In addition, alanine racemase has a second domain, which also contributes residues to the active site. YBL036c is a monomer and lacks the second domain entirely. Consequently, a number of residues found in the active site of alanine racemase are not present in the active site of YBL036c, raising the question of whether the biochemical function of alanine racemase has been preserved.

At this point, sequence and structure can tell us nothing further about function. Additional experimental approaches are needed (see Figure 4-1). The purified protein was first assayed to see if it had any alanine racemase activity, but it did not. More general methods of determining function must be tried. (Several of these are briefly described in section 4-4.)

Location of a protein within the cell is often informative about the cellular, if not biochemical, function. Sometimes, clues to location can be found in the sequence: for example, a carboxy-terminal KDEL sequence codes for retention of a protein in the endoplasmic reticulum. Other

References

Brent, R. and Finley, R.L. Jr.: **Understanding gene and allele function with two-hybrid methods.** *Annu. Rev. Genet.* 1997; **31**:663–704.

Eswaramoorthy, S. *et al.*: **Structure of a yeast hypothetical protein selected by a structural genomics approach.** *Acta Crystallogr. D. Biol. Crystallogr.* 2003, **59**:127–135.

Shaw, J.P. *et al.*: **Determination of the structure of alanine racemase from *Bacillus stearothermophilus* at 1.9-Å resolution.** *Biochemistry* 1997, **36**:1329–1342.

Templin, M. F. *et al.*: **Protein microarray technology.** *Trends Biotechnol.* 2002, **20**:160–166.

Tucker, C.L. *et al.*: **Towards an understanding of complex protein networks.** *Trends Cell Biol.* 2001, **11**:102–106.

van Roessel, P. and Brand, A.H.: **Imaging into the future: visualizing gene expression and protein interactions with fluorescent proteins.** *Nat. Cell Biol.* 2002, **4**:E15–E20.

von Mering, C. *et al.*: **Comparative assessment of large-scale data sets of protein–protein interactions.** *Nature* 2002, **417**:399–403.

Web site:
http://genome-www.stanford.edu/Saccharomyces/

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 182

sequence motifs specify transport into the nucleus, secretion from the cell, and so forth. YBL036c has no such motifs in its sequence, so more direct methods of determining localization must be used. The most common of these is the fusion of the protein of interest with a protein that can be visualized in the cell by antibody staining or intrinsic fluorescence. Fusion to green fluorescent protein (GFP), originally isolated from jellyfish, is a widely used strategy. Efforts are underway to apply these methods systematically to all the gene products in the yeast genome. By the GFP method, YBL036c was found distributed throughout the cell.

In eukaryotic cells in particular, the function of every protein is likely to depend in some manner on interaction with one or more other proteins. Demonstrating a physical interaction between two proteins can thus provide a clue to cellular or biochemical function if the function of one of them is known. Several different approaches to discovering such interactions have been developed. These include co-immunoprecipitation or affinity chromotography from cell extracts followed by mass spectroscopy to identify the interacting partner(s), and cell-based methods such as the yeast two-hybrid screen (see section 4-4). Application of the two-hybrid method to YBL036c detected one interacting partner, another protein of unknown function.

Proteins function in regulatory networks inside the cell; their expression patterns change with changes in external and internal conditions and proteins that perform similar functions often display similar patterns of expression. Thus, clues to the function of a gene product can come from analysis of the pattern of its expression under different conditions and comparison with the patterns of other proteins of known function. Two widely used methods for studying expression are two-dimensional gel electrophoresis, which measures protein levels directly, and DNA microarrays, which measure levels of mRNA. Both can be applied to whole genomes or subsets of the genome. Microarray analyses of yeast gene expression have been carried out by many different laboratories under hundreds of different conditions; most contain information about YBL036c. In general, YBL036c is expressed in all stages of the cell cycle and in all growth conditions tested. Its expression is broadly the same as that of a number of genes that code for proteins involved in amino-acid metabolism. Its expression is upregulated slightly in a variety of stress conditions.

If a gene is not essential for the survival of an organism, deleting it from the genome can often give rise to a phenotype suggestive of function. Microarray analysis of such a deletion strain should show changes in the expression of genes whose function is in some manner coupled with that of the gene that has been deleted. YBL036c is not an essential gene in yeast: the deletion strain is viable and shows no growth defect under a variety of conditions. However, there is a subtle phenotype when the yeast cells form spores. Instead of being dispersed, the spores clump together. Electron microscopy of the spore shows that the ascus containing the spores and the spores themselves have an abnormal wall structure. Since sporulation requires remodeling of cell-wall structures, this phenotype implies that YBL036c is involved in this process. Microarray analysis of the deletion strain supports this conclusion: genes involved in cell-wall biosynthesis show changes in expression levels when YBL036c is absent.

Although these genome-wide methods have suggested a cellular process in which YBL036c participates, a great deal of information is still needed to fully describe the workings of this gene product within the cell. The active-site architecture suggests that the protein may be an amino-acid racemase, but the substrate has yet to be identified. If YBL036c produces a D-amino acid, as such activity would indicate, the role of this product in cell-wall structure remains to be determined. As animal cells do not have cell walls, the cellular function of YBL036c homologs in those organisms might be somewhat different. It is clear from this example that the task of determining the function(s) of a gene is one that does not end with a single organism.

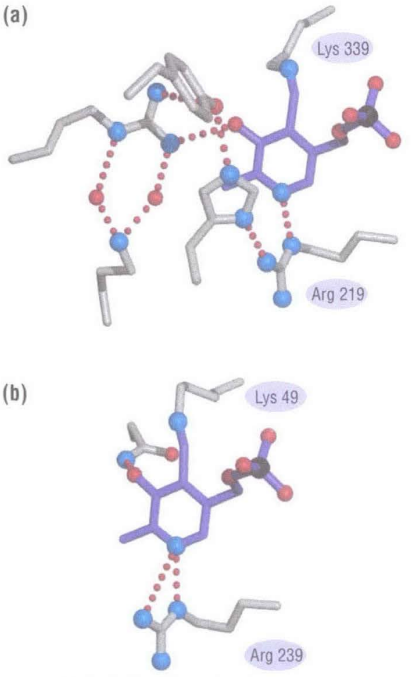


Figure 4-60 Comparison of the active sites of bacterial alanine racemase and YBL036c
(a) Alanine racemase; **(b)** YBL036c. Although many of the interactions between the pyridoxal phosphate cofactor (shown in purple) and protein side chains are different in the two active sites, two interactions are preserved: a covalent linkage to a lysine residue, and the interaction of a nitrogen atom in the pyridine ring of the cofactor with an arginine residue. The lysine interaction is diagnostic of all pyridoxal-phosphate-dependent enzymes; the interaction with arginine is diagnostic for pyridoxal-phosphate-dependent amino-acid racemases.

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 183

5

Structure Determination

Most of the protein structures described and discussed in this book have been determined either by X-ray crystallography or by nuclear magnetic resonance (NMR) spectroscopy. Although these techniques both depend on data derived from physical techniques for probing structure, their interpretation is not unambiguous and entails assumptions and approximations often depending upon knowledge of the protein from other sources, including its biology. This chapter briefly describes how structures are determined by X-ray crystallography and NMR, how the data are interpreted, and what contributes to the accuracy of a structure determination.

5-1 **The Interpretation of Structural Information**
5-2 **Structure Determination by X-Ray Crystallography and NMR**
5-3 **Quality and Representation of Crystal and NMR Structures**

Petitioner Merck, Ex. 1209, p. 184

5-1 The Interpretation of Structural Information

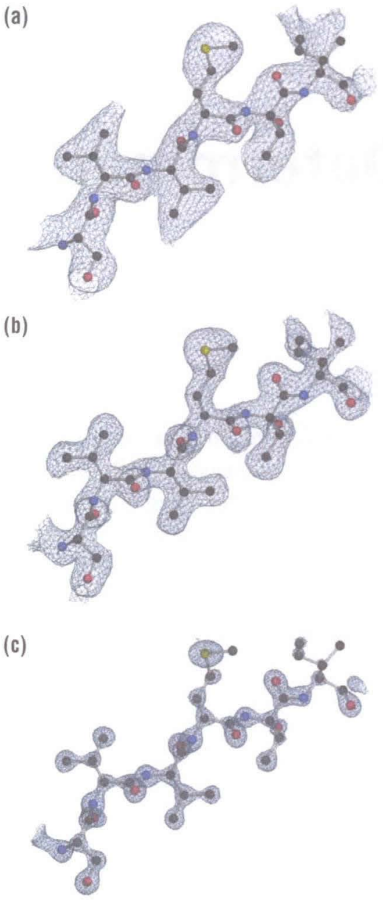


Figure 5-1 Portion of a protein electron density map at three different resolutions The peptide corresponding to the electron density map (serine, valine, valine, methionine, threonine, isoleucine) is superimposed. **(a)** At 3-Å resolution the fold of the polypeptide chain can be seen and the approximate positions of side chains can be determined. Interatomic distances can only be measured to ± about 0.5 Å. **(b)** At 2-Å resolution side chains are well delineated and the peptide carbonyls of the backbone are discernible, allowing the chain to be oriented with precision. Interatomic distances can be measured to a precision of about 0.2 Å. Approximately three times as many data are required for this resolution as were used at 3-Å resolution. **(c)** At 1-Å resolution atoms are visible and resolved. Interatomic distances can be measured to a precision of a few hundredths of an Ångstrom. Almost 30 times more data are required for this resolution as were used at 3-Å resolution. In favorable cases, the position of hydrogen atoms can be inferred at this resolution. Note that at high resolution (c) there is no electron density for the methyl group of the methionine side chain because it is disordered. At lower resolutions (a and b), it appears that such density is present, but this is actually the density of the heavier sulfur atom. Since at low resolution electron density is more diffuse, the entire thio-ether portion of the side chain can be fitted into the sulfur density. Kindly provided by Aaron Moulin.

Experimentally determined protein structures are the result of the interpretation of different types of data

Much of the content of this book depends on a detailed understanding of the structure of proteins at the atomic level. The atomic structures of biological macromolecules can be determined by several techniques: although a few have been determined by electron microscopy, the vast majority have been obtained by either single crystal *X-ray diffraction*, generally known as *X-ray crystallography*, or by *nuclear magnetic resonance spectroscopy*, generally known as NMR. How this is done, in both cases, is briefly explained in the next section. Here, we describe the kind of information these techniques produce, and how it can be interpreted.

Both X-ray crystallography and NMR produce information on the relative positions of the atoms of the molecule: these are termed **atomic coordinates**. Because X-rays are scattered from the electron cloud of the atoms, whereas NMR measures the interactions of atomic nuclei, X-ray crystallography provides the positions of all non-hydrogen atoms whereas NMR provides the positions of all atoms including hydrogen. However, the coordinate sets provided in this way represent the interpretation of the primary data obtained in each case. The precision and accuracy with which coordinates can be derived from these data depend on several factors that are different for the two methods. The objective end-product of a crystallographic structure determination is an **electron density map** (Figure 5-1), which is essentially a contour plot indicating those regions in the crystal where the electrons in the molecule are to be found. Human beings must interpret this electron density map in terms of an atomic model, aided by semi-automatic computational procedures. The objective end-product of an NMR structure determination is usually a set of distances between atomic nuclei that define both bonded and non-bonded close contacts in the molecule. These must be interpreted in terms of a molecular structure, a process that is aided by automated methods.

Because interpretation in each case requires assumptions and approximations, macromolecular structures can have errors. In most cases, these errors are small, and only affect a small part of the structure. However, such errors can be quite large: cases of completely incorrect structures, though rare, have been reported.

The choice of technique depends on several factors, including the molecular weight, solubility and ease of crystallization of the protein in question. For proteins and protein complexes with molecular weights above 50–100 kDa, X-ray crystallography is the method of choice. For smaller proteins or protein complexes, either method may be usable but X-ray diffraction will usually provide more precise structural information than NMR provided the species crystallizes easily and is well ordered in the crystal. For molecules that are hard to crystallize and can be dissolved at reasonably high concentrations without aggregation, NMR is the method of choice. In many cases however, both techniques can be used and provide complementary information, since crystallographic images are static but NMR can be used to study the flexibility of proteins and their dynamics over a wide range of time scales.

Both the accuracy and the precision of a structure can vary

It is essential to appreciate the distinction between the accuracy of an experimentally determined structure and its precision. The latter is much easier to assess than the former. We define the precision of a structure determination in terms of the reproducibility of its atomic coordinates. If the data allow the equilibrium position of each atom to be determined precisely, then the

Definitions

atomic coordinate: the position in three-dimensional space of an atom in a molecule relative to all other atoms in the molecule.

electron density map: a contour plot showing the distribution of electrons around the atoms of a molecule.

resolution: the level of detail that can be derived from a given process.

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 185

same structure determined independently elsewhere should yield atomic coordinates that agree very closely with the first set. Precise coordinates are usually reproducible to within a few tenths of an Ångstrom.

Accuracy refers to whether or not the structure is correct, but there are two ways to define correctness. A structure may be right but irrelevant: the protein may be in a conformation that is determined by the experimental conditions rather than its biological context. Such cases are rare, but they have occurred. It is also possible to interpret the experimental data incorrectly, yielding a structure that is not and could not possibly be correct, either in whole (rarely) or in part. Often it is possible to determine that a structure has been built incorrectly on the basis of the known rules of how proteins fold and their properties when they are folded. It is not possible to determine the correctness of minor details unless there is a biological experiment that can probe the structure at that position. For example, the involvement of a residue in catalysis predicted by the structure can be tested by mutating it to one that cannot perform the required catalytic function. Ultimately, the accuracy of a structure can best be assessed by the answer to a simple question: is the structure consistent with the body of biochemical and biological information about the protein?

Sometimes, part of a structure may be invisible to the experimental method used. This can arise because that part of the molecule is disordered and therefore does not contribute to the reflected X-rays in a crystallographic experiment or, in an NMR experiment, atoms may not be close enough to interact strongly.

The information content of a structure is determined by its resolution

Precision and accuracy of a structure are related: the more precise a structure determination is, the less likely it is to have gross errors. What links precision with accuracy is the concept of **resolution**.

Crystallographers express the resolution of a structure in terms of a distance: if a structure has been determined at 2-Å resolution then any atoms separated by more than about this distance will appear as separate maxima in the electron density contour plot, as we explain in the next section, and their positions can be obtained directly to high precision. Atoms closer together than the resolution limit will appear as a fused electron density feature, and their exact positions must be inferred from the shape of the electron density and knowledge of the chemical structure of amino acids or nucleotides (Figure 5-1). Since the average C–C single bond distance is 1.5 Å, the precision with which the individual atoms of, say, an inhibitor bound to an enzyme can be located will be much greater at 1.5-Å resolution than in a structure determined at, say, 3-Å resolution.

NMR spectroscopists express resolution somewhat differently. NMR structures are determined as ensembles of similar models (Figure 5-2): we explain in the next section how these are derived. However, since more than one closely related model will fit the data, the effective resolution of an NMR structure is given by the extent of the differences between these models, expressed as a root-mean-square deviation (RMSD) of their atomic coordinates. The smaller the deviation, the more precise (and, presumably accurate) the NMR structure is believed to be, and therefore the higher its effective resolution.

In the case of both X-ray and NMR structures, what sets the resolution limit is the intrinsic order of the protein plus the amount of data that the experimenter is able to measure. The greater the amount of data, the higher the resolution, and the higher the resolution, the more precise and accurate the information that can be extracted from the structure.

In favorable cases, the resolution of a macromolecular structure determination by crystallography is such that the relative positions of all non-hydrogen atoms are known to a precision of a few tenths of an Ångstrom.

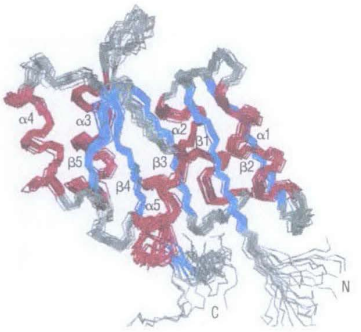


Figure 5-2 NMR structure ensemble The figure shows the superposition of the set of models derived from the internuclear distances measured for this protein in solution. Note that different portions of the structure are determined with different precision. Blue represents beta strands, red represents alpha helices, grey represents loops. This figure should not be taken to indicate the flexibility of segments of the protein: different regions may be poorly defined because there are insufficient data to constrain the structure, and not because the structures are mobile.

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 186

5-2 Structure Determination by X-Ray Crystallography and NMR

Protein crystallography involves summing the scattered X-ray waves from a macromolecular crystal

The steps in solving a protein crystal structure at high resolution are diagrammed in Figure 5-3. First, the protein must be crystallized. This is often the rate-limiting step in straightforward structure determinations, especially for membrane proteins. Then, the X-ray diffraction pattern from the crystal must be recorded. When X-rays strike a macromolecular crystal, the atoms in the molecules produce scattered X-ray waves which combine to give a complex diffraction pattern consisting of waves of different amplitudes. What is measured experimentally are the amplitudes and positions of the scattered X-ray waves from the crystal. The structure can be reconstructed by summing these waves, but each one must be in the correct registration with respect to every other wave, that is, the origin of each wave must be determined so that they sum to give some image instead of a sea of noise. This is called the **phase problem**. Phase values must be assigned to all of the recorded data; this can sometimes be done computationally, but is usually done experimentally by labeling the protein with one or more heavy atoms whose position in the crystal can be determined independently. The phased waves are then summed in three dimensions to generate an image of the electron density distribution of the molecule in the crystal. This can be done semi-automatically or by hand on a computer graphics system. A chemical model of part of the molecule is docked into the shape of each part of the electron part of the electron density (as shown in Figure 5-3). This fitting provides the first picture of the structure of the protein. The overall model is improved by an iterative process called refinement whereby the positions of the atoms in the model are tweaked until the calculated diffraction pattern from the model agrees as well as can be with the experimentally measured diffraction pattern from the actual protein. There is no practical limit to the size of the protein or protein complex whose structure can be determined by X-ray crystallography.

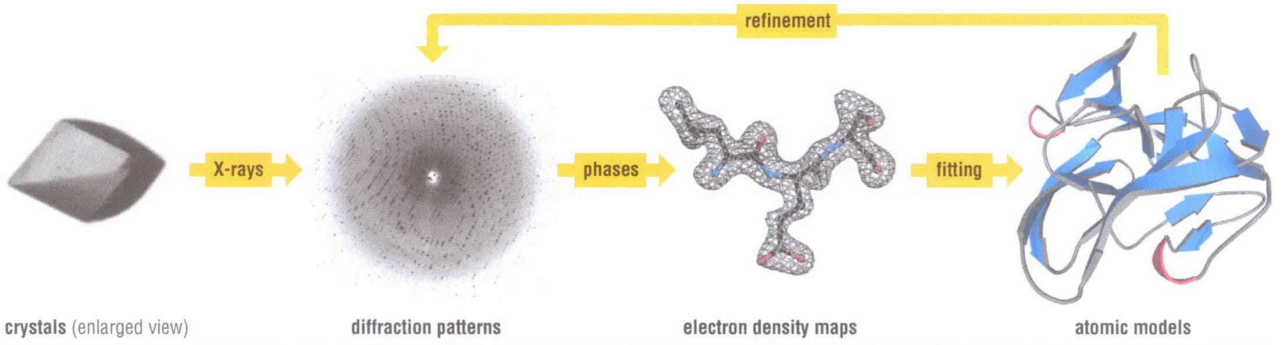


Figure 5-3 Structure determination by X-ray crystallography The first step in structure determination by X-ray crystallography is the crystallization of the protein. The source of the X-rays is often a synchrotron and in this case the typical size for a crystal for data collection may be $0.3 \times 0.3 \times 0.1$ mm. The crystals are bombarded with X-rays which are scattered from the planes of the crystal lattice and are captured as a diffraction pattern on a detector such as film or an electronic device. From this pattern, and with the use of reference—or phase—information from labeled atoms in the crystal, electron density maps (shown here with the corresponding peptide superimposed) are computed for different parts of the crystal. A model of the protein is constructed from the electron density maps and the diffraction pattern for the modeled protein is calculated and compared with the actual diffraction pattern. The model is then adjusted—or refined—to reduce the difference between its calculated diffraction pattern and the pattern obtained from the crystal, until the correspondence between model and reality is as good as possible. The quality of the structure determination is measured as the percentage difference between the calculated and the actual pattern.

Definitions

phase problem: in the measurement of data from an X-ray crystallographic experiment only the amplitude of the wave is determined. To compute a structure, the phase must also be known. Since it cannot be determined directly, it must be determined indirectly or by some other experiment.

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 187

NMR spectroscopy involves determining internuclear distances by measuring perturbations between assigned resonances from atoms in the protein in solution

Unlike crystallography, structure determination by NMR (Figure 5-4) is carried out on proteins in solution, but the protein must be soluble without aggregating at concentrations close to those of a protein in a crystal lattice. NMR structure determination requires two types of data. The first is measurement of nuclear magnetic resonances from protons and isotopically labeled carbons and nitrogens in the molecule. Different nuclei in a protein absorb electromagnetic energy (resonance) at different frequencies because their local electromagnetic environments differ due to the three-dimensional structure of the protein. These resonances must be assigned to atoms in specific amino acids in the protein sequence, a process that requires several specific types of experiments. The second set of data consists of internuclear distances that are inferred from perturbing the resonance of different atoms and observing which resonances respond; only atoms within 5 Å of each other show this effect, and its magnitude varies with the distance between them. The set of approximate internuclear distances is then used to compute a structure model consistent with the data. Since the distances are imprecise, many closely related models may be consistent with the observations, so NMR structures are usually reported as ensembles of atomic coordinates. In practice, if a structure is determined by both NMR and X-ray diffraction, it is usually found that the average of the NMR ensemble closely resembles the crystal structure. As a general rule, there are practical limitations to the determination of the structure of a complete protein by NMR: a molecular weight of about 50 kDa is considered a very large protein for NMR. In special cases, domains or portions of much larger proteins or complexes can be studied.

Unlike X-ray diffraction, which presents a static picture (an average in time and space) of the structure of a protein, NMR has the capability of measuring certain dynamic properties of proteins over a wide range of time scales.

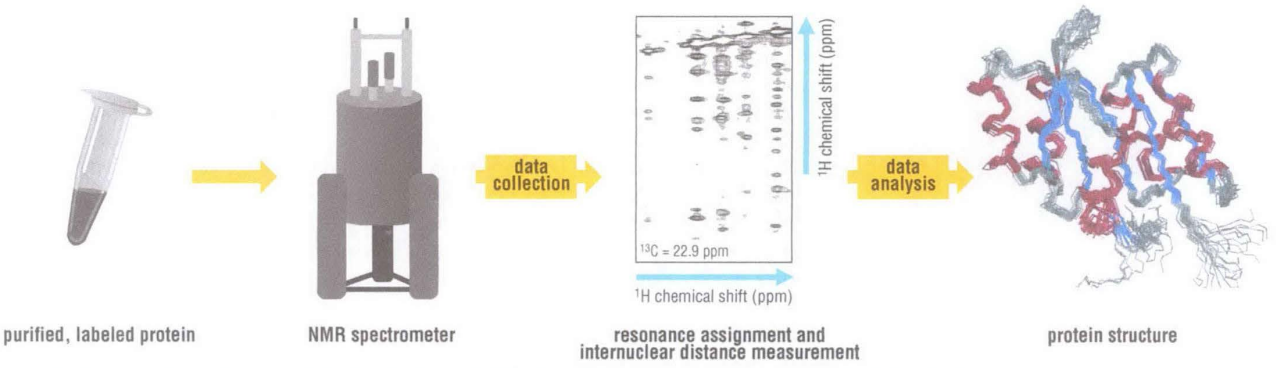


Figure 5-4 Structure determination by NMR For protein structure determination by NMR, a labeled protein is dissolved at very high concentration and placed in a magnetic field, which causes the spin of the hydrogen atoms to align along the field. Radio frequency pulses are then applied to the sample, perturbing the nuclei of the atoms which when they relax back to their original state emit radio frequency radiation whose properties are determined by the environment of the atom in the protein. This emitted radiation is recorded in the NMR spectrometer for pulses of differing types and durations (for simplicity, only one such record is shown here), and compared with a reference signal to give a measure known as the chemical shift. The relative positions of the atoms in the molecule are calculated from these data to give a series of models of the protein which can account for these data. The quality of the structure determination is measured as the difference between the different models.

References

Drenth, J.: *Principles of Protein X-Ray Crystallography* 2nd ed. (Springer-Verlag, New York, 1999).

Evans, J.N.S.: *Biomolecular NMR Spectroscopy* (Oxford University Press, Oxford, 1995).

Markley, J.L. *et al.*: **Macromolecular structure determination by NMR spectroscopy.** *Methods Biochem. Anal.* 2003, **44**:89–113.

Rhodes, G.: *Crystallography Made Crystal Clear: A Guide to Users of Macromolecular Models* 2nd ed. (Academic Press, New York and London, 1999).

Schmidt, A. and Lamzin, V.S.: ***Veni, vidi, vici*—atomic resolution unravelling the mysteries of protein function.** *Curr. Opin. Struct. Biol.* 2002,**12**:698–703.

Gorenstein, N.: *Nuclear Magnetic Resonance (NMR)* (Biophysics Textbooks Online): http://www.biophysics.org/btol/NMR.html

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 188

5-3 Quality and Representation of Crystal and NMR Structures

The quality of a finished structure depends largely on the amount of data collected

Both X-ray and NMR structure determination have statistical criteria for the quality of the atomic model produced. Crystallographers usually speak of R-factors, which represent the percentage disagreement between the observed diffraction pattern and that calculated from the final model. R-factors of around 20% or less are considered indicative of well determined structures that are expected to contain relatively few errors. NMR spectroscopists usually report overall root mean square deviation (RMSD) between the atoms in secondary structure elements in all coordinate sets in the ensemble of structures consistent with the experimental data. In practice, RMSDs of 0.7 Å are considered good, indicating a structure determination of high precision. RMSDs of around 1 Å are considered acceptable.

There is no substitute for high resolution. It makes the structure determination easier and more reliable. The closer one gets to true atomic resolution in X-ray crystallography (better than 1.5 Å), the less ambiguity one has in positioning every atom. Atomic resolution allows one to detect mistakes in the biochemically determined or genomically derived amino-acid sequence, to correct preliminary incorrect chain connectivity, and to identify unexpected chemical features in the molecule. Incorrect crystal structures are almost never reported from high-resolution data. Most of the mistakes in protein crystallography have been made because a medium-resolution structure has been misinterpreted or overinterpreted. In NMR, the general rule is the more internuclear distances measured the better. Most of the mistakes that have been made in NMR structure determination have resulted from either incorrect assignment of a set of resonances to a particular part of a protein or the failure to measure enough internuclear distances.

Different conventions for representing the structures of proteins are useful for different purposes

Atomic coordinate sets make for boring reading, so protein structures are presented visually. There are a number of different ways to render a protein structure, depending on the information that one wishes to convey. The fold of the polypeptide chain can be depicted as a wire model that follows the path of the backbone (Figure 5-5a), which is useful for example in comparisons of two conformations of the same molecule (see Figure 1-80); or it can be depicted as a ribbon diagram in which alpha helices and beta sheets are graphically stylized (Figure 5-5b): this not only makes the overall fold easily recognizable, but makes particular secondary structure elements or loops that may have particular functional significance easily recognizable. Detail of the structure at the atomic level can be rendered by means of a ball-and-stick model (Figure 5-5c) in which the balls are colored or sized by type of atom and covalent bonds are represented by perspective sticks; such drawings are to scale so relative bonded and non-bonded distances can be assessed, which is important for evaluating interactions. Atoms as volumes can be represented by space-filling drawings in which each atom is given a sphere scaled to its van der Waals radius (Figure 5-5d): this representation is particularly useful for assessing the fit of a ligand to a binding site (see Figure 2-8b). Finally, to emphasize the protein surface that is created by the space-filling nature of atoms, a surface topography image can be produced (Figure 5-5e). Such an image can be colored according to different local properties such as the electrostatic potential at different points in the molecule. In this book, all of these different methods of visualizing structures are used.

References

Holyoak, T. *et al.*: **The 2.4 Å crystal structure of the processing protease Kex2 in complex with an Ala–Lys–Arg boronic acid inhibitor.** *Biochemistry*, in the press.

Martz, E.: **Protein Explorer: easy yet powerful macromolecular visualization.** *Trends Biochem. Sci.* 2002, **27**:107–109.

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 189

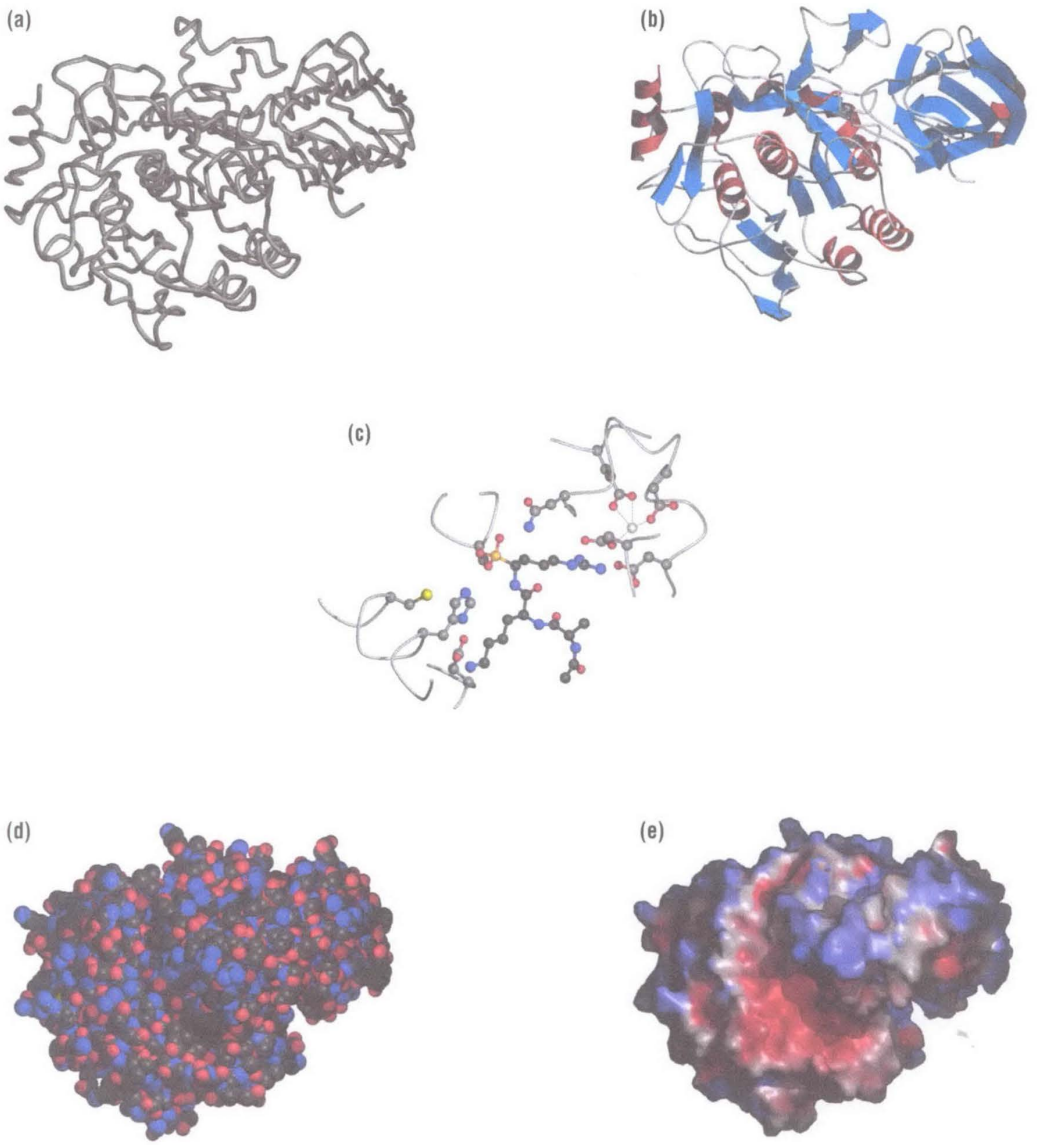


Figure 5-5 Different ways of presenting a protein structure The structure of the catalytic and P domains of the protease Kex2 is shown in different styles. **(a)** "Wire" diagram showing the path of the polypeptide backbone. **(b)** Ribbon diagram highlighting the secondary structure elements. Beta strands are depicted as arrows with the arrow head being at the carboxyl terminus. Alpha helices are drawn as coiled ribbons. **(c)** Ball-and-stick model of a small part of the protein structure, showing details of amino-acid interactions. **(d)** Space-filling representation in which every non-hydrogen atom is shown as a sphere of its van der Waals radius. **(e)** Surface representation (sometimes called a GRASP image after the program that computes it) in which the topography of the protein surface is shown and the electrostatic characteristics of the surface are highlighted in color (red for negative, blue for positive). Kindly provided by Todd Holyoak.

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 190

Glossary

acid: a molecule or chemical group that donates a proton, either to water or to some other base. (2-12)

acid-base catalysis: catalysis in which a proton is transferred in going to or from the transition state. When the acid or base that abstracts or donates the proton is derived directly from water (H^+ or OH^-) this is called specific acid-base catalysis. When the acid or base is not H^+ or OH^-, it is called general acid-base catalysis. Nearly all enzymatic acid-base catalysis is general acid-base catalysis. (2-12)

activation energy: the energy required to bring a species in a chemical reaction from the **ground state** to a state of higher free energy, in which it can transform spontaneously to another low-energy species. (2-6)

activation-energy barrier: the higher-energy region between two consecutive chemical species in a reaction. (2-6)

activation loop: a stretch of polypeptide chain that changes conformation when a kinase is activated by phosphorylation and/or protein binding. This segment may or may not be the one containing the residue that is phosphorylated to activate the kinase. Usually, in the inactive state, the activation loop blocks access to the active site. (3-13)

activation segment: see **activation loop**.

active site: asymmetric pocket on or near the surface of a macromolecule that promotes chemical catalysis when the appropriate **ligand** (substrate) binds. (2-1)

affinity: the tightness of a protein–ligand complex. (2-4)

alignment: procedure of comparing two or more sequences by looking for a series of characteristics (residue identity, similarity, and so on) that match up in both and maximize conservation, in order to assess overall similarity. (4-1)

allosteric activator: a ligand that binds to a protein and induces a conformational change that increases the protein's activity. (3-5)

allosteric inhibitor: a ligand that binds to a protein and induces a conformational change that decreases the protein's activity. (3-5)

allostery: the property of being able to exist in two structural states of differing activity. The equilibrium between these states is modulated by ligand binding. (3-5)

alpha/beta barrel: a parallel beta barrel formed usually of eight strands, each connected to the next by an alpha-helical segment. Also known as a **TIM barrel**. (1-18)

alpha/beta domain: a protein domain composed of beta strands connected by alpha helices. (1-17)

alpha+beta domain: a protein domain containing separate alpha-helical and beta-sheet regions. (1-17)

alpha/beta twist: a twisted parallel beta sheet with a saddle shape. Helices are found on one side of the sheet for the first half and the other side for the second half. (1-18)

alpha domain: a protein domain composed entirely of alpha helices. (1-17)

alpha helix: a coiled conformation, resembling a right-handed spiral staircase, for a stretch of consecutive amino acids in which the backbone –N–H group of every residue n donates a hydrogen bond to the C=O group of every residue n+4. (1-5)

alternative splicing: the production of different versions of the final protein sequence from a gene sequence by the removal during RNA processing of portions of the RNA containing or affecting coding sequences. (1-2)

amide bond: a chemical bond formed when a carboxylic acid condenses with an amino group with the expulsion of a water molecule. (1-3)

amphipathic: having both polar and nonpolar character and therefore a tendency to form interfaces between **hydrophobic** and **hydrophilic** molecules. (1-1)

amphipathic alpha helix: an alpha helix with a hydrophilic side and a hydrophobic side. (1-6)

anisotropic: behaving differently in different directions; dependent on geometry and direction. (2-4)

antiparallel beta sheet: a beta sheet, often formed from contiguous regions of the polypeptide chain, in which each strand runs in the opposite direction from its immediate neighbors. (1-7)

atomic coordinate: the position in three-dimensional space of an atom in a molecule relative to all other atoms in the molecule. (5-1)

autophosphorylation: phosphorylation of a protein kinase by itself. Autophosphorylation may occur when the active site of the protein molecule to be phosphorylated catalyzes this reaction (*cis* autophosphorylation) or when another molecule of the same kinase provides the active site that carries out the chemistry (*trans* autophosphorylation). Autophosphorylation *in trans* often occurs when kinase molecules dimerize, a process that can be driven by ligand binding as in the receptor tyrosine kinases. (3-13)

backbone: the regularly repeating part of a polymer. In proteins it consists of the amide –N–H, alpha carbon –C–H and the carbonyl –C=O groups of each amino acid residue in the **polypeptide** chain. Residues are linked to each other by means of **peptide bonds**. (1-0, 1-3)

base: (in a nucleic acid) the aromatic group attached to the sugar of a **nucleotide**. (1-2)

base: (in chemistry) a molecule or chemical group that accepts a proton, either from water or from some other acid. (2-12)

beta barrel: a beta sheet in which the last strand is hydrogen bonded to the first strand, forming a closed cylinder. (1-7)

beta domain: a protein domain containing only beta sheet. (1-17)

beta sandwich: a structure formed of two antiparallel beta sheets packed face to face. (1-17)

beta sheet: a **secondary structure** element formed by backbone hydrogen bonding between segments of extended polypeptide chain. (1-5)

beta turn: a tight turn that reverses the direction of the polypeptide chain, stabilized by one or more backbone hydrogen bonds. Changes in chain direction can also occur by loops, which are peptide chain segments with no regular conformations. (1-5)

bifunctional: having two distinct biochemical functions in one gene product. Bifunctional enzymes catalyze two distinct chemical reactions. (2-15)

BLAST: a family of programs for searching protein and DNA databases for sequence similarities by optimizing a specific similarity measure between the sequences being compared. (4-2)

catalyst: a substance that accelerates the rate of a reaction without itself being permanently altered. (2-6)

catalytic triad: a set of three amino acids that are hydrogen bonded together and cooperate in catalysis. (2-14)

cavity: a completely enclosed hole in the interior of a protein. Cavities may contain one or more disordered water molecules but some are believed to be completely empty. (2-3)

chameleon sequence: a sequence that exists in different conformations in different environments. (4-14)

chaperone: a protein that aids in the folding of another protein by preventing the unwanted association of the unfolded or partially folded forms of that protein with itself or with others. (1-9)

chromatin: the complex of DNA and protein that comprises eukaryotic nuclear chromosomes. The DNA is wound around the outside of highly conserved histone proteins, and decorated with other DNA-binding proteins. (3-20)

co-activator: a regulatory molecule that binds to a gene activator protein and assists its binding to DNA. (3-5)

codon: three consecutive **nucleotides** in a strand of DNA or RNA that represent either a particular amino acid or a signal to stop translating the transcript of the gene. The formula for translating the codons is given by the genetic code. (1-2)

coenzyme: a cofactor that is an organic or organometallic molecule and that assists catalysis. (2-13)

cofactor: a small, non-protein molecule or ion that is bound in the functional site of a protein and assists in ligand binding or catalysis or both. Some cofactors are bound covalently, others are not. (1-13, 2-13)

coiled coil: a protein or a region of a protein formed by a dimerization interaction between two alpha helices in which hydrophobic side chains on one face of each helix interdigitate with those on the other. (1-19)

competitive inhibitor: a species that competes with substrate for binding to the active site of an enzyme and thus inhibits catalytic activity. (3-0)

conservative substitution: replacement of one amino acid by another that has similar chemical and/or physical properties. (1-2)

conserved: identical in all sequences or structures compared. (4-1)

convergent evolution: evolution of structures not related by ancestry to a common function that is reflected in a common structure. (1-16, 4-5)

cooperative binding: interaction between two sites on a protein such that the binding of a ligand to the first one affects the properties—usually binding or catalytic—of the second one. (3-4)

cooperativity: interaction between two sites on a protein such that something that happens to the first one affects the properties of the second one. (3-4)

coordinate covalent bond: a bond formed when a lone pair of electrons from an atom in a ligand is donated to a vacant orbital on a metal ion. (1-13)

co-repressor: a regulatory molecule that binds to a gene repressor protein and assists its binding to DNA. (3-5)

cross-linked domain: a small protein domain with little or no secondary structure and stabilized by disulphide bridges or metal ions. (1-17)

decarboxylation: removal of carbon dioxide from a molecule. (2-11)

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 191

degenerate: having more than one **codon** for an amino acid. (1-2)

denaturant: a chemical capable of unfolding a protein in solution at ordinary temperatures. (1-12)

denatured state: the partially or completely unfolded form of a biological macromolecule in which it is incapable of carrying out its biochemical and biological functions. (1-12)

dimer: an assembly of two identical (homo-) or different (hetero-) subunits. In a protein, the subunits are individual folded polypeptide chains. (1-19)

dipole moment: an imaginary vector between two separated charges that may be full or partial. Molecules or functional groups having a dipole moment are said to be polar. (1-3)

disulfide bridge: a covalent bond formed when the reduced –S–H groups of two cysteine residues react with one another to make an oxidized –S–S– linkage. (1-4)

divergent evolution: evolution from a common ancestor. (4-5)

DNA microarray: an ordered array of nucleic acid molecules, either cDNA fragments or synthetic oligonucleotides, where each position in the array represents a single gene. (4-4)

domain: a compact unit of protein structure that is usually capable of folding stably as an independent entity in solution. Domains do not need to comprise a contiguous segment of peptide chain, although this is often the case. (1-14)

domain fold: the particular topographical arrangement of secondary structural elements that characterizes a single domain. Examples are an antiparallel arrangement of four helices in a four-helix bundle, or an open twisted beta sandwich with a particular sequence that binds nucleotides. (1-15)

domain swapping: the replacement of a structural element of one subunit of an oligomer by the same structural element of the other subunit, and vice versa. The structural element may be a secondary structure element or a whole domain. (2-4)

dominant-negative: dominant loss of function due to a single mutant copy of a gene. This can occur when the mutant subunit is able to oligomerize with normal subunits to form a non-functional protein, thereby producing a loss-of-function phenotype even in the presence of a normal copy of the gene. (1-20)

effector: a species that binds to a protein and modifies its activity. Effectors may be as small as a proton or as large as a membrane and may act by covalent binding, noncovalent binding, or covalent modification. (3-0)

effector ligand: a ligand that induces a change in the properties of a protein. (3-4)

electron density map: a contour plot showing the distribution of electrons around the atoms of a molecule. (5-1)

electrostatic interactions: noncovalent interaction between atoms or groups of atoms due to attraction of opposite charges. (1-4)

enthalpy: a form of energy, equivalent to work, that can be released or absorbed as heat at constant pressure. (1-12)

entropy: a measure of the disorder or randomness in a molecule or system. (1-12)

equilibrium: the state at which the rate of the forward reaction and the rate of the reverse reaction in a chemical transformation are equal. At equilibrium, the relative concentrations of reactants and products no longer change, although the reaction continues to occur. (2-6)

equilibrium constant: the ratio of the product of the concentrations of reaction products to the product of the concentrations of reaction reactants. For a reaction of the general form A + B = C + D, the equilibrium constant K_{eq} is [C][D]/[A][B], where [X] is the concentration of X, usually in moles per liter. This definition is a simplification that neglects effects at high concentrations. (2-6)

E-value: the probability that an **alignment** score as good as the one found between two sequences would be found in a comparison between two random sequences; that is, the probability that such a match would occur by chance. (4-1)

evolutionary distance: the number of observed changes in nucleotides or amino acids between two related sequences. (4-1)

exon: coding segment of a gene. The coding DNA of many eukaryotic genes is interrupted by segments of non-coding DNA (**introns**). (1-2)

extein: the sequences flanking an intein and which are religated after **intein** excision to form the functional protein. (3-17)

family: a group of homologous proteins that share a related function. Usually these will also have closely related sequences. Members of the same enzyme family catalyze the same chemical reaction on structurally similar substrates. (4-8)

four-helix bundle: a structure of four antiparallel alpha helices. Parallel bundles are possible but rare. (1-17)

free energy: a function, designed to produce a criterion for spontaneous change, that combines the **entropy** and **enthalpy** of a molecule or system. Free energy decreases for a spontaneous process, and is unchanged at equilibrium. (1-12)

functional motif: sequence or structural **motif** that is always associated with a particular biochemical function. (1-16, 4-2)

gated binding: binding that is controlled by the opening and closing of a physical obstacle to substrate or inhibitor access in the protein. (2-7)

gene knockout: inactivation of the function of a specific gene in a cell or organism, usually by recombination with a marker sequence but sometimes by antisense DNA, RNA interference, or by antibody binding to the gene product. The phenotype resulting from the knockout can often provide clues to the function of the gene. (4-4)

genetic code: the relationship between each of the 64 possible three-letter combinations of A, U, G and C (which stand for the RNA bases adenine, uracil, guanine and cytosine, respectively) and the 20 naturally occurring amino acids that make up proteins. U is the RNA equivalent of T (thymine) in DNA. (1-2)

genomics: the study of the DNA sequence and gene content of whole genomes. (4-0)

globin fold: a predominantly alpha-helical arrangement observed in certain heme-containing proteins. (1-17)

glycosylation: the post-translational covalent addition of sugar molecules to asparagine, serine or threonine residues on a protein molecule. Glycosylation can add a single sugar or a chain of sugars at any given site and is usually enzymatically catalyzed. (1-13, 3-18)

glycosylphosphatidylinositol anchor: a complex structure involving both lipids and carbohydrate molecules that is reversibly attached to some proteins to target them to the cell membrane. (3-19)

G protein: a member of a large class of proteins with GTPase activity that act as molecular switches in many different cellular pathways, controlling processes such as sensory perception, intracellular transport, protein synthesis and cell growth and differentiation. They undergo a large conformational change when a bound GTP is hydrolyzed to GDP. (3-6)

Greek-key motif: an arrangement of antiparallel strands in which the first three strands are adjacent but the fourth strand is adjacent to the first, with a long connecting loop. (1-17)

ground state: a species with low free energy; usually, the non-activated state of any substance. (2-6)

ground-state destabilization: raising the free energy, (relative to some reference state), of the ground state, usually referring to the bound substrate in the active site before any chemical change has occurred. Geometric or electronic strain are two ways of destabilizing the ground state. (2-8)

GTPase-activating protein (GAP): a protein that accelerates the intrinsic GTPase activity of switch GTPases. (3-7)

guanine-nucleotide-binding protein: see **G protein**.

guanine-nucleotide exchange factor (GEF): a protein that facilitates exchange of GDP for GTP in switch GTPases. (3-7)

hairpin turn: another name for **beta turn**.

helical parameters: set of numerical values that define the geometry of a helix. These include the number of residues per turn, the translational rise per residue, and the main-chain torsional angles. (1-6)

helix dipole: the macrodipole that is thought to be formed by the cumulative effect of the individual peptide dipoles in an alpha helix. The positive end of the dipole is at the beginning (amino terminus) of the helix; the negative end is at the carboxyl terminus of the helical rod. (1-6)

heptad repeat: a sequence in which hydrophobic residues occur every seven amino acids, a pattern that is reliably indicative of a **coiled-coil** interaction between two alpha helices in which the hydrophobic side chains of each helix interdigitate with those of the other. (1-19)

heterotetramer: an assembly of four subunits of more than one kind of polypeptide chain. (1-19)

heterotrimeric G protein: a GTPase switch protein composed of three different subunits, an α subunit with GTPase activity, and associated β and γ subunits, found associated with the cytoplasmic tails of G-protein-coupled receptors, where it acts to relay signals from the receptor to downstream targets. Exchange of bound GDP for GTP on the α subunit causes dissociation of the heterotrimer into a free α subunit and a $\beta\gamma$ heterodimer; hydrolysis of the bound GTP causes reassociation of the subunits. (3-8)

hexamer: an assembly of six identical or different subunits. In a protein the subunits are individual folded polypeptide chains. (1-19)

Hidden Markov Model: a probabilistic model of a sequence **alignment**. (4-1)

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 192

homologous: describes genes or proteins related by divergent evolution from a common ancestor. Homologous proteins, or homologs, will generally have similar sequences, structures and biochemical functions, although the sequence and/or functional similarity may be difficult to recognize. (4-1)

homology: the similarity seen between two gene or protein sequences that are both derived by evolution from a common ancestral sequence. (4-1)

homology modeling: a computational method for modeling the structure of a protein based on its sequence similarity to one or more proteins of known structure. (4-6)

homotrimer: an assembly of three identical subunits: in a protein, these are individual folded polypeptide chains. (1-19)

hydride ion: a hydrogen atom with an extra electron. (2-9)

hydrogen bond: a noncovalent interaction between the **donor atom**, which is bound to a positively polarized hydrogen atom, and the acceptor atom, which is negatively polarized. Though not covalent, the hydrogen bond holds the donor and **acceptor atom** close together. (1-4)

hydrolysis: breaking a covalent bond by addition of a molecule of water. (1-3)

hydrophilic: tending to interact with water. Hydrophilic molecules are polar or charged and, as a consequence, are very soluble in water. In polymers, hydrophilic **side chains** tend to associate with other hydrophilic side chains, or with water molecules, usually by means of hydrogen bonds. (1-1)

hydrophobic: tending to avoid water. Hydrophobic molecules are nonpolar and uncharged and, as a consequence, are relatively insoluble in water. In polymers, hydrophobic **side chains** tend to associate with each other to minimize their contact with water or polar side chains. (1-1)

hydrophobic effect: the tendency of nonpolar groups in water to self-associate and thereby minimize their contact surface area with the polar solvent. (1-9)

induced fit: originally, the change in the structure of an enzyme, induced by binding of the substrate, that brings the catalytic groups into proper alignment. Now generalized to the idea that specific ligands can induce the protein conformation that results in optimal binding interactions. (1-22, 2-2)

intein: a protein intron (intervening sequence). An internal portion of a protein sequence that is post-translationally excised in an autocatalytic reaction while the flanking regions are spliced together, making an additional protein product. (3-17)

interaction domain: a protein that recognizes another protein, usually via a specific recognition motif. (3-1)

intermediate: a species that forms transiently along the path from substrate to product. (2-6)

intron: an intervening sequence in a gene that does not correspond to any portion of the final protein sequence and is spliced out of the RNA transcript before translation. (1-2)

jelly roll fold: a beta sandwich built from two sheets with topologies resembling a Greek key design. The sheets pack almost at right-angles to each other. (1-17)

K_d: the dissociation constant for the binding of a ligand to a macromolecule. Typical values range from 10^{-3} M to 10^{-10} M. The lower the K_d, the tighter the ligand binds. (1-13)

ligand: small molecule or macromolecule that recognizes and binds to a specific site on a macromolecule. (2-1)

ligand-binding site: site on the surface of a protein at which another molecule binds. (2-1)

limited proteolysis: specific cleavage by a protease of a limited number of the peptide bonds in a protein substrate. The fragments thus produced may remain associated or may dissociate. (1-13)

lipid anchor: lipid attached to a protein that inserts into a membrane, thereby anchoring the protein to the bilayer. (3-2)

lipidation: covalent attachment of a fatty-acid group to a protein. (3-19)

lipid bilayer: the structure of cellular membranes, formed when two sheets of lipid molecules pack against each other with their hydrophobic tails forming the interior of the sandwich and their polar headgroups covering the outside. (1-6)

local alignment: alignment of only a part of a sequence with a part of another. (4-2)

mesophilic: favoring moderate temperatures. Mesophilic organisms normally cannot tolerate extremes of heat or cold. Mesophilic enzymes typically denature at moderate temperatures (over 40 °C or so). (1-12)

messenger RNA (mRNA): the RNA molecule transcribed from a gene sequence after removal of **introns** and editing. (1-2)

metastable: only partially stable under the given conditions. In the case of protein structures, a metastable fold exists in equilibrium with other conformations or with the unfolded state. (4-15)

methylation: modification, usually of a nitrogen or oxygen atom of an amino-acid side chain, by addition of a methyl group. Some bases on DNA and RNA can also be methylated. (3-20)

mixed beta sheet: beta sheet containing both parallel and antiparallel strands. (1-7)

monomer: a single subunit: in a protein, this is a folded peptide chain. (1-19)

motif: characteristic sequence or structure that in the case of a **structural motif** may comprise a whole domain or protein but usually consists of a small local arrangement of secondary structure elements which then coalesce to form domains. **Sequence motifs**, which are recognizable amino-acid sequences found in different proteins, usually indicate biochemical function. Structural motifs are less commonly associated with specific biochemical functions. (1-16)

multifunctional: having a number of distinct biochemical functions in one gene product. (2-15)

multiple sequence alignment: alignment of more than two sequences to maximize their overall mutual identity or similarity. (4-1)

myristoylation: irreversible attachment of a myristoyl group to a protein via an amide linkage. (3-19)

***N*-acetylation:** covalent addition of an acetyl group from acetyl-CoA to a nitrogen atom at either the amino-terminus of a polypeptide chain or in a lysine side-chain. The reaction is catalyzed by *N*-acetyltransferase. (1-13, 3-20)

native state: the stably folded and functional form of a biological macromolecule. (1-9)

negative cooperativity: binding of one molecule of a ligand to a protein makes it more difficult for a second molecule of that ligand to bind at another site. (3-4)

nitrosylation: modification of the –SH group of a cysteine residue by addition of nitric oxide produced by nitric oxide synthase. (3-20)

northern blot: technique for detecting and identifying individual RNAs by hybridization to specific nucleic acid probes, after separation of a complex mixture of mRNAs by electrophoresis and blotting onto a nylon membrane. (4-4)

nucleophile: a group that is electron-rich, such as an alkoxide ion ($-O^-$), and can donate electrons to an electron-deficient center. (2-14)

nucleotide: the basic repeating unit of a nucleic acid polymer. It consists of a **base** (A, U [in RNA, T in DNA], G or C), a sugar (ribose in RNA, deoxyribose in DNA) and a phosphate group. (1-2)

nucleotide-binding fold: an open parallel beta sheet with connecting alpha helices that is usually used to bind NADH or NADPH. It contains a characteristic sequence motif that is involved in binding the cofactor. Also known as the Rossmann fold. (1-18)

oligomer: an assembly of more than one subunit: in a protein, the subunits are individual folded polypeptide chains. (1-19)

oxidation: the loss of electrons from an atom or molecule. (2-10)

oxyanion hole: a binding site for an alkoxide in an enzyme active site. The "hole" is a pocket that fits the $-O^-$ group precisely, and has two hydrogen-bond-donating groups that stabilize the oxyanion with $-O^-\cdots H-X$ hydrogen bonds. (2-14)

packing motif: an arrangement of secondary structure elements defined by the number and types of such elements and the angles between them. The term motif is used in structural biology in a number of contexts and thus can be confusing. (1-10)

pairwise alignment: alignment of two sequences. (4-1)

palmitoylation: reversible attachment of a palmitoyl group to a protein via a thioester linkage. (3-19)

parallel beta sheet: a beta sheet, formed from non-contiguous regions of the polypeptide chain, in which every strand runs in the same direction. (1-7)

partner swapping: exchange of one protein for another in multiprotein complexes. (2-4)

pentamer: an assembly of five identical or different subunits: in a protein, these are individual folded polypeptide chains. (1-19)

peptide bond: another name for **amide bond**, a chemical bond formed when a carboxylic acid condenses with an amino group with the expulsion of a water molecule. The term peptide bond is used only when both groups come from amino acids. (1-3)

percent identity: the percentage of columns in an **alignment** of two sequences that contain identical amino acids. Columns that include gaps are not counted. (4-1)

phase problem: in the measurement of data from an X-ray crystallographic experiment only the amplitude of the wave is determined. To compute a structure, the phase must also be known. Since it cannot be determined directly, it must be determined indirectly or by some other experiment. (5-2)

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 193

phi torsion angle: see **torsion angle**.

phosphate-binding loop: see **P-loop**.

phosphorylation: covalent addition of a phosphate group, usually to one or more amino-acid side chains on a protein, catalyzed by protein kinases. (1-13)

phylogenetic tree: a branching diagram, usually based on the evolutionary distances between sequences, that illustrates the evolutionary history of a protein family or superfamily, or the relationships between different species of organism. (4-1)

pK_a value: strictly defined as the negative logarithm of the equilibrium constant for the acid-base equation. For ranges of pK_a between 0 and 14, it can be thought of as the pH of an aqueous solution at which a proton-donating group is half protonated and half deprotonated. pK_a is a measure of the proton affinity of a group: the lower the pK_a, the more weakly the proton is held. (2-12)

pleated sheet: another name for **beta sheet**.

P-loop: a conserved loop in GTPase- and ATPase-based nucleotide switch proteins that binds to phosphate groups in the bound nucleotide. (3-6)

polypeptide: a polymer of amino acids joined together by **peptide bonds**. (1-3)

positive cooperativity: binding of one molecule of a ligand to a protein makes it easier for a second molecule of that ligand to bind at another site. (3-4)

prenylation: irreversible attachment of either a farnesyl or geranylgeranyl group to a protein via thioether linkage. (3-19)

primary structure: the amino-acid sequence of a polypeptide chain. (1-2)

primary transcript: the RNA molecule directly transcribed from a gene, before removal of introns or other editing. (1-2)

profile: a table or matrix of information that characterizes a protein family or superfamily. It is typically composed of sequence variation or identity with respect to a reference sequence, expressed as a function of each position in the amino-acid sequence of a protein. It can be generalized to include structural information. Three-dimensional profiles express the three-dimensional structure of a protein as a table which represents the local environment and conformation of each residue. (4-2)

propinquity factor: another term for **proximity factor**.

proteasome: a multiprotein complex that degrades ubiquitinated proteins into short peptides. (3-11)

protein kinase: enzyme that transfers a phosphate group from ATP to the OH group of serines, threonines and tyrosines of target proteins. Kinases that phosphorylate carboxylates and histidines also occur as part of **two-component systems** in prokaryotes, fungi and plants, but not in animals. (3-2)

protein phosphatase: enzyme that specifically removes phosphate groups from phosphorylated serines, threonines or tyrosines on proteins. (3-12)

proteolytic cascade: a sequential series of protein cleavages by proteases, each cleavage activating the next protease in the cascade. (3-16)

protomer: the asymmetric repeating unit (or units) from which an oligomeric protein is built up. (1-21)

proximity factor: the concept that a reaction will be facilitated if the reacting species are brought close together in an orientation appropriate for chemistry to occur. (2-8)

pseudosymmetric: having approximate but not exact symmetry. A protein with two non-identical subunits of very similar three-dimensional structure is a pseudosymmetric dimer. (1-21)

psi torsion angle: see **torsion angle**.

quaternary structure: the subunit structure of a protein. (1-19)

Ramachandran plot: a two-dimensional plot of the values of the backbone torsion angles phi and psi, with allowed regions indicated for conformations where there is no steric interference. Ramachandran plots are used as a diagnosis for accurate structures: when the phi and psi torsion angles of an experimentally determined protein structure are plotted on such a diagram, the observed values should fall predominantly in the allowed regions. (1-5)

reaction sub-site: that part of the active site where chemistry occurs. (2-7)

redox reactions: reactions in which oxidation and reduction occur. (2-10)

reducing environment: a chemical environment in which the reduced states of chemical groups are favored. In a reducing environment, free –S–H groups are favored over –S–S– bridges. The interior of most cells is a highly reducing environment. (1-4)

reduction: the gain of electrons by an atom or molecule. (2-10)

residue: the basic building block of a polymer; the fragment that is released when the bonds that hold the polymer segments together are broken. In proteins, the residues are the amino acids. (1-1)

resolution: the level of detail that can be derived from a given process. (5-1)

resonance: delocalization of bonding electrons over more than one chemical bond in a molecule. Resonance greatly increases the stability of a molecule. It can be represented, conceptually, as if the properties of the molecule were an average of several structures in which the chemical bonds differ. (1-3)

reverse turn: another name for **beta turn**.

RGS protein: regulator of G-protein signaling protein; protein that binds to the free GTP-bound α subunit of a **heterotrimeric G protein** and stimulates its GTPase activity. (3-8)

RNA editing: enzymatic modification of the mRNA base sequence. It may involve changes in the bases or the insertion of entirely new stretches of bases. RNA editing produces a protein sequence that does not correspond precisely to the sequence of amino acids that would be predicted from the gene sequence by the genetic code. (1-2)

RNA interference (RNAi): abolition of the expression of a gene by a small (~22 base pair) double-stranded RNA. (4-4)

***S*-acylation:** reversible attachment of a fatty-acid group to a protein via a thioester linkage; **palmitoylation** is an example of *S*-acylation. (3-19)

salt bridge: a **hydrogen bond** in which both donor and acceptor atoms are fully charged. The bonding energy of a salt bridge is significantly higher than that of a hydrogen bond in which only one participating atom is fully charged or in which both are partially charged. (1-4)

scaffold protein: a protein that serves as a platform onto which other proteins assemble to form functional complexes. (2-5)

secondary structure: folded segments of a polypeptide chain with repeating, characteristic phi, psi backbone torsion angles, that are stabilized by a regular pattern of hydrogen bonds between the peptide –N–H and C=O groups of different residues. (1-5)

side chain: a chemical group in a polymer that protrudes from the repeating backbone. In proteins, the side chain, which is bonded to the alpha carbon of the backbone, gives each of the 20 amino acids its particular chemical identity. Glycine has no side chain, and the end of the side chain of proline is fused to the nitrogen of the backbone, creating a closed ring. (1-1)

single-nucleotide polymorphism (SNP): a mutation of a single base in a codon, usually one that does not affect the identity of the amino acid that it encodes. (1-2)

specificity sub-site: that part of the active site where recognition of the ligand takes place. (2-7)

stop codon: a codon that signals the end of the coding sequence and usually terminates **translation**. (1-2)

stress-response proteins: proteins whose synthesis is induced when cells are subjected to environmental stress, such as heat. (3-11)

structural domain: a compact part of the overall structure of a protein that is sufficiently independent of the rest of the molecule to suggest that it could fold stably on its own. (2-3)

substrate: the molecule that is transformed in a reaction. (2-6)

sumoylation: modification of the side chain of a lysine residue by addition of a small ubiquitin-like protein (SUMO). The covalent attachment is an amide bond between the carboxy-terminal carboxylate of SUMO and the NH_2 on the lysine side chain of the targeted protein. (3-20)

superfamily: proteins with the same overall fold but with usually less than 40% sequence identity. The nature of the biochemical functions performed by proteins in the same superfamily are more divergent than those within families. For instance, members of the same enzyme superfamily may not catalyze the same overall reaction, yet still retain a common mechanism for stabilizing chemically similar rate-limiting transition-states and intermediates, and will do so with similar active site residues. (4-8)

switch I region: a conserved sequence motif in GTPase- and ATPase-based nucleotide switch proteins that, with the **switch II region**, binds the terminal gamma-phosphate in the triphosphate form of the bound nucleotide and undergoes a marked conformational change when the nucleotide is hydrolyzed. (3-6)

switch II region: a conserved sequence motif in GTPase- and ATPase-based nucleotide switch proteins that, with the **switch I region**, binds the terminal gamma-phosphate in the triphosphate form of the bound nucleotide and undergoes a marked conformational change when the nucleotide is hydrolyzed. (3-6)

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 194

temperature-sensitive: losing structure and/or function at temperatures above physiological or room temperature. A temperature-sensitive mutation is a change in the amino-acid sequence of a protein that causes the protein to inactivate or fail to fold properly at such temperatures. (1-12)

temperature-sensitive mutants: organisms containing a genetic mutation that makes the resulting protein sensitive to slightly elevated temperatures. The temperature at which the mutant protein unfolds is called the restrictive temperature. The term is also used for the protein itself. (3-11)

tertiary structure: the folded conformation of a protein, formed by the condensation of the various secondary elements, stabilized by a large number of weak interactions. (1-10)

tetramer: an assembly of four identical or different subunits. (1-19)

thermophilic: favoring high temperatures. A thermophilic organism is one that requires high temperatures (above approximately 50 °C) for survival. A thermophilic enzyme is one that functions optimally and is stable at temperatures at which **mesophilic** proteins denature. (1-12)

TIM barrel: another name for the alpha/beta barrel fold. (1-18)

torsion angle: the angle between two groups on either side of a rotatable chemical bond. If the bond is the C_α–N bond of a peptide backbone the torsion angle is called **phi**. If the bond is the C_α–C backbone bond, the angle is called **psi**. (1-3)

transcription: the synthesis of RNA from the coding strand of DNA by DNA-dependent RNA polymerase. (1-2)

transition state: the species of highest free energy either in a reaction or a step of a reaction; the highest region on the **activation-energy barrier**. (2-6)

translation: the transfer of genetic information from the sequence of **codons** on **mRNA** into a sequence of amino acids and the synthesis on the ribosome of the corresponding polypeptide chain. (1-2)

trimer: an assembly of three identical or different subunits. (1-19)

two-component systems: signal transduction systems found in bacteria and some eukaryotes involving a membrane-bound histidine kinase and a cytoplasmic response regulator protein that is activated by phosphorylation. (3-15)

ubiquitin: a small protein that when attached to other proteins (**ubiquitination**), targets them for degradation to the **proteasome**. Sometimes ubiquitin tagging targets a protein to other fates such as endocytosis. (3-11)

ubiquitination: the attachment of ubiquitin to a protein. (3-11)

up-and-down structural motif: a simple fold in which beta strands in an antiparallel sheet are all adjacent in sequence and connectivity. (1-17)

van der Waals interaction: a weak attractive force between two atoms or groups of atoms, arising from the fluctuations in electron distribution around the nuclei. Van der Waals forces are stronger between less electronegative atoms such as those found in hydrophobic groups. (1-4)

yeast two-hybrid: a method for finding proteins that interact with another protein, based on activation of a reporter gene in yeast. (4-4)

zinc finger: a small, irregular domain stabilized by binding of a zinc ion. Zinc fingers usually are found in eukaryotic DNA-binding proteins. They contain signature metal-ion binding sequence motifs. (1-18)

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 195

References

Abel, K., Yoder, M.D., Hilgenfeld, R. and Jurnak, F.: **An alpha to beta conformational switch in EF-Tu.** *Structure* 1996, **4**:1153–1159. (3-9)

Aitken, A.: **Protein consensus sequence motifs.** *Mol. Biotechnol.* 1999, **12**:241–253. (1-16, 4-2)

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P.: *Molecular Biology of the Cell* 4th ed. (Garland, New York, 2002). (1-0, 1-2, 2-0)

Al-Lazikani, B., Jung, J., Xiang, Z. and Honig, B.: **Protein structure prediction.** *Curr. Opin. Chem. Biol.* 2001, **5**:51–56. (4-6)

Allen, K.N., Bellamacina, C.R., Ding, X., Jeffrey, C.J., Mattos, C.J., Petsko, G.A. and Ringe, D.: **An experimental approach to mapping the binding surfaces of crystalline proteins.** *J. Phys. Chem.* 1996, **100**:2605–2611. (4-9)

Almo, S.C., Smith, D.L., Danishefsky, A.T. and Ringe, D.: **The structural basis for the altered substrate specificity of the R292D active site mutant of aspartate aminotransferase from *E. coli*.** *Protein Eng.* 1994, **7**:405–412. (2-7)

Aloy, P., Querol, E., Aviles, F.X. and Sternberg, M.J.: **Automated structure-based prediction of functional sites in proteins: applications to assessing the validity of inheriting protein function from homology in genome annotation and to protein docking.** *J. Mol. Biol.* 2001, **311**:395–408. (4-9)

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J.: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res.* 1997, **25**:3389–3402. (4-2)

Anston, A.A., Dodson, E.J. and Dodson, G.G.: **Circular assemblies.** *Curr. Opin. Struct.* Biol. 1996, **6**:142–150. (1-19)

Argyle, E.: **A similarity ring for amino acids based on their evolutionary substitution rates.** *Orig. Life* 1980, **10**:357–360. (1-2)

Arrondo, J.L. and Goni, F.M.: **Structure and dynamics of membrane proteins as studied by infrared spectroscopy.** *Prog. Biophys. Mol. Biol.* 1999, **72**:367–405. (1-22)

Authier, F., Metioui, M., Fabrega, S., Kouach, M. and Briand, G.: **Endosomal proteolysis of internalized insulin at the C-terminal region of the B chain by cathepsin D.** *J. Biol. Chem.* 2002, **277**:9437–9446. (3-3)

Authier, F., Posner, B.I. and Bergeron, J.J.: **Endosomal proteolysis of internalized proteins.** *FEBS Lett.* 1996, **389**:55–60. (3-3)

Babbitt, P.C. and Gerlt, J.A.: **Understanding enzyme superfamilies. Chemistry as the fundamental determinant in the evolution of new catalytic activities.** *J. Biol. Chem.* 1997, **272**:30591–30594. (4-16)

Babbitt, P.C., Hasson, M.S., Wedekind, J.E., Palmer, D.R., Barrett, W.C., Reed, G.H., Rayment, I., Ringe, D., Kenyon, G.L. and Gerlt, J.A.: **The enolase superfamily: a general strategy for enzyme-catalyzed abstraction of the alpha-protons of carboxylic acids.** *Biochemistry* 1996, **35**:16489–16501. (4-16)

Babbitt, P.C., Mrachko, G.T., Hasson, M.S., Huisman, G.W., Kolter, R., Ringe, D., Petsko, G.A., Kenyon, G.L. and Gerlt, J.A.: **A functionally diverse enzyme superfamily that abstracts the alpha protons of carboxylic acids.** *Science* 1995, **267**:1159–1161. (4-11, 4-16)

Badger, J., Minor, I., Oliveira, M.A., Smith, T.J. and Rossmann, M.G.: **Structural analysis of antiviral agents that interact with the capsid of human rhinoviruses.** *Proteins* 1989, **6**:1–19. (2-1)

Baker, D. and Sali, A.: **Protein structure prediction and structural genomics.** *Science* 2001, **294**:93–96. (4-6)

Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A.: **The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution.** *Science* 2000, **289**:905–920. (2-5)

Barford, D., Hu, S.H. and Johnson, L.N.: **Structural mechanism for glycogen phosphorylase control by phosphorylation and AMP.** *J. Mol. Biol.* 1991, **218**:233–260. (3-12)

Barlow, D.J. and Thornton, J.M.: **Helix geometry in proteins.** *J. Mol. Biol.* 1988, **201**:601–619. (1-10)

Bartlett, G.J., Porter, C.T., Borkakoti, N. and Thornton, J.M.: **Analysis of catalytic residues in enzyme active sites.** *J. Mol. Biol.* 2002, **324**:105–121. (4-10)

Berisio, R., Vitagliano, L., Mazzarella, L. and Zagari, A.: **Recent progress on collagen triple helix structure, stability and assembly.** *Curr. Pharm. Des.* 2002, **9**:107–116. (2-5)

Beissinger, M., Sticht, H., Sutter, M., Ejchart, A., Haehnel, W., and Rosch, P.: **Solution structure of cytochrome c6 from the thermophilic cyanobacterium *Synechococcus elongatus*.** *EMBO J.* 1998, **2**:27–36. (2-4)

Bellamacina, C.R.: **The nicotinamide dinucleotide binding motif: a comparison of nucleotide binding proteins.** *FASEB J.* 1996, **10**:1257–1269. (1-18)

Benaroudj, N., Tarcsa, E., Cascio, P. and Goldberg, A.L.: **The unfolding of substrates and ubiquitin-independent protein degradation by proteasomes.** *Biochimie* 2001, **83**:311–318. (3-11)

Bennett, M.J., Choe, S. and Eisenberg, D.: **Domain swapping: entangling alliances between proteins.** *Proc. Natl Acad. Sci. USA* 1994, **91**:3127–3131. (2-4)

Bernstein, B.E. and Hol, W.G.: **Crystal structures and products bound to the phosphoglycerate kinase active site reveal the catalytic mechanism.** *Biochemistry* 1998, **37**:4429–4436. (2-9)

Bernstein, B.E. Michels, P.A. and Hol, W.G.: **Synergistic effects of substrate-induced conformational changes in phosphoglycerate kinase activation.** *Nature* 1997, **385**:275–278. (2-9)

Bitetti-Putzer, R., Joseph-McCarthy, D., Hogle, J. M. and Karplus, M.: **Functional group placement in protein binding sites: a comparison of GRID and MCSS.** *J. Comput Aided Mol Des.* 2001, **15**:935–960. (4-9)

Bond, C.J., Wong, K.B., Clarke, J., Fersht, A.R. and Daggett, V.: **Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: description of the folding pathway.** *Proc. Natl Acad. Sci. USA* 1997, **94**:13409–13413. (1-9)

Bonneau, R., Tsai, J., Ruczinski, I., Chivian, D., Rohl, C., Strauss, C.E. and Baker, D.: **Rosetta in CASP4: Progress in *ab initio* protein structure prediction.** *Proteins* 2001, **45(S5)**:119–126. (4-7)

Bork, P., Holm, L. and Sander, C.: **The immunoglobulin fold. Structural classification, sequence patterns and common core.** *J. Mol. Biol.* 1994, **242**:309–320. (1-17)

Bosshard, H.R., Durr, E., Hitz, T. and Jelesarov, I.: **Energetics of coiled coil folding: the nature of the transition state.** *Biochemistry.* 2001, **40**:3544–3552. (1-19)

Bourne, Y., Arvai, A.S., Bernstein, S.L., Watson, M.H., Reed, S.I., Endicott, J.E., Noble, M.E., Johnson, L.N. and Tainer, J.A.: **Crystal structure of the cell cycle-regulatory protein suc1 reveals a beta-hinge conformational switch.** *Proc. Natl Acad. Sci. USA* 1995, **92**:10232–10236. (2-4)

Bowie, J.U., Luthy, R. and Eisenberg, D.A.: **A method to identify protein sequences that fold into a known three-dimensional structure.** *Science* 1991, **253**:164–170. (4-7)

Branden, C. and Tooze, J.: *Introduction to Protein Structure* 2nd ed. (Garland, New York, 1999). (1-15, 1-17, 1-18)

Brazhnik, P., de la Fuente, A. and Mendes, P.: **Gene networks: how to put the function in genomics.** *Trends Biotechnol.* 2002, **20**:467–472. (4-0)

Brenner, S.: **Theoretical biology in the third millennium.** *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 1999, **354**:1963–1965. (4-3)

Brent, R. and Finley, R.L. Jr.: **Understanding gene and allele function with two-hybrid methods.** *Annu. Rev. Genet.* 1997, **31**:663–704. (4-17)

Brizuela, L., Richardson, A., Marsischky, G., and Labaer, J.: **The FLEXGene repository: exploiting the fruits of the genome projects by creating a needed resource to face the challenges of the post-genomic era.** *Arch. Med. Res.* 2002, **33**:318–324. (4-3)

Brown, N.R., Noble, M.E., Endicott, J.A. and Johnson, L.N.: **The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases.** *Nat. Cell Biol.* 1999, **1**:438–443. (3-14)

Brown, N.R., Noble, M.E., Lawrie, A.M., Morris, M.C., Tunnah, P., Divita, G., Johnson, L.N. and Endicott, J.A.: **Effects of phosphorylation of threonine 160 on cyclin-dependent kinase 2 structure and activity.** *J. Biol. Chem.* 1999, **274**:8746–8756. (3-14)

Bruckner, K., Perez, L., Clausen, H. and Cohen, S.: **Glycosyltransferase activity of Fringe modulates Notch-Delta interactions.** *Nature* 2000, **406**:411–415. (3-18)

Bruice, T.C. and Benkovic, S.J.: **Chemical basis for enzyme catalysis.** *Biochemistry* 2000, **39**:6267–6274. (2-6)

Bullock, T.L., Clarkson, W.D., Kent, H.M. and Stewart, M.: **The 1.6 angstroms resolution crystal structure of nuclear transport factor 2 (ntf2).** *J. Mol. Biol.* 1996, **260**:422–431. (4-5)

Bunn, H.F. and Forget, B.G.: *Hemoglobin: Molecular, Genetic and Clinical Aspects* (Saunders, New York, 1986). (1-20)

Burchett, S.A.: **Regulators of G protein signaling: a bestiary of modular protein binding domains.** *J. Neurochem.* 2000, **75**:1335–1351. (1-14)

Burley, S.K. and Petsko, G.A.: **Weakly polar interactions in proteins.** *Adv. Prot. Chem.* 1988, **39**:125–189. (1-4)

Butikofer, P., Malherbe, T., Boschung, M. and Roditi, I.: **GPI-anchored proteins: now you see 'em, now you don't.** *FASEB J.* 2001, **15**:545–548. (3-19)

Byerly, D.W., McElroy, C.A. and Foster, M.P.: **Mapping the surface of *Escherichia coli* peptide deformylase by NMR with organic solvents.** *Protein Science* 2002, **11**:1850–1853. (4-9)

Campbell, I.D. and Downing, A.K.: **Building protein structure and function from modular units.** *Trends Biotechnol.* 1994, **12**:168–172. (1-14)

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 196

Capaldi, R.A. and Aggeler, R.: **Mechanism of the F(1)F(0)-type ATP synthase, a biological rotary motor.** *Trends Biochem. Sci.* 2002, **27**:154–160. (3-10)

Cardozo, T., Batalov, S. and Abagyan, R.: **Estimating local backbone structural deviation in homology models.** *Comput. Chem.* 2000, **24**:13–31. (4-6)

Casey, P.J.: **Lipid modifications** in *Encyclopedic Reference of Molecular Pharmacology.* Offermanns, S. and Rosenthal, W. eds (Springer-Verlag, Heidelberg, 2003) in the press. (3-19)

Casey, P.J.: **Protein lipidation in cell signaling.** *Science* 1995, **268**:221–225. (3-19)

Chan, T.-F., Carvalho, J., Riles, L. and Zheng, X.F.: **A chemical genomics approach toward understanding the global functions of the target of rapamycin protein (TOR).** *Proc. Natl Acad. Sci. USA* 2000, **97**:13227–13232. (4-0)

Chatterjee, S. and Mayor, S.: **The GPI-anchor and protein sorting.** *Cell. Mol. Life Sci.* 2001, **58**:1969–1987. (3-19)

Chevalier, B.S. and Stoddard, B.L.: **Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility.** *Nucleic Acids Res.* 2001, **29**:3757–3774. (3-17)

Chook, Y.M., Ke, H., and Lipscomb, W.N.: **Crystal structures of the monofunctional chorismate mutase from *Bacillus subtilis* and its complex with a transition state analog.** *Proc. Natl Acad. Sci. USA* 1993, **90**:8600–8603. (2-8)

Chothia, C.: **Asymmetry in protein structure.** *Ciba Foundation Symp.* 1991, **162**:36–49. (1-18)

Chou, P.Y. and Fasman, G.D.: **Prediction of protein conformation.** *Biochemistry* 1974, **13**:222–245. (1-8)

Cline, M., Hughey, R. and Karplus, K.: **Predicting reliable regions in protein sequence alignments.** *Bioinformatics* 2002, **18**:306–314. (4-6)

Cohen, C. and Parry, D.A.: **Alpha-helical coiled coils and bundles: how to design an alpha-helical protein.** *Proteins* 1990, **7**:1–15. (1-19)

Cohen, B.I., Presnell, S.R. and Cohen, F.E.: **Origins of structural diversity within sequentially identical hexapeptides.** *Protein Sci.* 1993, **2**:2134–2145. (4-15)

Colas, P. and Brent, R.: **The impact of two-hybrid and related methods on biotechnology.** *Trends Biotechnol.* 1998, **16**:355–363. (4-4)

Cook, A., Lowe, E.D., Chrysina, E.D., Skamnaki, V.T., Oikonomakos, N.G. and Johnson, L.N.: **Structural studies on phospho-CDK2/cyclin A bound to nitrate, a transition state analogue: implications for the protein kinase mechanism.** *Biochemistry* 2002, **41**:7301–7311. (3-14)

Cozzone, A.J.: **Regulation of acetate metabolism by protein phosphorylation in enteric bacteria.** *Annu. Rev. Microbiol.* 1998, **52**:127–164. (3-12)

Cramer, W.A, Engelman, D.M., Von Heijne, G. and Rees, D.C.: **Forces involved in the assembly and stabilization of membrane proteins.** *FASEB J.* 1992, **6**:3397–3402. (1-9)

Creighton, T.E.: *Proteins: Structure and Molecular Properties* 2nd ed. (Freeman, New York, 1993). (1-1, 1-19)

Cronin, C.N. and Kirsch, J.F.: **Role of arginine-292 in the substrate specificity of aspartate aminotransferase as examined by site-directed mutagenesis.** *Biochemistry*. 1988, **27**:4572–4579. (2-7)

Cutforth, T. and Gaul, U.: **A methionine aminopeptidase and putative regulator of translation initiation is required for cell growth and patterning in *Drosophila*.** *Mech Dev.* 1999, **82**:23–28. (4-13)

Cyert, M.S.: **Regulation of nuclear localization during signaling.** *J. Biol. Chem.* 2001, **276**:20805–20808. (3-2)

Dalal, S. and Regan, L.: **Understanding the sequence determinants of conformational switching using protein design.** *Protein Sci.* 2000, **9**:1651–1659. (4-15)

Daggett, V.: **Long timescale simulations.** *Curr. Opin. Struct. Biol.* 2000, **10**:160–164. (1-22)

Darnell, J.E. Jr.: **STATs and gene regulation.** *Science* 1997, **277**:1630–1635. (2-4)

Datta, B.: **MAPs and POEP of the roads from prokaryotic to eukaryotic kingdoms.** *Biochimie* 2000, **82**:95–107. (4-13)

Dayhoff, M.O, Barker, W.C. and Hunt, L.C.: **Establishing homologies in protein sequences.** *Methods Enzymol.* 1983, **91**:524–545. (1-2)

Dean, A.M. and Koshland, D.E. Jr.: **Electrostatic and steric contributions to regulation at the active site of isocitrate dehydrogenase.** *Science* 1990, **249**:1044–1046. (3-12)

Deane, C.M., Allen, F.H., Taylor, R. and Blundell, T.L.: **Carbonyl-carbonyl interactions stabilize the partially allowed Ramachandran conformations of asparagine and aspartic acid.** *Protein Eng.* 1999, **12**:1025–1028. (1-5)

de la Cruz, X. and Thornton, J.M.: **Factors limiting the performance of prediction-based fold recognition methods.** *Protein Sci.* 1999, **8**:750–759. (1-16, 4-7)

Deleage, G., Blanchet, C. and Geourjou, C.: **Protein structure prediction. Implications for the biologist.** *Biochimie* 1997, **79**:681–686. (1-8)

Dengler, U., Siddiqui, A.S. and Barton, G.J.: **Protein structural domains: analysis of the 3Dee domains database.** *Proteins* 2001, **42**:332–344. (1-14)

Dennis, S., Kortvelyesi, T. and Vajda, S.: **Computational mapping identifies the binding sites of organic solvents on proteins.** *Proc. Natl Acad. Sci. USA* 2002, **99**:4290–4295. (4-9)

Desai, A. and Mitchison, T.J.: **Microtubule polymerization dynamics.** *Annu. Rev. Cell Dev. Biol.* 1997, **13**:83–117. (2-0)

Dhillon, A.S. and Kolch, W.: **Untying the regulation of the Raf-1 kinase.** *Arch. Biochem. Biophys.* 2002, **404**:3–9. (3-2)

Di Gennaro, J.A., Siew, N., Hoffman, B.T., Zhang, L., Skolnick, J., Neilson, L.I. and Fetrow, J.S.: **Enhanced functional annotation of protein sequences via the use of structural descriptors.** *J. Struct. Biol.* 2001, **134**:232–245. (4-10)

Dill, K.A. and Bromberg, S.: *Molecular Driving Forces: Statistical Thermodynamics in Chemistry and Biology* (Garland Science, New York and London 2003). (1-12)

Ding, X., Rasmussen, B.F., Petsko, G.A. and Ringe, D.: **Direct structural observation of an acyl-enzyme intermediate in the hydrolysis of an ester substrate by elastase.** *Biochemistry* 1994, **33**:9285–9293. (2-2)

Dinner, A.R., Sali, A., Smith, L.J., Dobson, C.M. and Karplus, M.: **Understanding protein folding via free-energy surfaces from theory and experiment.** *Trends Biochem. Sci.* 2000, **25**:331–339. (1-9)

Dobson, C.M.: **Protein misfolding, evolution and disease.** *Trends Biochem Sci.* 1999, **24**:329–332. (4-15)

Domingues, F.S., Lackner, P., Andreeva, A. and Sippl, M.J.: **Structure-based evaluation of sequence comparison and fold recognition alignment accuracy.** *J. Mol. Biol.* 2000, **297**:1003–1013. (4-3)

Doolittle R.F.: **Convergent evolution: the need to be explicit.** *Trends Biochem. Sci.* 1994, **19**:15–18. (4-12)

Dorn, G.W. 2nd and Mochly-Rosen, D.: **Intracellular transport mechanisms of signal transducers.** *Annu. Rev. Physiol.* 2002, **64**:407–429. (3-2)

Doyle, D.A., Morais Cabral, J., Pfuetzner, R.A., Kuo, A., Gulbis, J.M., Cohen, S.L., Chait, B.T. and MacKinnon, R.: **The structure of the potassium channel: molecular basis of K^+ conduction and selectivity.** *Science* 1998, **280**:69–77. (1-11)

Drenth, J.: *Principles of Protein X-Ray Crystallography* 2nd ed. (Springer-Verlag, New York, 1999). (5-2)

Dunitz, J.D.: **Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions.** *Chem. Biol.* 1995, **2**:709–712. (1-4)

Eaton, W.A., Munoz, V., Hagen, S.J., Jas, G.S., Lapidus, L.J., Henry, E.R. and Hofrichter, J.: **Fast kinetics and mechanisms in protein folding.** *Annu. Rev. Biophys. Biomol. Struct.* 2000, **29**:327–259. (1-9)

Eilers, M., Shekar, S.C., Shieh, T., Smith, S.O. and Fleming, P.J.: **Internal packing of helical membrane proteins.** *Proc. Natl Acad. Sci. USA* 2000, **97**:5796–5801. (1-10)

Elia, A.E., Cantley, L.C. and Yaffe, M.B.: **Proteomic screen finds pSer/pThr-binding domain localizing Plk1 to mitotic substrates.** *Science* 2003, **299**: 1228–1231. (3-1)

Elion, E.A.: **The Ste5p scaffold.** *J. Cell Sci.* 2001, **114**:3967–3978. (2-5)

Elofsson, A., Fischer, D., Rice, D.W., Le Grand, S.M. and Eisenberg, D.: **A study of combined structure/sequence profiles.** *Fold. Des.* 1996, **1**:451–461. (4-2)

Engh, R.A., Huber, R., Bode, W. and Schulze, A.J.: **Divining the serpin inhibition mechanism: a suicide substrate "springe"?** *Trends Biotechnol.* 1995, **13**:503–510. (4-15)

English, A.C., Groom, C.R. and Hubbard, R.E.: **Experimental and computational mapping of the binding surface of a crystalline protein**. *Protein Eng.* 2001, **14**:47–59. (4-9)

Eswaramoorthy, S., Gerchman, S., Graziano, V., Kycia, H., Studier, F.W. and Swaminathan, S.: **Structure of a yeast hypothetical protein selected by a structural genomics approach.** *Acta Crystallogr. D. Biol. Crystallogr.* 2003, **59**:127–135. (4-17)

Evans J.N.S.: *Biomolecular NMR Spectroscopy* (Oxford University Press, Oxford, 1995). (5-2)

Ewing, T.J., Makino, S., Skillman A.G. and Kuntz, I.D.: **DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases.** *J. Comput. Aided Mol. Des.* 2001, **15**:411–428. (4-10)

Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C.J., Hofmann, K. and Bairoch, A.: **The PROSITE database, its status in 2002.** *Nucleic Acids Res.* 2002, **30**:235–238. (4-2)

Fan, J.S. and Zhang, M.: **Signaling complex organization by PDZ domain proteins.** *Neurosignals* 2002, **11**:315–321. (3-1)

Ferguson, A.D., Hofmann, E., Coulton, J.W., Diederichs, K. and Welte, W.: **Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide.** *Science* 1998, **282**:2215–2220. (1-11)

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 197

Ferreira, S.T. and De Felice, F.G.: **Protein dynamics, folding and misfolding: from basic physical chemistry to human conformational diseases.** *FEBS Lett.* 2001, **498**:129–134. (1-12)

Fersht, A.: *Structure and Mechanism in Protein Science. A Guide to Enzyme Catalysis and Protein Folding.* (Freeman, New York, 1999). (1-9)

Fersht, A.R.: **The hydrogen bond in molecular recognition.** *Trends Biochem. Sci.* 1987, **12**:301–304. (1-4)

Fetrow, J.S., Siew, N., Di Gennaro, J.A., Martinez-Yamout, M., Dyson, H.J. and Skolnick, J.: **Genomic-scale comparison of sequence- and structure-based methods of function prediction: does structure provide additional insight?** *Protein Sci.* 2001, **10**:1005–1014. (4-6)

Fetrow, J.S., Siew, N. and Skolnick, J.: **Structure-based functional motif identifies a potential disulfide oxidoreductase active site in the serine/threonine protein phosphatase-1 subfamily.** *FASEB J.* 1999, **13**:1866–1874. (4-10)

Fischer, D. and Eisenberg, D.: **Protein fold recognition using sequence-derived predictions.** *Protein Sci.* 1996, **5**:947–955. (4-7)

Garrington, T.P. and Johnson, G.L.: **Organization and regulation of mitogen-activated protein kinase signaling pathways.** *Curr. Opin. Cell Biol.* 1999, **2**:211–218. (3-2)

Gasch, A.P., Huang, M., Metzner, S., Botstein, D., Elledge, S.J. and Brown, P.O.: **Genomic expression programs in the response of yeast cells to environmental changes.** *Mol. Biol. Cell* 2000, **11**:4241–4257. (4-4)

Gaucher, E.A., Gu, X., Miyamoto, M.M. and Benner, S.A.: **Predicting functional divergence in protein evolution by site-specific rate shifts.** *Trends Biochem. Sci.* 2002, **27**:315–321. (4-2)

Gerlt, J.A. and Babbitt, P.C.: **Can sequence determine function?** *Genome Biol.* 2000, **1**: reviews 0005.1–0005.10. (4-11)

Gerlt, J.A. and Babbitt, P.C.: **Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies.** *Annu. Rev. Biochem.* 2001, **70**:209–246. (4-8, 4-11)

Gerstein, M. and Honig, B.: **Sequences and topology.** *Curr. Opin. Struct. Biol.* 2001, **11**:327–329. (4-1)

Geyer, M. and Wittinghofer A.: **GEFs, GAPs, GDIs and effectors: taking a closer (3D) look at the regulation of Ras-related GTP-binding proteins.** *Curr. Opin. Struct. Biol.* 1997, **7**:786–792. (3-7)

Goldberg, A.L., Elledge, S.J. and Harper, J.W.: **The cellular chamber of doom.** *Sci. Am.* 2001, **284**:68–73. (3-11)

Goldstein, L.S.: **Molecular motors: from one motor many tails to one motor many tales.** *Trends Cell Biol.* 2001, **11**:477–482. (3-10)

Gonzalez, L. Jr., Woolfson, D.N. and Alber, T.: **Buried polar residues and structural specificity in the GCN4 leucine zipper.** *Nat. Struct. Biol.* 1996, **3**:1011–1018. (1-19)

Goodford, P.J.: **A computational procedure for determining energetically favorable binding sites on biologically important macromolecules.** *J. Med. Chem.* 1985, **28**:849–857. (4-9)

Goodsell, D.S.: **Inside a living cell.** *Trends Biochem. Sci.* 1991, **16**:203–206. (3-0)

Goodsell, D.S.: **Structural symmetry and protein function.** *Annu. Rev. Biophys. Biomol. Struct.* 2000, **29**:105–153. (1-21)

Greasley, S.E., Horton, P., Ramcharan, J., Beardsley, G.P., Benkovic, S.J. and Wilson, I.A.: **Crystal structure of a bifunctional transformylase and cyclohydrolase in purine biosynthesis.** *Nat. Struct. Biol.* 2001, **8**:402–406. (2-15)

Greenfield, N.J. and Fasman, G.D.: **Computed circular dichroism spectra for the evaluation of protein conformation.** *Biochemistry* 1969, **8**:4108–4116. (1-12)

Gribskov, M., McLachlan, A.D. and Eisenberg, D.: **Profile analysis: detection of distantly related proteins.** *Proc. Natl Acad. Sci. USA* 1987, **84**:4355–4358. (4-2)

Griffith, E.C., Su, Z., Turk, B.E., Chen, S., Chang, Y.H., Wu, Z., Biemann, K. and Liu, J.O.: **Methionine aminopeptidase (type 2) is the common target for angiogenesis inhibitors AGM-1470 and ovalicin.** *Chem Biol.* 1997, **4**:461–471. (4-13)

Guttmacher A.E. and Collins, F.S.: **Genomic medicine—a primer.** *N. Engl. J. Med.* 2002, **347**:1512–1520. (4-0)

Haga, A., Niinaka, Y. and Raz, A.: **Phosphohexose isomerase/autocrine motility factor/neuroleukin/maturation factor is a multifunctional phosphoprotein.** *Biochim. Biophys. Acta.* 2000, **1480**:235–244. (4-13)

Hall, T.M., Porter, J.A., Young, K.E., Koonin, E.V., Beachy, P.A. and Leahy, D.J.: **Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins.** *Cell* 1997, **91**:85–97. (3-17)

Hamad, N.M., Elconin, J.H., Karnoub, A.E., Bai, W., Rich, J.N., Abraham, R.T., Der, C.J. and Counter, C.M.: **Distinct requirements for Ras oncogenesis in human versus mouse cells.** *Genes Dev.* 2002, **16**:2045–2057. (3-7)

Hamm, H.E. and Gilchrist, A.: **Heterotrimeric G proteins.** *Curr. Opin. Cell Biol.* 1996, **8**:189–196. (3-8)

Hammes, G.G.: **Multiple conformational changes in enzyme catalysis.** *Biochemistry* 2002, **41**:8221–8228. (2-2)

Han, G.W., Lee, J.Y., Song, H.K., Chang, C., Min, K., Moon, J., Shin, D.H., Kopka, M.L., Sawaya, M.R., Yuan, H.S., Kim, T.D., Choe, J., Lim, D., Moon, H.J. and Suh, S.W.: **Structural basis of non-specific lipid binding in maize lipid-transfer protein complexes revealed by high-resolution X-ray crystallography.** *J. Mol. Biol.* 2001, **308**:263–278. (2-4)

Harrington, D.J., Adachi, K. and Royer, W.E. Jr.: **The high resolution structure of deoxyhemoglobin S.** *J. Mol. Biol.* 1997, **272**:398–407. (1-20)

Hasson, M.S., Muscate, A., McLeish, M., Polovnikova, L.S., Gerlt, J.A., Kenyon. G.L., Petsko, G.A. and Ringe, D.: **The crystal structure of benzoylformate decarboxylase at 1.6Å resolution: diversity of catalytic residues in thiamine diphosphate dependent enzymes.** *Biochemistry* 1998, **37**:9918–9930. (4-5)

Hasson, M.S., Schlichting, I., Moulai, J., Taylor, K., Kenyon, G.L., Babbitt, P.C., Gerlt, J.A. Petsko, G.A. and Ringe, D.: **Evolution of an enzyme active site: the structure of a new crystal form of muconate lactonizing enzyme compared with mandelate racemase and enolase.** *Proc. Natl Acad. Sci. USA* 1998, **95**:10396–10401. (4-11)

Hawkins, A.R. and Lamb, H.K.: **The molecular biology of multidomain proteins. Selected examples.** *Eur. J. Biochem.* 1995, **232**:7–18. (1-14)

Hegyi, H. and Bork, P.: **On the classification and evolution of protein molecules.** *J. Prot. Chem.* 1997, **16**:545–551. (1-14)

Hegyi, H. and Gerstein, M.: **Annotation transfer for genomics: measuring functional divergence in multi-domain proteins.** *Genome Res.* 2001, **11**:1632–1640. (4-3)

Heller, H., Schaefer, M. and Schulten, K.: **Molecular dynamics simulation of a bilayer of 200 lipids in the gel and in the liquid crystal phases.** *J. Phys. Chem.* 1993, **97**:8343–8360. (1-11)

Herskowitz, I.: **Functional inactivation of genes by dominant negative mutations.** *Nature* 1987, **329**:219–222. (1-20)

Highbarger, L.A., Gerlt, J.A., and Kenyon, G.L.: **Mechanism of the reaction catalyzed by acetoacetate decarboxylase. Importance of lysine 116 in determining the pK_a of active-site lysine.** *Biochemistry* 1996, **35**:41–46. (2-1)

Hochstrasser, M.: **SP-RING for SUMO: New functions bloom for a ubiquitin-like protein.** *Cell* 2001, **107**:5–8. (3-20)

Hodgson, D.A. and Thomas, C.M. (eds): **Signals, switches, regulons, and cascades: control of bacterial gene expression.** *61st Symposium of the Society for General Microbiology* (Cambridge University Press, Cambridge, 2002). (3-0)

Hogg, T., Mesters, J.R. and Hilgenfeld, R.: **Inhibitory mechanisms of antibiotics targeting elongation factor Tu.** *Curr. Protein Pept. Sci.* 2002, **3**:121–131. (3-9)

Hol, W.G.: **The role of the alpha helix dipole in protein function and structure.** *Prog. Biophys. Mol. Biol.* 1985, **45**:149–195. (1-6)

Holmes, K.C. and Geeves, M.A.: **The structural basis of muscle contraction.** *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 2000, **355**:419–431. (3-10)

Holyoak, T., Wilson, M.A., Fenn, T.D., Petsko, G.A., Fuller, R.S. and Ringe, D.: **The 2.4 Å crystal structure of the processing protease Kex2 in complex with an Ala–Lys–Arg boronic acid inhibitor.** *Biochemistry* in the press. (5-3)

Hook, V.Y., Azaryan, A.V., Hwang, S.R. and Tezapsidis, N.: **Proteases and the emerging role of protease inhibitors in prohormone processing.** *FASEB J.* 1994, **8**:1269–1278. (3-16)

Houdusse, A., Szent-Gyorgyi, A.G. and Cohen, C.: **Three conformational states of scallop myosin S1.** *Proc. Natl Acad. Sci. USA.* 2000, **97**:11238–11243. (2-5)

Houry, W.A., Frishman, D., Eckerskorn, C., Lottspeich, F. and Hartl, F.U.: **Identification of *in vivo* substrates of the chaperonin GroEL.** *Nature* 1999, **402**:147–154. (4-0)

Huang, X., Holden, H.M. and Raushel, F.M.: **Channeling of substrates and intermediates in enzyme-catalyzed reactions.** *Annu. Rev. Biochem.* 2001, **70**:149–180. (2-16)

Hunter, T.: **Signaling—2000 and beyond.** *Cell* 2000, **100**:113–127. (3-12)

Hurley, J.H., Dean, A.M., Koshland, D.E. Jr. and Stroud, R.M.: **Catalytic mechanism of NADP(+)-dependent isocitrate dehydrogenase: implications from the structures of magnesium-isocitrate and NADP+ complexes.** *Biochemistry* 1991, **30**:8671–8678. (2-15)

Hurley, J.H., Dean, A.M., Sohl, J.L., Koshland, D.E. Jr. and Stroud, R.M.: **Regulation of an enzyme by phosphorylation at the active site.** *Science* 1990, **249**:1012–1016. (3-12)

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 198

Huse, M. and Kuriyan, J.: **The conformational plasticity of protein kinases.** *Cell* 2002, **109**:275–282. (3-13)

Hyde, C.C., Ahmed, S.A., Padlan, E.A., Miles, E.W. and Davies, D.R.: **Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from *Salmonella typhimurium*.** *J. Biol. Chem.* 1988, **263**:17857–17871. (2-16)

Imada, K., Sato, M., Tanaka N., Katsube, Y., Matsuura, Y. and Oshima, T.: **Three-dimensional structure of a highly thermostable enzyme, 3-isopropylmalate dehydrogenase of *Thermus thermophilus* at 2.2 Å resolution.** *J. Mol. Biol.* 1991, **222**:725–738. (2-3)

Imperiali, B. and O'Connor, S.E.: **Effect of *N*-linked glycosylation on glycopeptide and glycoprotein structure.** *Curr. Opin. Chem. Biol.* 1999, **3**:643–649. (3-18)

Irving, J.A., Whisstock, J.C., and Lesk, A.M.: **Protein structural alignments and functional genomics.** *Proteins* 2001, **42**:378–382. (4-5)

Ishima, R. and Torchia, D.A.: **Protein dynamics from NMR.** *Nat. Struct. Biol.* 2000, **7**:740–743. (1-22)

Jaenicke, R.: **Stability and stabilization of globular proteins in solution.** *J. Biotechnol.* 2000, **79**:193–203. (1-12)

James, C.L. and Viola, R.E.: **Production and characterization of bifunctional enzymes. Substrate channeling in the aspartate pathway.** *Biochemistry* 2002, **41**:3726–3731. (2-16)

Jansen, R. and Gerstein, M.: **Analysis of the yeast transcriptome with structural and functional categories: characterizing highly expressed proteins.** *Nucleic Acids Res.* 2000, **28**:1481–1488. (1-0)

Jeffery, C.J.: **Moonlighting proteins.** *Trends Biochem. Sci.* 1999, **24**:8–11. (4-13)

Jeffery, C.J., Bahnson, B.J., Chien, W., Ringe, D. and Petsko, G.A.: **Crystal structure of rabbit phosphoglucose isomerase, a glycolytic enzyme that moonlights as neuroleukin, autocrine motility factor, and differentiation mediator.** *Biochemistry* 2000, **39**:955–964. (4-13)

Jencks, W.P.: **Binding energy, specificity, and enzymic catalysis: the Circe effect.** *Adv. Enzymol. Relat. Areas Mol. Biol.* 1975, **43**:219–410. (2-8)

Jencks, W.P.: *Catalysis and Enzymology* (Dover Publications, New York, 1987). (2-6)

Jensen, R.B. and Shapiro, L.: **Proteins on the move: dynamic protein localization in prokaryotes.** *Trends Cell Biol.* 2000, **10**:483–488. (3-0)

Jenuwein, T. and Allis, C.D.: **Translating the histone code.** *Science*, **293**:1074–1080. (3-20)

Jeong, J. and McMahon, A.P.: **Cholesterol modification of Hedgehog family proteins.** *J. Clin. Invest.* 2002, **110**:591–596. (3-17)

Johnson, L.N. and O'Reilly, M.: **Control by phosphorylation.** *Curr. Opin. Struct. Biol.* 1996, **6**:762–769. (3-12)

Jones, D.T. and Swindells, M.B.: **Getting the most from PSI-BLAST.** *Trends Biochem. Sci* 2002, **27**:161–164. (4-2)

Jones, D.T., Taylor, W.R. and Thornton, J.M.: **The rapid generation of mutation data matrices from protein sequences.** *Comput. Appl. Biosci.* 1992, **8**:275–282. (1-2)

Jones, S. and Thornton, J.M.: **Principles of protein-protein interactions** *Proc. Natl Acad. Sci. USA* 1996, **93**:13–20. (1-19)

Jones, S. and Thornton, J.M.: **Prediction of protein–protein interaction sites using patch analysis.** *J. Mol. Biol.* 1997, **272**:133–143. (4-10)

Jouanguy, E., Lamhamedi-Cherradi, S., Lammas, D., Dorman, S.E., Fondaneche, M.C., Dupuis, S., Doffinger, R., Altare, F., Girdlestone, J., Emile, J.F., Ducoulombier, H., Edgar, D., Clarke, J., Oxelius, V.A., Brai, M., Novelli, V., Heyne, K., Fischer, A., Holland, S.M., Kumararatne, D.S., Schreiber, R.D. and Casanova, J.L.: **A human IFNGR1 small deletion hotspot associated with dominant susceptibility to mycobacterial infection.** *Nat. Genet.* 1999, **21**:370–378. (1-20)

Kabsch, W. and Sander, C.: **On the use of sequence homologies to predict protein structure: identical pentapeptides can have completely different conformations.** *Proc. Natl Acad. Sci. USA* 1984, **81**:1075–1078. (4-15)

Kalafatis, M., Egan, J.O., van 't Veer, C., Cawthern, K.M. and Mann, K.G.: **The regulation of clotting factors.** *Crit. Rev. Eukaryot. Gene. Expr.* 1997, **7**:241–280. (3-16)

Kallal, L. and Benovic, J.L.: **Using green fluorescent proteins to study G-protein-coupled receptor localization and trafficking.** *Trends Pharmacol Sci.* 2000, **21**:175–180. (4-4)

Kantrowitz, E.R. and Lipscomb, W.N.: ***Escherichia coli* aspartate transcarbamylase: the relation between structure and function.** *Science* 1988, **241**:669–674. (3-5)

Karpusas, M., Branchaud, B. and Remington, S.J.: **Proposed mechanism for the condensation reaction of citrate synthase: 1.9Å structure of the ternary complex with oxaloacetate and carboxymethyl coenzyme A.** *Biochemistry* 1990, **29**:2213–2219. (2-9)

Kauzmann, W.: **Some factors in the interpretation of protein denaturation.** *Adv. Protein Chem.* 1959, **14**:1–63. (1-12)

Kawabata, T., Fukuchi, S., Homma, K., Ota, M., Araki, J., Ito, T., Ichiyoshi, N. and Nishikawa, K.: **GTOP: a database of protein structures predicted from genome sequences.** *Nucleic Acids Res.* 2002, **30**:294–298. (4-2)

Karplus, M. and Petsko, G.A.: **Molecular dynamics simulations in biology.** *Nature* 1990, **347**:631–639. (1-22)

Kerrebrock, A.W., Moore, D.P., Wu, J.S. and Orr-Weaver, T.L.: **Mei-S332, a *Drosophila* protein required for sister-chromatid cohesion, can localize to meiotic centromere regions.** *Cell* 1995, **83**:247–256. (4-4)

Kim, S.W., Cha, S.S., Cho, H.S., Kim, J.S., Ha, N.C., Cho, M.J., Joo, S., Kim, K.K., Choi, K.Y. and Oh, B.H.: **High-resolution crystal structures of delta5-3-ketosteroid isomerase with and without a reaction intermediate analogue.** *Biochemistry* 1997, **36**: 14030–14036. (4-5)

Kirsch, J.F., Eichele, G., Ford, G.C., Vincent, M.G., Jansonius, J.N., Gehring, H. and Christen, P.: **Mechanism of action of aspartate aminotransferase proposed on the basis of its spatial structure.** *J. Mol. Biol.* 1984, **174**:497–525. (4-12)

Kisselev, A.F. and Goldberg, A.L.: **Proteasome inhibitors: from research tools to drug candidates.** *Chem. Biol.* 2001, **8**:739–758. (3-11)

Klabunde, T., Sharma, S., Telenti, A., Jacobs, W.R. Jr. and Sacchettini, J.C.: **Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing.** *Nat. Struct. Biol.* 1998, **5**:31–36. (3-17)

Knighton, D.R., Kan, C.C., Howland, E., Janson, C.A., Hostomska, Z., Welsh, K.M. and Matthews, D.A.: **Structure and kinetic channelling in bifunctional dihydrofolate reductase-thyidylate synthase.** *Nat. Struct. Biol.* 1994, **1**:186–194. (2-15)

Ko, Y.H. and Pedersen, P.L.: **Cystic fibrosis: a brief look at some highlights of a decade of research focused on elucidating and correcting the molecular basis of the disease.** *J. Bioenerg. Biomembr.* 2001, **33**:513–521. (4-15)

Koebnik, R., Locher, K.P. and Van Gelder, P.: **Structure and function of bacterial outer membrane proteins: barrels in a nutshell.** *Mol. Microbiol.* 2000, **37**:239–253. (1-11)

Koonin E.V., Wolf, Y.I. and Karev, G.P.: **The structure of the protein universe and genome evolution.** *Nature* 2002, **420**:218–223. (4-0)

Kornitzer, D. and Ciechanover, A.: **Modes of regulation of ubiquitin-mediated protein degradation.** *J. Cell Physiol.* 2000, **182**:1–11. (3-0)

Koshland, D.E. Jr. and Hamadani, K.: **Proteomics and models for enzyme cooperativity.** *J. Biol. Chem.* 2002, **277**:46841–46844. (3-4)

Koshland, D.E. Jr., Nemethy, G. and Filmer, D.: **Comparison of experimental binding data and theoretical models in proteins containing subunits.** *Biochemistry* 1966, **5**:365–385. (2-2)

Kossiakoff, A.A. and De Vos, A.M.: **Structural basis for cytokine hormone–receptor recognition and receptor activation.** *Adv. Protein Chem.* 1998, **52**:67–108. (2-3)

Kouzarides, T.: **Histone methylation in transcriptional control.** *Curr. Opin. Genet. Dev.* 2002, **12**:198–209. (3-20)

Krause, K.L., Volz, K.W., and Lipscomb, W.N.: **2.5 Å structure of aspartate carbamoyltransferase complexed with the bisubstrate analog *N*-(phosphonoacetyl)-L-aspartate.** *J. Mol. Biol.* 1987, **193**:527–553. (2-8)

Krem, M.M., Rose, T. and Di Cera, E.: **Sequence determinants of function and evolution in serine proteases.** *Trends Cardiovasc. Med.* 2000, **10**:171–176. (4-8)

Kunin, V., Chan, B., Sitbon, E., Lithwick, G. and Pietrokovski, S.: **Consistency analysis of similarity between multiple alignments: prediction of protein function and fold structure from analysis of local sequence motifs.** *J. Mol. Biol.* 2001, **307**:939–949. (4-2)

Kuriyan, J. and Cowburn, D.: **Modular peptide recognition domains in eukaryotic signaling.** *Annu. Rev. Biophys. Biomol. Struct.* 1997, **26**:259–288. (3-1)

Kurosaki, T.: **Regulation of B-cell signal transduction by adaptor proteins.** *Nat. Rev. Immunol.* 2002, **2**:354–363. (3-2)

Kyte, J. and Doolittle, R.F.: **A simple method for displaying the hydropathic character of a protein.** *J. Mol. Biol.* 1982, **157**:105–132. (1-11)

Lahiri, S.D., Zhang, G., Dunaway-Mariano, D. and Allen, K.N.: **Caught in the act: the structure of phosphorylated beta-phosphoglucomutase from *Lactococcus lactis*.** *Biochemistry* 2002, **41**:8351–8359. (2-14)

Landro, J.A., Gerlt, J.A., Kozarich, J.W., Koo, C.W., Shah, V.J., Kenyon, G.L., Neidhart, D.J., Fujita, S. and Petsko, G.A.: **The role of lysine 166 in the mechanism of mandelate racemase from *Pseudomonas putida*: mechanistic and crystallographic evidence for stereospecific alkylation by (R)-alpha-phenylglycidate.** *Biochemistry* 1994, **33**:635–643. (2-1)

Laney, J.D. and Hochstrasser, M.: **Substrate targeting in the ubiquitin system.** *Cell* 1999, **97**:427–430. (3-11)

Laskowski, R.A., Luscombe, N.M., Swindells, M.B. and Thornton, J.M.: **Protein clefts in molecular recognition and function.** *Protein Sci.* 1996, **5**:2438–2452. (4-10)

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 199

Lau, E.Y., Kahn, K., Bash, P.A., and Bruice, T.C.: **The importance of reactant positioning in enzyme catalysis: a hybrid quantum mechanics/molecular mechanics study of a haloalkane dehalogenase.** *Proc. Natl Acad. Sci. USA* 2000, **97**:9937–9942. (2-8)

Leclerc, E., Peretz, D., Ball, H., Sakurai, H., Legname, G., Serban, A., Prusiner, S.B., Burton, D.R. and Williamson, R.A.: **Immobilized prion protein undergoes spontaneous rearrangement to a conformation having features in common with the infectious form.** *EMBO J.* 2001, **20**:1547–1554. (4-15)

Lee, A.Y., Gulink, S.V. and Erickson, J.W.: **Conformational switching in an aspartic proteinase.** *Nat. Struct. Biol.* 1998, **5**:866–871. (3-3)

Lee, M. and Goodbourn, S.: **Signalling from the cell surface to the nucleus.** *Essays Biochem.* 2001, **37**:71–85. (3-12)

Leon, O. and Roth, M.: **Zinc fingers: DNA binding and protein-protein interactions.** *Biol. Res.* 2000, **33**:21–30. (1-18)

Lesk, A.M.: *Introduction to Protein Architecture* (Oxford University Press, Oxford, 2001). (4-5)

Lesk, A.M. and Chothia, C.: **Solvent accessibility, protein surfaces and protein folding.** *Biophys. J.* 1980, **32**:35–47. (1-10)

Liang, J., Edelsbrunner, H. and Woodward, C.: **Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design.** *Protein Sci.* 1998, **7**:1884–1897. (4-9)

Liang, P.H. and Anderson, K.S.: **Substrate channeling and domain-domain interactions in bifunctional thymidylate synthase-dihydrofolate reductase.** *Biochemistry* 1998, **37**:12195–12205. (2-15)

Liepinsh, E. and Otting, G.: **Organic solvents identify specific ligand binding sites on protein surfaces.** *Nat. Biotechnol.* 1997, **15**:264–268. (4-9)

Liu, J., Schmitz, J.C., Lin, X., Tai, N., Yan, W., Farrell, M., Bailly, M., Chen, T. and Chu, E.: **Thymidylate synthase as a translational regulator of cellular gene expression.** *Biochim. Biophys. Acta* 2002, **1587**:174–182. (2-15)

Liu, S., Widom, J., Kemp, C.W., Crews, C.M. and Clardy, J.: **Structure of human methionine aminopeptidase-2 complexed with fumagillin.** *Science* 1998, **282**:1324–1327. (4-13)

Liu, Y., Hart, J.P., Schlunegger, M.P. and Eisenberg, D.: **The crystal structure of a 3D domain-swapped dimer of RNase A at a 2.1-Å resolution.** *Proc. Natl Acad. Sci. USA* 1998, **95**:3437–3442. (2-4)

Lockless, S.W. and Ranganathan, R.: **Evolutionarily conserved pathways of energetic connectivity in protein families.** *Science* 1999, **286**:295–299. (4-6)

Lubetsky, J.B., Dios, A., Han, J., Aljabari, B., Ruzsicska, B., Mitchell, R., Lolis, E. and Al-Abed, Y.: **The tautomerase active site of macrophage migration inhibitory factor is a potential target for discovery of novel anti-inflammatory agents.** *J. Biol. Chem.* 2002, **277**:24976–24982. (4-13)

Lundqvist, T., Rice, J., Hodge, C.N., Basarab, G.S., Pierce, J. and Lindqvist, Y.: **Crystal structure of scytalone dehydratase—a disease determinant of the rice pathogen, *Magnaporthe grisea*.** *Structure* 1994, **2**:937–944. (4-5)

Macias, M.J., Wiesner, S. and Sudol, M.: **WW and SH3 domains, two different scaffolds to recognize proline-rich ligands.** *FEBS Lett.* 2002, **513**:30–37. (3-1)

Mahadeva, R., Dafforn, T.R., Carrell, R.W. and Lomas, D.A.: **6-mer peptide selectively anneals to a pathogenic serpin conformation and blocks polymerization. Implications for the prevention of Z alpha(1)-antitrypsin-related cirrhosis.** *J. Biol. Chem.* 2002, **277**:6771–6774. (4-15)

Malcolm, B.A, Rosenberg, S., Corey, M.J, Allen, J.S., de Baetselier, A. and Kirsch, J.F.: **Site-directed mutagenesis of the catalytic residues Asp-52 and Glu-35 of chicken egg white lysozyme.** *Proc. Natl Acad. Sci. USA* 1989, **86**:133–137. (2-12)

Manning, G., Whyte, D.B., Martinez, R., Hunter, T. and Sudarsanam, S.: **The protein kinase complement of the human genome.** *Science* 2002, **298**:1912–1934. (3-12)

Markley, J.L., Ulrich, E.L., Westler W.M. and Volkman, B.F.: **Macromolecular structure determination by NMR spectroscopy.** *Methods Biochem. Anal.* 2003, **44**:89–113. (5-2)

Marmorstein, R., Carey, M., Ptashne, M. and Harrison, S.C.: **DNA recognition by GAL4: structure of a protein-DNA complex.** *Nature* 1992, **356**:408–414. (2-3)

Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F. and Sali, A.: **Comparative protein structure modeling of genes and genomes.** *Annu. Rev. Biophys.* 2000, **29**:291–325. (4-6)

Martin, R.B.: **Peptide bond characteristics.** *Met. Ions Biol. Syst.* 2001, **38**:1–23. (1-3)

Martin, T.F.: **PI(4,5)P(2) regulation of surface membrane traffic.** *Curr. Opin. Cell Biol.* 2001, **13**:493–499. (3-2)

Martz, E.: **Protein Explorer: easy yet powerful macromolecular visualization.** *Trends Biochem. Sci.* 2002, **27**:107–109. (5-3)

Matthews, B.W. and Bernhard, S.A.: **Structure and symmetry of oligomeric enzymes.** *Annu. Rev. Biophys. Bioeng.* 1973, **2**:257–317. (1-21)

Mattos, C., Petsko, G.A. and Karplus, M.: **Analysis of two-residue turns in proteins.** *J. Mol. Biol.* 1994, **238**:733–747. (1-5)

Mattos, C. and Ringe, D.: **Locating and characterizing binding sites on proteins.** *Nat. Biotechnol.* 1996, **14**:595–599. (4-9)

Mattos, C. and Ringe, D.: **Proteins in organic solvents.** *Curr. Opin. Struct. Biol.* 2001, **11**:761–764. (4-9)

McBride, A.E. and Silver, P. A.: **State of the Arg: protein methylation at arginine comes of age.** *Cell* 2001, **106**:5–8. (3-20)

McKessar, S.J., Berry, A.M., Bell, J.M., Turnidge, J.D. and Paton, J.C.: **Genetic characterisation of vanG, a novel vancomycin resistance locus of *Enterococcus faecalis*.** *Antimicrob. Agents Chemother.* 2000, **44**:3224–3228. (1-8)

Mesecar, A.D., Stoddard, B.L., and Koshland, D.E. Jr.: **Orbital steering in the catalytic power of enzymes: small structural changes with large catalytic consequences.** *Science* 1997, **277**:202–206. (2-8)

Mezei, M.: **Chameleon sequences in the PDB.** *Protein Eng.* 1998, **11**:411–414. (4-14)

Michal, G., ed.: *Boehringer Mannheim Biochemical Pathways Wallcharts*, Roche Diagnostics Corporation, Roche Molecular Biochemicals, P.O. Box 50414, Indianapolis, IN 46250-0414, USA. (1-0)

Miller, B.G. and Wolfenden, R.: **Catalytic proficiency: the unusual case of OMP decarboxylase.** *Annu. Rev. Biochem.* 2002, **71**:847–885. (2-6)

Miller, R.T., Jones, D.T. and Thornton, J.M.: **Protein fold recognition by sequence threading: tools and assessment techniques.** *FASEB J.* 1996, **10**:171–178. (4-7)

Milner-White, E.J.: **Description of the quaternary structure of tetrameric proteins. Forms that show either right-handed or left-handed symmetry at the subunit.** *Biochem. J.* 1980, **187**:297–302. (1-21)

Minor, D.L. Jr. and Kim, P.S.: **Context-dependent secondary structure formation of a designed protein sequence.** *Nature* 1996, **380**:730–734. (4-14)

Miranker, A. and Karplus, M.: **Functionality maps of binding sites: a multiple copy simultaneous search method.** *Proteins* 1991, **11**:29–34. (4-9)

Modis, Y., Trus, B.L. and Harrison, S.C.: **Atomic model of the papillomavirus capsid.** *EMBO J.* 2002, **21**:4754–4762. (2-4)

Morris, A.J. and Malbon, C.C.: **Physiological regulation of G protein-linked signaling.** *Physiol. Rev.* 1999, **79**:1373–1430. (1-13)

Mount, D.W.: *Bioinformatics: Sequence and Genome analysis* (Cold Spring Harbor Laboratory Press, New York, 2001). (4-1)

Murzin, A.G.: **How far divergent evolution goes in proteins.** *Curr. Opin. Struct. Biol.* 1998, **8**:380–387. (4-5)

Myers, J.K. and Oas, T.G.: **Reinterpretation of GCN4-p1 folding kinetics: partial helix formation precedes dimerization in coiled coil folding.** *J. Mol. Biol.* 1999, **289**:205–209. (1-19)

Nagano, N., Orengo, C.A. and Thornton, J.M.: **One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions.** *J. Mol. Biol.* 2002, **321**:741–765. (4-11)

Neidhart, D.J., Kenyon, G.L., Gelt, J.A. and Petsko, G.A.: **Mandelate racemase and muconate lactonizing enzyme are mechanistically distinct and structurally homologous.** *Nature* 1990, **347**:692–694. (4-11)

Nogales, E.: **Structural insights into microtubule function.** *Annu. Rev. Biochem.* 2000, **69**:277–302. (2-0)

Nyborg, J. and Liljas, A.: **Protein biosynthesis: structural studies of the elongation cycle.** *FEBS Lett.* 1998, **430**:95–99. (3-9)

O'Donovan, C., Apweiler, R. and Bairoch, A.: **The human proteomics initiative (HPI).** *Trends Biotechnol.* 2001, **19**:178–181. (4-0)

Oliver, S.G.: **Functional genomics: lessons from yeast.** *Philos Trans. R. Soc. Lond. B. Biol. Sci.* 2002, **357**:17–23. (4-0)

Okada, T., Ernst, O.P., Palczewski, K. and Hofmann, K.P.: **Activation of rhodopsin: new insights from structural and biochemical studies.** *Trends Biochem. Sci.* 2001, **26**:318–324. (3-8)

Ondrechen, M.J., Clifton, J.G. and Ringe, D.: **THEMATICS: a simple computational predictor of enzyme function from structure.** *Proc. Natl Acad. Sci. USA* 2001, **98**:12473–12478. (4-10)

Palade, G.E.: **Protein kinesis: the dynamics of protein trafficking and stability.** *Cold Spring Harbor Symp. Quant. Biol.* 1995, **60**:821–831. (1-13)

Pannifer, A.D., Wong, T.Y., Schwarzenbacher, R., Renatus, M., Petosa, C., Bienkowska, J., Lacy, D.B., Collier, R.J., Park, S., Leppla, S.H., Hanna, P. and Liddington, R.C.: **Crystal structure of the anthrax lethal factor.** *Nature* 2001, **414**:229–233. (2-1)

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 200

Park, H.W. and Beese, L.S.: **Protein farnesyltransferase.** *Curr. Opin. Struct. Biol.* 1997, **7**:873–880. (3-19)

Parkinson, J.S. and Kofoid, E.C.: **Communication modules in bacterial signaling proteins.** *Annu. Rev. Genet.* 1992, **26**:71–112. (3-15)

Parekh, R.B. and Rohlff, C.: **Post-translational modification of proteins and the discovery of new medicine.** *Curr. Opin. Biotechnol.* 1997, **8**:718–723. (3-18)

Patterson, S.D.: **Proteomics: the industrialization of protein chemistry.** *Curr. Opin. Biotechnol.* 2000, **11**:413–418. (4-4)

Patthy, L.: **Genome evolution and the evolution of exon-shuffling.** *Gene* 1999, **238**:103–114. (1-15)

Patthy, L.: Protein Evolution (Blackwell Science, Oxford, 1999). (4-5)

Pauling, L. and Corey, R.B.: **Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets.** *Proc. Natl Acad. Sci. USA* 1951, **37**:729–740. (1-7)

Pauling, L.C., Corey, R.B., and Branson, H.R.: **The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain.** *Proc. Natl Acad. Sci. USA.* 1951, **37**:205–211. (1-6)

Pauling, L.C.: *The Nature of the Chemical Bond and the Structure of Molecules and Crystals*, 3rd ed. (Cornell Univ. Press, Ithaca, New York, 1960). (1-3, 1-4)

Paulus, H.: **Protein splicing and related forms of protein autoprocessing.** *Annu. Rev. Biochem.* 2000, **69**:447–496. (3-17)

Pawson, A.J. (ed): *Protein Modules in Signal Transduction* (Springer, Berlin, New York, 1998). (3-0)

Peisach, E., Wang, J., de los Santos, T., Reich, E. and Ringe, D.: **Crystal structure of the proenzyme domain of plasminogen.** *Biochemistry* 1999, **38**:11180–11188. (3-16)

Pellegrini, M., Marcotte, E.M., Thompson, M.J., Eisenberg, D. and Yeates, T.O.: **Assigning protein functions by comparative genome analysis: protein phylogenetic profiles.** *Proc. Natl Acad. Sci. USA* 1999, **96**:4285–4288. (4-2)

Penn, R.B., Pronin, A.N., and Benovic J.L.: **Regulation of G protein-coupled receptor kinases.** *Trends Cardiovasc. Med.* 2000, **10**:81–89. (3-2)

Pereira-Leal, J.B., Hume, A.N. and Seabra, M.C.: **Prenylation of Rab GTPases: molecular mechanisms and involvement in genetic disease.** *FEBS Lett.* 2001, **498**:197–200. (3-19)

Perham, R.N.: **Self-assembly of biological macromolecules.** *Philos. Trans. R. Soc. Lond. B.* 1975, **272**:123–136. (1-19)

Perler, F.B.: **Protein splicing of inteins and hedgehog autoproteolysis: structure, function, and evolution.** *Cell* 1998, **92**:1–4. (3-17)

Perona, J.J. and Craik, C.S.: **Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold**. *J. Biol. Chem.* 1997, **272**:29987–29990. (4-8)

Perutz, M.F.: **Mechanisms of cooperativity and allosteric regulation in proteins.** *Q. Rev. Biophys.* 1989, **22**:139–237. (3-0)

Petrescu, A.J., Butters, T.D., Reinkensmeier, G., Petrescu, S., Platt, F.M., Dwek, R.A. and Wormald, M.R.: **The solution NMR structure of glucosylated *N*-glycans involved in the early stages of glycoprotein biosynthesis and folding**. *EMBO J.* 1997, **16**:4302–4310. (3-18)

Petsko, G.A.: **Structural basis of thermostability in hyperthermophilic proteins, or "there's more than one way to skin a cat".** *Methods Enzymol.* 2001, **334**:469–478. (3-11)

Petsko, G.A. and Ringe, D.: **Fluctuations in protein structure from X-ray diffraction.** *Annu. Rev. Biophys. Bioeng.* 1984, **13**:331–371. (1-22)

Petsko, G.A., Kenyon, G.L., Gerlt, J.A., Ringe, D. and Kozarich, J.W.: **On the origin of enzymatic species.** *Trends Biochem. Sci.* 1993, **18**:372–376. (4-11)

Phizicky, E.M. and Fields, S.: **Protein-protein interactions: methods for detection and analysis.** *Microbiol. Rev.* 1995, **59**:94–123. (4-4)

Ponting, C.P., Schultz, J., Copley, R.R., Andrade, M.A. and Bork, P.: **Evolution of domain families.** *Adv. Protein Chem.* 2000, **54**:185–244. (1-16)

Popot, J.L. and Engelman, D.M.: **Helical membrane protein folding, stability and evolution.** *Annu. Rev. Biochem.* 2000, **69**:881–922. (1-11)

Popot, J.L. and Engelman, D.M.: **Membrane protein folding and oligomerization: the two-stage model.** *Biochemistry* 1990, **29**:4031–4037. (1-11)

Poulos, T.L., Finzel, B.C. and Howard, A.J.: **High-resolution crystal structure of cytochrome p450 cam.** *J. Mol. Biol.* 1987, **195**:687–700. (2-3)

Quevillon-Cheruel, S., Collinet, B., Zhou, C.Z., Minard, P., Blondeau, K., Henkes, G., Aufrere, R., Coutant, J., Guittet, E., Lewit-Bentley, A., Leulliot, N., Ascone, I., Sorel, I., Savarin, P., De La Sierra Gallay, I.L., De La Torre, F., Poupon, A., Fourme, R., Janin, J. and Van Tilbeurgh, H.: **A structural genomics initiative on yeast proteins.** *J. Synchrotron. Radiat.* 2003, **10**:4–8. (4-0)

Radzicka, A. and Wolfenden, R.: **A proficient enzyme.** *Science* 1995, **267**:90–93. (2-8)

Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V.: **Stereochemistry of polypeptide chain configurations.** *J. Mol. Biol.* 1963, **7**:95–99. (1-5)

Ramakrishnan, V.: **Ribosome structure and the mechanism of translation.** *Cell* 2002, **108**:557–572. (3-9)

Rasmussen, B.F., Stock, A.M., Ringe, D. and Petsko, G.A.: **Crystalline ribonuclease A loses function below the dynamical transition at 220 K.** *Nature* 1992, **357**:423–424. (2-2)

Rattan, S.I., Derventzi, A. and Clark, B.F.: **Protein synthesis, posttranslational modification, and aging.** *Annls N.Y. Acad. Sci.* 1992, **663**:48–62. (1-13)

Reardon, D. and Farber, G.K.: **The structure and evolution of alpha/beta barrel proteins.** *FASEB J.* 1995, **9**:497–503. (1-18)

Reva, B., Finkelstein, A. and Topiol, S.: **Threading with chemostructural restrictions method for predicting fold and functionally significant residues: application to dipeptidylpeptidase IV (DPP-IV).** *Proteins* 2002, **47**:180–193. (4-10)

Reymond, M.A., Sanchez, J.C., Hughes, G.J., Gunther, K., Riese, J., Tortola, S., Peinado, M.A., Kirchner, T., Hohenberger, W., Hochstrasser, D.F. and Kockerling, F.: **Standardized characterization of gene expression in human colorectal epithelium by two-dimensional electrophoresis.** *Electrophoresis* 1997, **18**:2842–2848. (4-4)

Rhodes, G.: *Crystallography Made Crystal Clear: A Guide to Users of Macromolecular Models* 2nd ed. (Academic Press, New York and London, 1999). (5-2)

Richards, F.M. and Richmond, T.: **Solvents, interfaces and protein structure**. *Ciba. Found. Symp.* 1997, **60**:23–45. (1-10)

Richardson, J.S. and Richardson, D.C.: **Principles and patterns of protein conformation** in *Prediction of Protein Structure and the Principles of Protein Conformation* 2nd ed. Fasman, G.D. ed. (Plenum Press, New York, 1990). (1-5, 1-15, 1-17, 1-18)

Richardson, J.S.: **Introduction: protein motifs.** *FASEB J.*1994, **8**:1237–1239. (1-15)

Richardson, J.S.: **Looking at proteins: representation, folding, packing and design.** *Biophys. J.* 1992, **63**:1185–1209. (1-15)

Richardson, J.S.: **The anatomy and taxonomy of protein structure.** *Adv. Prot. Chem.* 1981, **34**:167–339. (1-7, 1-14)

Ringe, D.: **What makes a binding site a binding site?** *Curr. Opin. Struct. Biol.* 1995, **5**:825–829. (2-1, 2-3, 2-4)

Ringe, D. and Petsko, G.A.: **Mapping protein dynamics by X-ray diffraction.** *Prog. Biophys. Mol. Biol.* 1985, **45**:197–235. (1-22)

Rochet, J.C. and Lansbury, P.T. Jr.: **Amyloid fibrillogenesis: themes and variations.** *Curr. Opin. Struct. Biol.* 2000, **10**:60–68. (2-5)

Rockwell, N.C., Krysan, D.J., Komiyama, T. and Fuller, R.S.: **Precursor processing by kex2/furin proteases.** *Chem. Rev.* 2002, **102**:4525–4548. (3-16)

Rodbell, M.: **The complex regulation of receptor-coupled G-proteins.** *Adv. Enzyme Regul.* 1997, **37**:427–435. (3-8)

Rose, G.D. and Roy, S.: **Hydrophobic basis of packing in globular proteins.** *Proc. Natl Acad. Sci. USA* 1980, **77**:4643–4647. (1-10)

Ross, E.M. and Wilkie, T.M.: **GTPase-activating proteins for heterotrimeric G proteins: regulators of G protein signaling (RGS) and RGS-like proteins.** *Annu. Rev. Biochem.* 2000, **69**:795–827. (3-8)

Rudd, P.M. and Dwek, R.A.: **Glycosylation: heterogeneity and the 3D structure of proteins.** *Crit. Rev. Biochem. Mol. Biol.* 1997, **32**:1–100. (3-18)

Rutenber, E., Fauman, E.B., Keenan, R.J., Fong, S., Furth, P.S., Ortiz de Montellano, P.R., Meng, E., Kuntz, I.D., De Camp, D.L. and Salto, R.: **Structure of a non-peptide inhibitor complexed with HIV-1 protease. Developing a cycle of structure-based drug design.** *J. Biol. Chem.*1993, **268**:15343–15436. (2-2)

Sablin, E.P., and Fletterick, R.J.: **Nucleotide switches in molecular motors: structural analysis of kinesins and myosins.** *Curr. Opin. Struct. Biol.* 2001, **11**:716–724. (3-6, 3-10)

Salem, G.M., Hutchinson, E.G., Orengo, C.A. and Thornton, J.M.: **Correlation of observed fold frequency with the occurrence of local structural motifs.** *J. Mol. Biol.* 1999, **287**:969–981. (1-15)

Sato, T.K., Overduin, M. and Emr, S.D.: **Location, location, location: membrane targeting directed by PX domains.** *Science* 2001, **294**:1881–1885. (3-0, 3-2)

Scheffzek, K., Ahmadian, M.R. and Wittinghofer, A.: **GTPase-activating proteins: helping hands to complement an active site.** *Trends Biochem. Sci.* 1998, **23**:257–262. (3-7)

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 201

Schena, M., Shalon, D., Davis, R.W. and Brown, P.O.: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1995, **270**:467–470. (4-4)

Schmidt, A. and Lamzin, V.S.: ***Veni, vidi, vici*—atomic resolution unraveling the mysteries of protein function.** *Curr. Opin. Struct. Biol.* 2002, **12**:698–703. (5-2)

Scott, J.E.: **Molecules for strength and shape.** *Trends Biochem. Sci.* 1987, **12**:318–321. (1-6)

Sette, C., Inouye, C.J., Stroschein, S.L., Iaquinta, P.J. and Thorner, J.: **Mutational analysis suggests that activation of the yeast pheromone response mitogen-activated protein kinase pathway involves conformational changes in the Ste5 scaffold protein.** *Mol. Biol. Cell.* 2000, **11**:4033–4049. (2-5)

Sharp, K.A. and Englander, S.W.: **How much is a stabilizing bond worth?** *Trends Biochem. Sci.* 1994, **19**:526–529. (1-4, 1-12)

Shaw, J.P., Petsko, G.A. and Ringe, D.: **Determination of the structure of alanine racemase from *Bacillus stearothermophilus* at 1.9-Å resolution.** *Biochemistry* 1997, **36**:1329–1342. (4-17)

Sheinerman, F.B., Norel, R. and Honig, B.: **Electrostatic aspects of protein-protein interactions.** *Curr. Opin. Struct. Biol.* 2000, **10**:153–159. (4-10)

Sherlock, G., Hernandez-Boussard, T., Kasarskis, A., Binkley, G., Matese, J.C., Dwight, S.S., Kaloper, M., Weng, S., Jin, H., Ball, C.A., Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D. and Cherry, J.M.: **The Stanford Microarray Database.** *Nucleic Acids Res.* 2001, **29**:152–155. (4-4)

Sherman, M.Y. and Goldberg, A.L.: **Cellular defenses against unfolded proteins: a cell biologist thinks about neurodegenerative diseases.** *Neuron* 2001, **29**:15–32. (3-11)

Shuker, S.B., Hajduk, P.J., Meadows, R.P. and Fesik, S.W.: **Discovering high-affinity ligands for proteins: SAR by NMR.** *Science* 1996, **274**:1531–1534. (4-9)

Siezen, R.J. and Leunissen, J.A.: **Subtilases: the superfamily of subtilisin-like serine proteases.** *Protein Sci.* 1997, **6**:501–523. (4-8)

Sindelar, C.V., Budny, M.J., Rice, S., Naber, N., Fletterick, R. and Cooke, R.: **Two conformations in the human kinesin power stroke defined by X-ray crystallography and EPR spectroscopy.** *Nat. Struct. Biol.* 2002, **9**:844–848. (3-10)

Silverman, R.B.: *The Organic Chemistry of Enzyme-Catalyzed Reactions* Revised ed. (Academic Press, New York, 2002). (2-6, 2-10, 2-11, 2-14)

Simons, K.T., Kooperberg, C., Huang, E. and Baker, D.: **Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions.** *J. Mol. Biol.* 1997, **268**:209–225. (4-7)

Simons, K.T. Strauss, C. and Baker, D.: **Prospects for *ab initio* protein structural genomics.** *J. Mol. Biol.* 2001, **306**:1191–1199. (4-7)

Smith, C.A., Calabro, V. and Frankel, A.D.: **An RNA-binding chameleon.** *Mol. Cell* 2000, **6**:1067–1076. (4-14)

Smith, D.L., Almo, S.C., Toney, M.D. and Ringe, D.: **2.8-Å-resolution crystal structure of an active-site mutant of aspartate aminotransferase from *Escherichia coli*.** *Biochemistry* 1989, **28**:8161–8167. (4-12)

Snel, B., Bork, P. and Huynen, M.A.: **The identification of functional modules from the genomic association of genes.** *Proc. Natl Acad. Sci. USA* 2002, **99**:5890–5895. (4-2)

Song, H., Hanlon, N., Brown, N.R., Noble, M.E., Johnson, L.N. and Barford, D.: **Phosphoprotein-protein interactions revealed by the crystal structure of kinase-associated phosphatase in complex with phosphoCDK2.** *Mol. Cell* 2001, **7**:615–626. (3-14)

Spearman, J. C.: *The Hydrogen Bond and Other Intermolecular Forces* (The Chemical Society, London, 1975). (1-4)

Spiro, R.G.: **Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds.** *Glycobiology* 2002, **12**:43R–56R. (3-19)

Spoerner, M., Herrmann, C., Vetter, I.R., Kalbitzer, H.R. and Wittinghofer, A.: **Dynamic properties of the Ras switch I region and its importance for binding to effectors.** *Proc. Natl Acad. Sci. USA* 2001, **98**:4944–4949. (3-7)

Sprang, S.R.: **G protein mechanisms: insights from structural analysis.** *Annu. Rev. Biochem.* 1997, **66**:639–678. (3-7)

Spudich, J.A.: **The myosin swinging cross-bridge model.** *Nat. Rev. Mol. Cell Biol.* 2001, **2**:387–392. (3-6, 3-10)

Stamler, J.S., Lamas, S., and Fang, F.C.: **Nitrosylation: the prototypic redox-based signaling mechanism.** *Cell* 2001, **106**:675–683. (3-20)

Steiner, D.F.: **The proprotein convertases.** *Curr. Opin. Chem. Biol.* 1998, **2**:31–39. (3-16)

Steitz, T.A.: **DNA polymerases: structural diversity and common mechanisms.** *J. Biol. Chem.* 1999, **274**:17395–17398. (4-8)

Stenmark, H., Aasland, R. and Driscoll, P.C.: **The phosphatidylinositol 3-phosphate-binding FYVE finger.** *FEBS Lett.* 2002, **513**:77–84. (3-1)

Stock, A.M., Martinez-Hackert, E., Rasmussen, B.F., West, A.H., Stock, J.B., Ringe, D. and Petsko, G.A.: **Structure of the Mg^{2+}-bound form of CheY and mechanism of phosphoryl transfer in bacterial chemotaxis.** *Biochemistry* 1993, **32**:13375–13380. (3-15)

Stock, A.M., Robinson, V.L. and Goudreau, P.N.: **Two-component signal transduction.** *Annu. Rev. Biochem.* 2000, **69**:183–215. (3-15)

Stock, D., Nederlof, P.M., Seemuller, E., Baumeister, W., Huber, R., and Lowe, J.: **Proteasome: from structure to function.** *Curr. Opin. Biotechnol.* 1996, **7**:376–385. (3-11)

Strater, N., Schnappauf, G., Braus, G., and Lipscomb, W.N.: **Mechanisms of catalysis and allosteric regulation of yeast chorismate mutase from crystal structures.** *Structure* 1997, **5**:1437–1452. (2-8)

Sudarsanam, S.: **Structural diversity of sequentially identical subsequences of proteins: identical octapeptides can have different conformations.** *Proteins* 1998, **30**:228–231. (4-14)

Sugio, S., Petsko, G.A., Manning, J.M., Soda, K. and Ringe, D.: **Crystal structure of a D-amino acid aminotransferase: how the protein controls stereoselectivity.** *Biochemistry* 1995, **34**:9661–9669. (4-12)

Sun, Y.J., Chou, C.C., Chen, W.S., Wu, R.T., Meng, M. and Hsiao, C.D.: **The crystal structure of a multifunctional protein: phosphoglucose isomerase/autocrine motility factor/neuroleukin.** *Proc. Natl Acad. Sci. USA* 1999, **96**:5412–5417. (4-13)

Swindells, M.B., MacArthur, M.W. and Thornton, J.M.: **Intrinsic phi, psi propensities of amino acids, derived from the coil regions of known structures.** *Nat. Struct. Biol.* 1995, **2**:596–603. (1-8)

Swope, M.D. and Lolis, E.: **Macrophage migration inhibitory factor: cytokine, hormone, or enzyme?** *Rev. Physiol. Biochem. Pharmacol.* 1999, **139**:1–32. (4-13)

Szwajkajzer, D. and Carey, J.: **Molecular constraints on ligand-binding affinity and specificity.** *Biopolymers* 1997, **44**:181–198. (2-4)

Tainer, J.A., Roberts, V.A. and Getzoff, E.D.: **Protein metal-binding sites.** *Curr. Opin. Biotechnol.* 1992, **3**:378–387. (1-13)

Takai, Y., Sasaki, T. and Matozaki, T.: **Small GTP-binding proteins.** *Physiol. Rev.* 2001, **81**:153–208. (3-7)

Tan, S. and Richmond T.J.: **Crystal structure of the yeast MATalpha2/MCM1/DNA ternary complex.** *Nature* 1998, **391**:660–666. (4-14)

Tefferi, A., Bolander, M., Ansell, S., Wieben, E.D. and Spelsberg, T.: **Primer on medical genomics parts I–IV.** *Mayo. Clin. Proc.* 2002, **77**:927–940. (4-0)

Templin, M.F., Stoll, D., Schrenk, M., Traub, P.C., Vohringer, C.F. and Joos, T.O.: **Protein microarray technology.** *Trends Biotechnol.* 2002, **20**:160–166. (4-17)

Theodosiou, A. and Ashworth, A.: **MAP kinase phosphatases.** *Genome Biol.* 2002, **3**:reviews3009. (3-13)

Thoden, J.B., Holden, H.M., Wesenberg, G., Raushel, F.M. and Rayment, I.: **Structural and carbamoyl phosphate synthetase: a journey of 96 Å from substrate to product.** *Biochemistry* 1997, **36**:6305–6316. (2-16)

Thornton, J.W. and DeSalle, R.: **Gene family evolution and homology: genomics meets phylogenetics.** *Annu. Rev. Genomics Hum. Genet.* 2000, **1**:41–73. (1-14)

Thornton, J.M., Orengo, C.A., Todd, A.E. and Pearl, F.M.: **Protein folds, functions and evolution.** *J. Mol. Biol.* 1999, **293**:333–342. (1-15)

Todd, A.E., Orengo, C.A., Thornton, J.M.: **Evolution of function in protein superfamilies, from a structural perspective.** *J. Mol. Biol.* 2001, **307**:1113–1143. (4-8)

Tong, A.H., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Page, N., Robinson, M., Raghibizadeh, S., Hogue, C.W., Bussey, H., Andrews, B., Tyers, M. and Boone, C.: **Systematic genetic analysis with ordered arrays of yeast deletion mutants.** *Science* 2001, **294**:2364–2368. (4-0)

Topham, C.M., McLeod, A., Eisenmenger, F., Overington, J.P., Johnson, M.S. and Blundell, T.L.: **Fragment ranking in modelling of protein structure. Conformationally constrained environmental amino acid substitution tables.** *J. Mol. Biol.* 1993, **229**:194–220. (1-2)

Tsou, C.L.: **Active site flexibility in enzyme catalysis.** *Annls. N.Y. Acad. Sci.* 1998, **864**:1–8. (2-2)

Tucker, C.L., Gera, J.F. and Uetz, P.: **Towards an understanding of complex protein networks.** *Trends Cell Biol.* 2001, **11**:102–106. (4-17)

Turner, M.: **Cellular memory and the histone code.** *Cell* 2002, **111**:285–291. (3-20)

Vale, R.D. and Milligan, R.A.: **The way things move: looking under the hood of molecular motor proteins.** *Science* 2000, **288**:88–95. (3-10)

van Roessel, P. and Brand, A.H.: **Imaging into the future: visualizing gene expression and protein interactions with fluorescent proteins.** *Nat. Cell Biol.* 2002, **4**:E15–E20. (4-17)

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 202

Vetter, I.R. and Wittinghofer, A.: **Nucleoside triphosphate-binding proteins: different scaffolds to achieve phosphoryl transfer.** *Q. Rev. Biophys.* 1999, **32**:1–56. (3-7)

Vetter, I.R. and Wittinghofer, A.: **The guanine nucleotide-binding switch in three dimensions.** *Science* 2001, **294**:1299–1304. (3-6)

Vocadlo, D.J., Davies, G.J., Laine, R. and Withers, S.G.: **Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate.** *Nature* 2001, **412**:835–838. (2-12)

Voet, D. and Voet, G.: *Biochemistry* 2nd ed. (Wiley, New York, 1995). (1-0, 1-3, 2-12, 2-13)

Vogeley, L., Palm, G.J., Mesters, J.R. and Hilgenfeld, R.: **Conformational change of elongation factor Tu (EF-Tu) induced by antibiotic binding. Crystal structure of the complex between EF-Tu.GDP and aurodox.** *J. Biol. Chem.* 2001, **276**:17149–17155. (3-9)

von Heijne, G.: **Recent advances in the understanding of membrane protein assembly and structure.** *Q. Rev. Biophys.* 1999, **32**:285–307. (1-11)

von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P. and Snel, B.: **STRING: a database of predicted functional associations between proteins.** *Nucleic Acids Res.* 2003, **31**:258–261. (4-2)

von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S. and Bork, P.: **Comparative assessment of large-scale data sets of protein–protein interactions.** *Nature* 2002, **417**:399–403. (4-0, 4-17)

Wade, R.C., Gabdoulline, R.R., Ludemann, S.K. and Lounnas, V.: **Electrostatic steering and ionic tethering in enzyme-ligand binding: insights from simulations.** *Proc. Natl Acad. Sci. USA* 1998, **95**:5942–5949. (2-7)

Wall, M.E., Gallagher, S.C. and Trewhella, J.: **Large-scale shape changes in proteins and macromolecular complexes.** *Annu. Rev. Phys. Chem.* 2000, **51**:355–380. (1-22)

Wallace, A.C., Laskowski, R.A. and Thornton, J.M.: **Derivation of 3D coordinate templates for searching structural databases: application to Ser-His-Asp catalytic triads in the serine proteinases and lipases.** *Protein Sci.* 1996, **5**:1001–1013. (4-10)

Walsh, C.: *Enzymatic Reaction Mechanisms* (Freeman, San Francisco, 1979). (2-6, 2-14)

Walther, D., Eisenhaber, F. and Argos, P.: **Principles of helix-helix packing in proteins: the helical lattice superposition model.** *J. Mol. Biol.* 1996, **255**:536–553. (1-10)

Watanabe, H., Takehana, K., Date, M., Shinozaki, T. and Raz, A.: **Tumor cell autocrine motility factor is the neuroleukin/phosphohexose isomerase polypeptide.** *Cancer Res.* 1996, **56**:2960–2963. (4-13)

Weber, P.C. and Salemme, F.R.: **Structural and functional diversity in 4 alpha-helical proteins.** *Nature* 1980, **287**:82–84. (1-17)

Weiss, M.S., Blanke, S.R., Collier, R.J. and Eisenberg, D.: **Structure of the isolated catalytic domain of diphtheria toxin.** *Biochemistry* 1995, **34**:773–781. (3-3)

Wells, L., Vosseller, K. and Hart, G.W.: **Glycosylation of nucleoplasmic proteins: signal transduction and *O*-GlcNac.** *Science* 2001, **291**:2376–2378. (3-18)

West, A.H. and Stock, A.M.: **Histidine kinases and response regulator proteins in two-component signaling systems.** *Trends Biochem. Sci.* 2001, **26**:369–376. (3-15)

White, A., Ding, X., van der Spek, J.C., Murphy, J.R. and Ringe, D.: **Structure of the metal-ion-activated diphtheria toxin repressor/tox operator complex.** *Nature* 1998, **394**:502–506. (3-5)

Wilkinson, M.G. and Millar, J.B.: **Control of the eukaryotic cell cycle by MAP kinase signaling pathways.** *FASEB J.* 2000, **14**:2147–2157. (3-12)

Williams, R.W., Chang, A., Juretic, D. and Loughran, S.: **Secondary structure predictions and medium range interactions.** *Biochim. Biophys. Acta* 1987, **916**:200–204. (1-8)

Wilson, C., Kreychman, J. and Gerstein, M.: **Assessing annotation transfer for genomics: quantifying the relations between protein sequence, structure and function through tradition and probabilistic scores.** *J. Mol. Biol.* 2000, **297**:233–249. (4-1)

Wittinghofer, A.: **Signal transduction via Ras.** *Biol. Chem.* 1998, **379**:933–937. (3-7)

Workman, J.L. and Kingston, R.E.: **Alteration of nucleosome structure as a mechanism of transcriptional regulation.** *Annu. Rev. Biochem.* 1998, **67**:545–579. (3-20)

Wolan, D.W., Greasley, S.E., Beardsley, G.P. and Wilson, I.A.: **Structural insights into the avian AICAR transformylase mexhanism.** *Biochemistry* 2002, **41**:15505–15513. (2-15)

Wrba, A., Schweiger, A., Schultes, V., Jaenicke R. and Zavodsky, P.: **Extremely thermostable D-glyceraldehyde-3-phosphate dehydrogenase from the eubacterium *Thermotoga maritima*.** *Biochemistry* 1990, **29**:7584–7592. (2-2)

Wu, N., Mo, Y., Gao, J. and Pai, E.F.: **Electrostatic stress in catalysis: structure and mechanism of the enzyme orotidine monophosphate decarboxylase.** *Proc. Natl Acad. Sci. USA* 2000, **97**:2017–2022. (2-8)

Wurgler-Murphy, S.M. and Saito, H.: **Two-component signal transducers and MAPK cascades.** *Trends Biochem. Sci.* 1997, **22**:172–176. (3-15)

Wyrick, J.J. and Young, R.A.: **Deciphering gene expression regulatory networks.** *Curr. Opin. Genet. Dev.* 2002, **12**:130–136. (3-0)

Xia, D., Yu, C.A., Kim, H., Xia, J.Z., Kachurin, A.M., Zhang, L., Yu, L. and Deisenhofer, J.: **Crystal structure of the cytochrome bc1 complex from bovine heart mitochondria.** *Science* 1997, **277**:60–66. (1-11)

Xu, D., Xu, Y. and Uberbacher, E.C.: **Computational tools for protein modeling.** *Curr. Protein Pept. Sci.* 2000, **1**:1–21. (5-3)

Xu, W., Doshi, A., Lei, M., Eck, M.J. and Harrison, S.C.: **Crystal structures of c-Src reveal features of its autoinhibitory mechanism.** *Mol. Cell* 1999, **3**:629–638. (3-13)

Xu, W., Seiter, K., Feldman, E., Ahmed, T. and Chiao, J.W.: **The differentiation and maturation mediator for human myeloid leukemia cells shares homology with neuroleukin or phosphoglucose isomerase.** *Blood* 1996, **87**:4502–4506. (4-13)

Yaffe, M.B.: **How do 14-3-3 proteins work? Gatekeeper phosphorylation and the molecular anvil hypotheses.** *FEBS Lett.* 2002, **513**:53–57. (3-1)

Yaffe, M.B.: **Phosphotyrosine-binding domains in signal transduction.** *Nat. Rev. Mol. Cell Biol.* 2002, **3**:177–186. (3-1)

Yaffe, M.B. and Elia, A.E.: **Phosphoserine/threonine binding domains.** *Curr. Opin. Cell Biol.* 2001, **13**:131–138. (3-1)

Yang, A.S. and Honig, B.: **An integrated approach to the analysis and modeling of protein sequences and structures. I-III.** *J Mol Biol.* 2000, **301**:665–678. (4-6)

Yang, W.Z., Ko, T.P., Corselli, L., Johnson, R.C. and Yuan, H.S.: **Conversion of a beta-strand to an alpha-helix induced by a single-site mutation observed in the crystal structure of Fis mutant Pro26Ala.** *Protein Sci.* 1998, **7**:1875–1883. (4-14)

Yaswen, L., Diehl, N., Brennan, M.B. and Hochgeschwender, U.: **Obesity in the mouse model of pro-opiomelanocortin deficiency responds to peripheral melanocortin.** *Nat. Med.* 1999, **5**:1066–1070. (4-4)

Ye, S. and Goldsmith, E.J.: **Serpins and other covalent protease inhibitors.** *Curr. Opin. Struct. Biol.* 2001, **11**:740–745. (4-15)

Yennawar, N., Dunbar, J., Conway, M., Huston, S. and Farber, G.: **The structure of human mitochondrial branched-chain aminotransferase.** *Acta. Crystallogr. D. Biol. Crystallogr.* 2001, **57**:506–515. (4-12)

Yoshimura, T., Jhee, K.H. and Soda, K.: **Stereospecificity for the hydrogen transfer and molecular evolution of pyridoxal enzymes.** *Biosci. Biotechnol. Biochem.* 1996, **60**:181–187. (4-12)

Young, M.A., Gonfloni, S., Superti-Furga, G., Roux, B. and Kuriyan, J.: **Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation.** *Cell* 2001, **105**:115–126. (3-13)

Zahn, R., Liu, A., Lührs, T., Riek, R., von Schroetter, C., Lopez Garcia F., Billeter, M., Calzolai, L., Wider, G. and Wüthrich, K.: **NMR structure of the human prion protein.** *Proc. Natl Acad. Sci. USA* 2000, **97**:145–150. (4-15)

Zhang, B., Rychlewski L, Pawlowski K, Fetrow JS, Skolnick, J. and Godzik, A.: **From fold predictions to function predictions: automation of functional site conservation analysis for functional genome predictions.** *Protein Sci.* 1999, **8**:1104–1115. (4-10)

Zhang, T., Bertelsen, E. and Alber, T.: **Entropic effects of disulfide bonds on protein stability**. *Nat. Struct. Biol.* 1994, **1**:434–438. (1-13)

Zhu, Z.Y. and Blundell, T.L.: **The use of amino acid patterns of classified helices and strands in secondary structure prediction.** *J. Mol. Biol.* 1996, **260**:261–276. (1-8)

Zielenkiewicz, P. and Rabczenko, A.: **Methods of molecular modelling of protein-protein interactions.** *Biophys. Chem.* 1988, **29**:219–214. (1-19)

Zolnierowicz, S. and Bollen, M.: **Protein phosphorylation and protein phosphatases. De Panne, Belgium, September 19–24, 1999.** *EMBO J.* 2000, **19**:483–488. (3-12)

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 203

Index

References to figures are in bold, for example **F4-33** refers to Figure 4-33.

3-10 helix **F1-14**
14-3-3 domain 88, **F3-2**

acetals, cleavage of 72–73
acetonitrile **F4-32, F4-33**
N-acetylation 28, 29, 126–127, **F3-48**
acetylcholine receptor 40
acetylcholinesterase 92
N-acetylgalactosamine 122
N-acetylglucosamine 122, **F3-41,** 123
N-acetylglucosaminyltransferase I 123
acetyltransferases **F3-48,** 127
acid-base catalysis 52–53, 74–75, **F2-35**
acids 74
aconitase **F4-47**
F-actin **F1-1**
actin filaments 60, 61
  formation 47
  myosin movement along 106–107, **F3-17**
activation energy 62, 63
activation-energy barrier 62, 63
active sites 52
  alanine racemase **F4-60**
  amino acid aminotransferase **F4-45**
  aspartate aminotransferase **F2-24, F4-45**
  aspartate transcarbamoylase **F2-25**
  bifunctional enzymes 80–81, **F2-42, F2-43**
  chemistry 74–75
  chymotrypsin **F2-40**
  citrate synthase **F2-28**
  enolase **F4-55**
  exclusion of water 68–69, **F2-30**
  galactonate dehydratase **F4-58**
  geometry 64–65, 150–151, **F4-35**
  ground-state destabilization by 67
  homologous, in prediction of function 162
  identifying residues involved in 150–151, **F4-35**
  location 57
  lysozyme **F2-36**
  mandelate racemase **F2-3, F4-41, F4-55**
  muconate lactonizing enzyme **F4-42, F4-55**
  beta-phosphoglucomutase **F2-41**
  phosphoglycerate kinase **F2-29**
  plasmin **F3-32**
  plasminogen **F3-32**
  promoting proximity 66–67, **F2-25**
  reaction sub-site 64, 65
  specialized microenvironments 52–53, **F2-3**
  specificity sub-site 64, 65
  strategies for identifying 148–149
  transition state stabilization by 68, **F2-27, F2-28**
  tunnel linking two 82, **F2-44**
  YBL036c **F4-60**
S-acylation 124, 125
acylglycerols 73
addition reactions 72, **F2-32**
adenosine deaminase **F2-20**
adenosine triphosphate *see* ATP
adenylate kinase **F2-7**
ADP ribosylation factor (ARF) 125, **F3-46**
adrenocorticotropic hormone (ACTH, corticotropin) 118, **F3-33**
affinity 58, 59
aggregates, protein 160–161
AICAR transformylase-IMP cyclohydrolase (ATIC) 81, **F2-43**
alanine 4, **F1-3**
alanine racemase 30, **F1-42, F1-53**
  compared with YBL036c 164, **F4-59, F4-60**
alanine-scanning mutagenesis 150
aldolase 72
aldol cleavage 72
aldol condensation 72, **F2-31**
aldose reductase **F1-47**
alignment, sequence *see* sequence(s), alignment
aliphatic carbon groups (–C–H) 75
alkaline phosphatase **F2-20,** 79
allosteric activator 96
allosteric effectors 56, 86–87, 96
allosteric inhibitor 96
allosteric proteins 86, 96
allostery 96–97, **F3-9**
alpha$_1$-antitrypsin, Z-variant 161
alpha/beta barrel 38–39, **F1-61**
alpha+beta domain 36, 39, **F1-62**
alpha/beta domain 36, 38–39, **F1-61**
alpha/beta twist 38–39, **F1-61**
alpha domain 36
alpha helix 12, 14–15, **F1-13**
  amino-acid preferences 18, **F1-20**
  amphipathic 14, 15, **F1-15**
  chameleon sequences 158, **F4-49,** 159
  coiled-coil interactions 41, **F1-67**
  conformational parameters **F1-14**
  connected to beta sheets 16–17, **F1-18**
  isolated 18, 24
  in membrane proteins *see* membrane proteins, alpha-helical
  membrane spanning 25, **F1-31**
  packing arrangements 23, **F1-28**
alternative splicing 6
Alzheimer's disease 60, 160
amide bonds 8
  cleavage 72–73, **F2-33**
D-amino acid aminotransferase (DaAT) **F1-74**
  cofactor binding 28, **F1-39**
  compared with L-amino acid aminotransferase **F4-46**
  compared with L-aspartate aminotransferase **F1-48,** 154–155, **F4-44, F4-45**
amino acids 4
  branched chain, in beta sheets 17, 18
  conformational preferences 18, **F1-20**
  conservative substitutions 6–7, **F1-6**
  genetic code 6, **F1-4**
  physical-chemical properties 4, **F1-3**
  sequence *see* sequence(s)
  see also specific amino acids
aminoacyl-tRNAs 104, **F3-16**
5-aminoimidazole-4-carboxamide-ribonucleotide (AICAR) transformylase 81, **F2-43**
aminopeptidases 155
amino terminus **F1-7**
aminotransferases 28, 47, 140, 154–155
amphipathic molecules 4
  alpha helices 14, 15, **F1-15**
  amino acids 4, **F1-3**
  beta sheets 17
amplification, signal 87, 110
amyloid fibrils 60, 158
  formation 160–161, **F4-53**
amyloidoses 160–161
anisotropic 58, 59
ANK (ankyrin repeat) domain **F3-2**
anthrax toxin lethal factor **F2-2**
antibiotics 104–105, 122
Arabidopsis thaliana 136, **F4-10**
arginine
  methylation 126, **F3-37**
  structure and side chain 4, **F1-3**
armadillo repeat (ARM) domain **F3-2**
asparagine
  deamidation 8
  structure and side chain 4, **F1-3**
aspartate aminotransferase 53
  active site **F2-24, F4-45**
  ligand-induced conformational change 47, **F1-80**
  compared with D-amino acid aminotransferase **F1-48,** 154–155, **F4-44, F4-45**
aspartate semi-aldehyde dehydrogenase 38, **F1-61**
aspartate transcarbamoylase (ATCase) 97
  allosteric regulation 96, **F3-10**
  proximity of reactants 66–67, **F2-25**
aspartic acid (aspartate)
  phosphorylation 110–111, 117
  structure and side chain 4, **F1-3**
assemblies, protein
  structural proteins 60–61
  see also quaternary structure
asymmetric protein complexes 44, **F1-73**
atomic coordinates 168
ATP
  in allosteric regulation 96, **F3-10**
  formation 73
  hydrolysis 73, 98
ATPases 98
  in motor switches **F3-17,** 107
  structure **F3-12,** 99
  switch mechanism 99
ATP-citrate lyase 95
aurodox 104–105
autocrine motility factor (AMF) 156–157
autophosphorylation 112
autoproteolysis by inteins 120–121
autoregulation 88

backbone
  amino acid **F1-3**
  polypeptide 3, **F1-2,** 8
bacteria *see* prokaryotes
bacteriochlorophyll A protein 37, **F1-59**
bacteriocin AS-48 **F4-27**
ball-and-stick model 172, **F5-5**
barnase **F1-22**
bases 6, 74–75
benzoylformate decarboxylase (BFD) 141, **F4-20**
bepridil 95
beta-amyloid protein 60
beta barrel 16, **F1-19,** 36, 37
  in membrane proteins 25, **F1-32**
beta domain 36–37
beta-propeller domain 36, **F1-57**
beta sandwich 36, 37
beta sheet (pleated sheet) 12, 16–17
  amino-acid preferences 18, **F1-20**
  amphipathic 17
  antiparallel 16, **F1-17,** 37
  chameleon sequences 158, **F4-49,** 159
  crossover connection between 38, **F1-60**
  in membrane proteins 24, 25, **F1-32**
  mixed 16, **F1-17**
  in oligomerization 42, **F1-70**
  parallel 16, **F1-17**
  transmembrane 18
beta turn 12, 13, **F1-12**
  amino-acid preferences 18, **F1-20**
  connecting beta sheets 16–17
BH domain **F3-2**
binding **F1-1,** 3, 50, 51, 52–59
  competitive 94
  cooperative 94–95, **F3-8**
  displacement of water and 59
  docking models 151
  gated 64, **F2-23**
  interactions, role of complementarity 40–41, **F1-66,** 54, **F2-4**
  protein flexibility and 54–55, **F2-6**
binding sites, ligand 52, 58–59
  identifying residues 150–151
  location and shape 56–57
  for macromolecules 56, **F2-8, F2-9**

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 204

for small ligands 56–57
specialized microenvironments 52–53
strategies for identifying 148–149
biotin **F2-37**
bisubstrate analog 66–67
BLAST 132, 134, 135, 136
blood coagulation 119, **F3-34**
bovine pancreatic trypsin inhibitor (BPTI) **F1-37**
bovine spongiform encephalopathy (BSE) 160
bromo domain **F3-2**

C1 domain **F3-2**
C2 domain **F3-2**
Caenorhabditis elegans 136, **F4-10**
calcium (Ca^{2+}) 28
calsequestrin motif **F4-7**
camphor **F2-10**
carbamoyl phosphate synthetase 83, **F2-45**
carbon–carbon bonds, formation and cleavage 72, **F2-31**
carbon dioxide, loss of 73
carbonic anhydrase **F2-20**
carboxyl terminus **F1-7**
carboxypeptidase 79
cascades, regulatory 87, 110, **F3-21**
CASP (critical assessment of techniques for protein structure prediction) 144
catalases 135
catalysis **F1-1,** 3, 50, 51, 62–83
acid-base 52–53, 74–75, **F2-35**
energetics 63, **F2-21**
factors contributing to 62–63
identifying residues involved 150–151
multi-step reactions 78–79, **F2-40**
types of reactions 70–73
see also enzymes
catalyst 62
catalytic triad 34–35, **F1-52,** 78, 79
see also under serine proteases
cathepsin D **F3-5,** 93
cavities, binding 56–57, 148
Cdc42 59
Cdk2 96, 114–115, **F3-27**
activation loop (T loop) 114, **F3-27,** 115
substrate-binding site 115, **F3-28**
Cdk-activating kinase (CAK) **F3-27, F3-28**
cells
internal frameworks, connectors and scaffolds 60, 61, **F2-18**
internal structure **F3-3**
protein expression 138, **F4-15,** 165
protein localization *see* localization, intracellular
reducing environment 10, 92
cellular compartments
as mechanism of regulating function 90
pH and redox environments 92–93
protein targeting 86, 91, **F3-4**
chameleon sequences 158–159, **F4-49**
channels 25, **F1-33**
chaperones 20
chelates 28
CheY **F3-30**
cholera toxin **F1-74**
chorismate mutase 67, **F2-26**
choroideremia 125
Chou-Fasman method of predicting secondary structure 18
chromatin 126, 127
chromo domain **F3-2**
chymotrypsin **F2-20,** 118, **F3-31, F4-29**
alpha 118, **F3-31**
catalytic triad **F1-52**
reaction steps 78–79, **F2-40**
chymotrypsinogen 143, **F4-24**
activation 118, **F3-31**
circular dichroism spectra **F1-35**
citrate synthase 68, **F2-28,** 72, **F2-32**
classification, protein 32–33
clefts, binding 56–57
co-activators 96, 97
coagulation, blood 119, **F3-34**
coat proteins, vesicular 125, **F3-46**
codons 6, **F1-4**
stop 6, 7
coenzyme A (CoA) 72, **F2-37**
coenzymes 76–77, **F2-37**
cofactors
enzyme 76–77, **F2-37**
decarboxylation reactions 73, **F2-34**
redox reactions 70–71, **F2-31**
stabilizing tertiary structure 28, **F1-39**
coiled coils 40–41, 41, **F1-67**
collagen 15, **F1-16,** 60, **F2-17**
collective motions 46, **F1-76**
competitive binding 94
competitive inhibitors 86, 94, **F3-7**
complementarity
in binding interactions 40–41, **F1-66,** 52, **F2-2**
to transition states 68
complement cascade 110
complexes, protein *see* quaternary structure
computational methods
identifying catalytic residues 150–151
mapping binding sites 148–149
concavity, binding sites 56–57, 58
condensation 73
conformational changes
in Cdk activation 114–115, **F3-28**
exclusion of water 68–69, **F2-30**
ligand-induced 46, **F1-76**
induced fit model 54, **F2-4**
large-scale 47, **F1-80,** 55, **F2-7**
local, in domain structure 46, **F1-77**
regulating protein function 96–97
molecular switches 98
motor proteins **F3-17**
pH-dependent **F3-5,** 93
phosphorylation-induced 111, **F3-22**
prion proteins 160
switch proteins 98, 99, **F3-13**
transition state stabilization 68, **F2-29**
conformational states 47, **F1-79**
conservation, evolutionary 142–143, **F4-23**
conservative substitutions 6–7, **F1-6**
conserved (sequences) 132, 133
cooperative binding 94–95, **F3-8**
cooperativity 86, 94–95
negative 86, 94, 95
positive 86, 94, 95
coordinate covalent bonds 28, **F1-38**
copper (Cu^{2+}) 77, **F2-38**
co-repressors 96, 97
covalent bonds 10, **F1-10**
atomic motions 46–47
coordinate 28, **F1-38**
stabilizing tertiary structure 28
see also disulfide bridges; peptide bonds
covalent modification 86–87, 110
Creutzfeldt-Jakob disease (CJD) 160
crixivan **F2-5**
cross-linking interactions
in domains 36, 39
in quaternary structure 42
structural assemblies of proteins 60–61
see also disulfide bridges; metal ions
crossover connection, between beta sheets 38, **F1-60**
crystallins 31, 147
beta 31
gamma 31, **F1-44**
C terminus *see* carboxyl terminus
cyclins 96, 114–115
cyclin-dependent kinases (CDKs) 112, 114–115
see also Cdk2
cyclodextrin glycosyltransferase **F4-49**
cyclophilin **F2-20, F4-47**
cysteine
disulfide bonds *see* disulfide bridges
nitrosylation 127, **F3-50**
structure and side chain 4, **F1-3**
cytidine triphosphate (CTP) 96, 97
cytochrome b562 **F1-28, F1-51,** 36
cytochrome c 28, **F1-39**
cytochrome c6 **F2-12**
cytochrome P450 **F2-10,** 71
cytochromes 36
cytokines, multifunctional 156–157

DctB protein, *Rhizobium meliloti* 25, **F1-31**
deacetylases **F3-48,** 127
DEAD box helicase motif **F4-7**
death domain (DD) **F3-2**
decarboxylation 72, 73, **F2-34**
decoy structures, Rosetta method 145, **F4-26**
degradation, protein 78–79, 87, 108–109
see also peptide bonds, hydrolysis; proteolysis
dehydration 73
3-dehydroquinate dehydratase 44, **F1-74**
deletions, sequence 33, 132
deletion strains, yeast 165
denaturants 26, 27, 55
denatured state **F1-22,** 26, 27, **F1-35**
5′-deoxyadenosylcobalamin **F2-37**
1-deoxymannojirimycin 122
dephosphorylation 115
DH domain **F4-6**
diabetes, type II 160
Dictyostelium 116
differentiation and maturation mediator (DMM) 156–157
dihydrofolate reductase (DHFR) **F1-24,** 81, **F4-31**
dimers 40, 44, **F1-74**
diphtheria toxin
gene repressor protein **F2-9,** 97, **F3-11**
mechanism of action 93, **F3-6**
dipole moment 8, 9, **F1-8**
disulfide bridges 10, **F1-10,** 28, **F1-37**
cross-linking of domains 39, **F1-63**
intracellular environment and 92
disulfide oxidoreductases 150
DNA 6
binding sites 56, **F2-9**
DNA-binding proteins 88, 97
DNA microarrays 138, **F4-14,** 165
DNA polymerases **F1-1, F2-33,** 147, **F4-30**
DNase **F2-33**
DOCK 151
docking programs 151
dodecamer 44, **F1-74**
dolichol pyrophosphate (DPP) **F3-42**
domains 30–31
arrangements 32–33, **F1-46**
binding sites at interface between 57, **F2-11**
classes 36–39
conformational changes 46, 47
evolution by gene duplication and fusion 31
folds 32
hydrophobic cores 30
interaction 86, 88, **F3-2**
interrupted 30, **F1-42**
prediction of protein function from 134, **F4-6**
in protein classification 32
proteins with multiple 31, **F1-43, F1-44**
size 30
structural 56, 57, 142
domain swapping 58, 59, **F2-14**

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 205

dominant-negative effect 42, 43, **F1-72**
Drosophila melanogaster 136, **F4-10**
Hedgehog protein (HH-C) **F3-39**
drugs, disrupting protein function 97
DtxR 97, **F3-11**

EF-1 104, **F4-47**
effector molecules (ligands) 86–87, 94–97
competitive binding and cooperativity 94–95
ligand-induced conformational change and allostery 96–97
EF-G 105
EF-hand domain 88, **F3-1, F4-6**
EF-Ts 104, **F3-16**
EF-Tu 104–105, **F3-16,** 158, 161
Eg5 **F3-12**
Ehlers-Danlos syndrome 61
elastin 60–61
electron density map 168, **F5-1**
electrons
transfer 70–71, **F2-31**
unpaired 76
electrostatic interactions 10, **F1-10**
in binding interactions 52–53
in catalysis 64–65, **F2-23**
see also hydrogen bonds; van der Waals interactions
elimination reactions 72, **F2-32**
elongation factor *see* EF
endonucleases, homing 120–121
endoplasmic reticulum (ER)
glycosylation in 122, **F3-42**
protein targeting to 91, **F3-4**
endorphin 118, **F3-33**
endosomal vesicles 92–93
energetics, catalysis 63, **F2-21**
enkephalin 118, **F3-33**
enolase 141, 162, **F4-55**
Enterococcus faecalis protein, prediction of secondary structure **F1-21**
enthalpy 26
entropy 26
enzymes 62
activation by limited proteolysis 118, **F3-31, F3-32**
active sites *see* active sites
allosteric regulation 96, **F3-10**
bifunctional 80–82, **F2-42, F2-43, F2-44**
cofactors *see* cofactors, enzyme
competitive inhibition 94, **F3-7**
divergent and convergent evolution 140–141, 154–155
multifunctional 80–83
non-enzymatic functions 83
phosphorylation 111, **F3-23; F3-22**
predicting function of uncharacterized 162–163
reaction acceleration rates 62, **F2-20**
with structural role 61
substrates *see* substrates
temperature dependence 55, **F2-6**
trifunctional 83, **F2-45**
types of reactions catalyzed 70–73
see also catalysis; specific enzymes
equilibrium 62
equilibrium constant 62
Erk2 **F4-2**
ERKs **F4-5**
Escherichia coli **F4-10**
f857 protein 162–163
galactonate utilization pathway **F4-56**
protein kinases 111
two-component signaling systems 116
esters, cleavage of 72–73
eukaryotes
cell structure 90, **F3-3**
genes 6, **F1-5**
glycosylation 122
protein degradation 108–109, **F3-19**
two-component signaling systems 116
E-values 132, **F4-3,** 133
EVH1 domain 88
evolution 140–141
convergent 34, 35, 140–141, 146–147, 154–155
divergent 140, **F4-19,** 141, 154, 155
by gene fusions and duplications 31
homology modeling and 142–143, **F4-23**
by sequence insertions and deletions 33
evolutionary distance 132, 133
exons 6, **F1-5**
insertion, in protein evolution 33
expression, gene and protein 138, 165
exteins 120

f857 162–163
F-actin **F1-1**
families
interaction domains 88, **F3-2**
protein 146
functional variation within 147
profiles 135, **F4-8**
farnesylation 91, 124
FASTA 132
F-box domain **F3-2**
feedback inhibition, competitive 94, **F3-7**
ferredoxins 39, 147
FHA domain 88, **F3-2**
FhuA protein **F1-32**
fibrinogen 60
fibronectin domain **F3-2**
fibrous proteins 30
Fis 158, **F4-50**
FixJ **F3-30**
flavin adenine dinucleotide (FAD) 70, **F2-31, F2-37**
flavin mononucleotide (FMN) 70
flexibility, protein 27, **F1-36,** 46–47
binding in interior cavities and 57, **F2-10**
binding interactions and 54–55, **F2-6**
cooperativity and 95
variability 55, **F2-7**
flipping, aromatic ring 46–47
fluctuations 46, **F1-76**
folded proteins
secondary structure *see* secondary structure
thermodynamics 26–27
weak interactions stabilizing 10–11, **F1-10,** 26–27
folding 3, 20–21, **F1-23**
forces driving 20–21
helical membrane proteins 21
intermediates 20, **F1-22**
see also tertiary structure
folds, domain 32
four-helix bundle **F1-51,** 36, **F1-54,** 153
frameworks, intracellular 60
free energy
ligand binding 59
stabilization 26
free radicals 76
FtsH **F4-47**
fumarase **F2-20,** 72, **F2-32**
function, protein **F1-1,** 3, 50–51
context dependence 90
convergent evolution 146–147, 154–155
diverse structures with one 154–155
diversity of specific structures 152–153
flexibility and 54–55
gene product with unknown 164–165
levels 50, **F2-1**
proteins with more than one 156–157, **F4-47**
regulation *see* regulation of protein function
and structure 33
time and distance scales **F4-1**
of uncharacterized genes 162–163
functional genomics 130
experimental methods **F4-1,** 138–139
sequence analysis methods 132–137
fungi, two-component signaling systems 116
"fuzzy functional form" (FFF) 150
FYVE domain 88, **F3-2**

Gal4 transcription factor **F2-9**
galactonate dehydratase (GalD) **F1-45**
active site **F4-58**
characterization 162–163, **F4-57**
quaternary structure **F1-74**
galactonate utilization pathway **F4-56**
beta-galactosidase **F4-49**
GAP1 88, **F3-14**
GAPs *see* GTPase-activating proteins
gated binding 64, **F2-23**
G domain 98–99, **F3-12**
GEFs *see* guanine-nucleotide exchange factors
gene(s) 6–7
duplication 31
expression patterns 138, 165
fusion 31
numbers per organism 136, **F4-10**
product of unknown function 164–165
proteins regulating 88, 97
relatedness 132
uncharacterized, functions for 162–163
gene chips 138
gene knockouts 138, 139, **F4-15,** 157
genetic code 6, **F1-4**
degeneracy 6
organization 6–7, **F1-6**
genetic diseases 42–43
genetic engineering 120
genome
sequencing 136, **F4-9**
sizes 136, **F4-10**
genomics 130
functional *see* functional genomics
structural 146
geranylgeranyl modification 124
Glc3Man9GlcNAc2 122, 123, **F3-43**
globin fold domain 36, **F1-55**
globular proteins 30
glucose-6-phosphate isomerase (phosphoglucose isomerase) 156–157, **F4-47**
glutamic acid (glutamate) 4, **F1-3**
glutamine 4, **F1-3**
glutathione (GSH) 94
D-glyceraldehyde-3-phosphate dehydrogenase (GAPDH) **F2-6**
glycine **F1-3,** 18
glycogen phosphorylase 111, **F3-22**
glycoproteins 122
glycosylation 28–29, 122–123
glycosylphosphatidylinositol (GPI) anchor 124, **F3-45,** 125
glycosyltransferases 140–141
Golgi complex
oligosaccharide processing 122, **F3-42**
vesicular transport 125, **F3-46**
G-protein-coupled receptors (GPCRs) 102–103, **F3-15**
G proteins (GTPases) 50, 98–99, 100–105
α subunit 102
βγ subunits 102
heterotrimeric 88, 102–103
scaffold proteins and 61, **F2-18**
small 100–101
lipid modification 124, 125, **F3-46**
regulation of signaling 100–101, **F3-14**
targeting to membrane 91, **F3-4,** 124, 125
see also Ras
structure 98–99, **F3-12**
switch mechanism 99, **F3-13**

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 206

versus motor proteins **F3-17,** 107
see also EF-Tu
GRASP (surface topography) image 172, **F5-5**
Grb2 **F3-21,** 111
Greek-key motif 34, 36, 37, **F1-58**
green fluorescent protein (GFP) 77, 139, **F4-17**
GRID 148, **F4-31,** 149
GroES co-chaperonin **F1-74**
ground state 62, 63
destabilization 66, 67
growth hormone, human *see* human growth hormone
Gstat **F4-23,** 143
GTP 98
in EF-Tu switching 104–105, **F3-16**
in G protein switching 100, 101, **F3-14,** 102
in microtubule assembly 50-51, **F2-1**
hydrolysis 50, **F2-1,** 98, 100
GTPase-activating proteins (GAPs) 88, 100, **F3-14,** 104
GTPases *see* G proteins
guanine-nucleotide exchange factors (GEFs) 100, **F3-14,** 102
guanosine triphosphate *see* GTP
GYF domain **F3-2**

hairpin turns *see* beta turns
haloperidol **F2-5**
Hck **F3-24**
Hedgehog proteins (HH) 120–121, **F3-39**
helical parameters 14, **F1-14**
helix
alpha *see* alpha helices
cap **F1-13**
dipole 14, 15
helix-turn-helix motif 34, **F1-50, F4-6,** 137
heme 71, **F2-31, F2-37**
heme-containing proteins 28, **F1-39,** 36
hemoglobin 40, 42, 44
cooperative binding 95
oxygen binding 47
sickle cell 42–43, **F1-71,** 44
heptad repeat 40–41, 41
heptamer **F1-74**
Hermansky-Pudlak syndrome 125
heterodimers 40, **F1-65**
heterogeneous nuclear ribonucleoproteins (hnRNPs) 126
heteropentamers **F1-65**
heterotetramers 40, **F1-65**
hexamers 40, **F1-74**
Hidden Markov Model 132, 133
histidine
phosphorylation 110–111, 116–117.
structure and side chain 4, **F1-3**
histidine protein kinase (HK) 116, **F3-29**
histones, post-translational modifications 126, 127
HIV protease **F1-1**
HIV protease inhibitors **F2-5**
homeobox domain **F4-6**
homodimers 40, **F1-65**
homologous genes/proteins 132, 140–141
functional diversity 152–153
prediction of function 162–163, **F4-55**
homology 132
modeling 142–143, **F4-22**
homotrimers 40
hormones, polypeptide 118–119
human genome 136, **F4-10**
human growth hormone **F1-28, F1-51,** 36
binding sites **F2-8**
receptor complex 44, **F1-73**
human immunodeficiency virus (HIV)
protease **F1-1**
reverse transcriptase **F4-30**
hydration shell 23, **F1-26**
hydride ions 68, 69
hydrogen bonds 4, 10, **F1-10**
acceptor atom 10
in alpha helices 14–15, **F1-13**
in beta sheets 16, **F1-17, F1-18**
in binding interactions 59, **F2-15**
donor atom 10
in oligomer formation 42
in phosphorylated proteins 111
in secondary structure elements 12
water molecules 11
hydrolases 33, 147
hydrolysis 8, 72–73
hydronium ions (H_3O^+) 74–75
hydropathy plots 25, **F1-31**
hydrophilic 4
hydrophilic amino acids 4, **F1-3**
hydrophobic 4
hydrophobic amino acids 4, **F1-3**
hydrophobic cores, in domains 30
hydrophobic effect 4, 20
forces driving 26, **F1-34**
in oligomer formation 42
in protein folding 20–21
hydrophobic interactions, in ligand binding 59, **F2-15**
hydrophobic surface, binding sites 58, **F2-12,** 148
hydroxide ion (OH^-) 75, 92
hypoxia-inducible factor (HIF-1a) 109

icosahedron 44, **F1-74**
IκB 109
immunoglobulin A **F3-30**
immunoglobulins 36, **F1-56,** 37
hypervariable loops 22–23, **F1-25**
IMP cyclohydrolase 81, **F2-43**
inclusion bodies 161
induced fit 47, 54, **F2-4**
inhibitors 86–87
allosteric 96
competitive 86, 94, **F3-7**
insertions, sequence 33, 132
insulin 44, **F1-74**
inteins 120–121
mechanism of action **F3-38,** 121
structure 120–121, **F3-37**
interaction domains 86, 88, **F3-2**
interleukin-4 and its receptor **F1-66**
intermediates
in enzymatic reactions 62, 63, 78, **F2-40**
shuttling over long distances 83, **F2-45**
shuttling through internal tunnel 82, **F2-44**
folding 20, **F1-22**
introns 6, **F1-5**
iron (Fe^{2+}, Fe^{3+})
diphtheria toxin expression and 97, **F3-11**
as enzyme cofactor 77, **F2-38**
iron–sulfur proteins 39
isocitrate dehydrogenase (ICDH) 80, **F2-42,** 111, **F3-23**
isoleucine **F1-3**
3-isopropylmalate dehydrogenase **F2-11**

JAK kinase **F2-13**
JAK/STAT signaling pathway 127
jelly roll fold 36, 37, **F1-59**
JNKs **F4-5**

K_d 28, 55
KDEL sequence 91, 164–165
KDGP aldolase **F1-74**
keratin 40
kinases 33
kinesin 106–107, **F3-17**
kirromycin antibiotics 104–105
knockouts, gene 138, 139, **F4-15,** 157
Kss1 **F4-2, F4-4, F4-5**
kT 55

Lac repressor 30, **F1-41,** 31
lactate dehydrogenase **F1-74, F2-30**
Lactococcus lactis, beta-phosphoglucomutase 79, **F2-41**
lambda repressor **F1-50**
Lck **F3-24**
Leishmania major, dihydrofolate reductase-thymidylate synthase 81
lethal factor, anthrax toxin **F2-2**
leucine 4, **F1-3**
leucine zipper 41, **F1-68**
lifetime, protein 87, 108–109
ligands 51, 52
binding sites *see* binding sites, ligand
effector 86–87, 94–97
finding candidate 151
induced conformational changes *see* conformational changes, ligand-induced
induced fit model of binding 54, **F2-4**
macromolecular 56
regulation of protein function 86
small-molecule 56–57, 58
LIM domain **F3-2**
lipid anchor (tail) 86, 90, 91, **F3-4,** 124–125
lipidation 124
lipid bilayer 14, 24, **F1-29**
alpha helices spanning 14, **F1-30,** 25
protein folding in 21
lipid modifications, protein targeting by 124–125, **F3-44**
lipid rafts 91, 124
lipid-transport protein (LTP) **F2-15**
lipoamide (lipoic acid) **F2-37**
beta-lipotropin 118, **F3-33**
localization, intracellular 139, **F4-17,** 164–165
mechanisms *see* targeting, protein
as regulatory mechanism 90
lock-and-key analogy 54
LON **F4-47**
loops
movements 46, **F1-77,** 47
at protein surfaces 22–23, **F1-25**
sequence insertions and deletions 33
LRR domain **F3-2**
lymphoma proprotein convertase 124
lysine
carbamylation 8
methylation 126, **F3-37**
structure and side chain 4, **F1-3**
lysine tyrosylquinone 77, **F2-39**
lysozyme 75, **F2-36,** 160
lysyl oxidase 60–61

macrophage inhibitory factor (MIF) **F4-47,** 157, **F4-48**
magnesium (Mg^{2+})
as enzyme cofactor 77, **F2-38**
in GTPase and ATPase function **F3-12,** 99
malate dehydrogenase 70, **F2-31**
mandelate 152
mandelate racemase (MR)
active site **F2-3, F4-41**
biochemical function **F4-38**
homologs 141, 152–153, **F4-40,** 162, **F4-55**
manganese (Mn^{2+}) **F2-38**
mannose 122, **F3-41**
MAPKs *see* mitogen-activated protein kinases
MARCKS motif **F4-7**
MATα2 protein 158–159, **F4-51**
MCM1 158–159, **F4-51**
MCSS (multiple conformations simultaneous search) 148–149
melanocyte-stimulating hormone (MSH) 118, **F3-33**
membrane proteins 24–25
all beta sheet 25, **F1-32**
alpha-helical 24, **F1-30,** 25
isolated alpha helix 18, 24
membrane spanning **F1-30,** 25

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 207

folding 21
secondary structure elements 18, 24
targeting 91, **F3-4,** 124–125, **F3-44**
membranes 24, **F1-29**
pores/channels 24, 25, **F1-33**
Menkes syndrome 61
mesophilic proteins 26, 27, **F2-6**
messenger RNA (mRNA) 6, **F1-5**
regulation 87
metal ions
cross-linking of domains 39, **F1-64**
enzyme cofactors 77, **F2-38**
in metalloproteases 79
in redox reactions 71
stabilizing tertiary structure 28, **F1-38**
metalloproteases 79
metastable protein folds 160–161
Methanococcus jannaschii **F4-10**
Methanopyrus kandleri 81
methionine 4, **F1-3**
methionine aminopeptidase 2 (MetAP2) **F4-47,** 157
methionine aminopeptidases 109
methylation 126, **F3-37,** 127
methyltransferases 126
microarrays 138, **F4-14,** 165
microfilaments 47
microtubules 50, **F2-1**
kinesin movement along 106–107, **F3-17**
mitogen-activated protein kinase kinase (MAPKK-2) **F2-2**
mitogen-activated protein kinases (MAPKs)
cascade 61, **F2-18, F3-21**
phylogenetic tree **F4-5**
modular structure of proteins 32–33
molybdenum (Mo) **F2-38**
molybdenum cofactor biosynthesis protein C **F1-74**
monomers 40
"moonlighting" 156–157
motifs 34–35
determining intracellular localization 164–165
functional 34, 134–135, **F4-7**
identification from sequence 34–35
sequence 34
structural 34, 35
motions, protein 46–47, **F1-76**
motor proteins **F2-1,** 61, 106–107
MSCS (multiple solvent crystal structures) 149
muconate lactonizing enzyme (MLE)
active site **F4-41**
biochemical function 152, **F4-39**
homologs 141, 152–153, **F4-40,** 162, **F4-55**
multidomain proteins 31, **F1-43, F1-44**
classification 32–33
multifunctional proteins 156–157, **F4-47**
muscle 61
mutagenesis, site-directed 150
mutations
causing abnormal polymerization 42–43, **F1-72**
disrupting protein function 97
oncogenes 101
surface loops 22–23
temperature-sensitive (ts) 26, 27, 108
MutS **F4-27**
Mycobacterium xenopi **F3-37**
Mycoplasma genitalium **F4-10**
myoglobin **F1-1, F1-39,** 36, **F1-55**
sperm whale **F1-78**
myohemerythrin **F1-51,** 36, **F1-54**
myosin 61, 106–107, **F3-17**
myristoylation 124

N-acetylation 28, 29, 126–127, **F3-48**
NAD *see* nicotinamide adenine dinucleotide
NAD-dependent dehydrogenases 69, **F2-30**
native state 20, **F1-22**
neuraminidase 36, **F1-57, F1-74**
neurofibromin **F3-14**
neuroleukin 156, 157
NFκB 109
nickel (Ni^{2+}) **F2-38**
nicotinamide adenine dinucleotide (NAD) **F2-37,** 77
-dependent dehydrogenases 68–69, **F2-30,** 155
in oxidation/reduction reactions 70, **F2-31**
nicotinamide adenine dinucleotide phosphate (NADP) 70, **F2-31, F2-37,** 77
nitric oxide (NO) 127
nitrosylation 126, 127, **F3-50**
N-linked sugars 122, **F3-41**
northern blot 138
N terminus **F1-7**
nuclear magnetic resonance (NMR) spectroscopy
binding-site studies 149
interpretation of information produced 168
quality of model produced 172
resolution 169, 172
structure determination by 171, **F5-4**
structure ensemble 169, **F5-2**
nuclear transport factor-2 141, **F4-21**
nucleases 72–73
nucleophile 4, 78, 79
nucleotide-binding fold 38, 39
nucleotides 6
hydrolysis, protein switches based on 98–105
see also ATP; GTP

octamer **F1-74**
oligomers 40, **F1-65**
formation 42
oligosaccharides 73, 122
processing 122, **F3-42**
O-linked sugars 122, **F3-41**
omega torsion angle **F1-9**
oncogenes 101
operons 162–163
organic solvents, binding-site probes **F4-32,** 149
orotidine 5′-monophosphate decarboxylase (OMP decarboxylase, ODCase) 62, **F2-19, F2-20,** 67
osteonectin motif **F4-7**
oxidation 70
oxidation/reduction (redox) reactions 70–71, **F2-31**
oxyanion hole **F2-40,** 79

p38 **F4-5**
p67 157
p120GAP 88, **F3-14**
packing motif 22, 23, **F1-28**
PAK1 protein kinase 59
PALA (*N*-phosphonoacetyl-L-aspartate) 66–67, **F2-25**
palmitoylation 124
papilloma virus capsid protein **F2-14**
parasites 125
Parkinson's disease 160
partner swapping 58–59, **F2-13**
PBD domain 88
PDZ domain **F3-2, F4-23,** 143
pentamers 40, 44, **F1-74**
peptide bonds 8–9
formation 8, **F1-7,** 73, **F2-33**
hydrolysis 8, **F1-7, F2-33,** 78–79, **F2-40**
properties 9, **F1-8, F1-9**
rearrangements, in protein splicing 120, **F3-35**
percent identity 132, 136–137
pericyclic rearrangement 67, **F2-26**
Pfam database **F4-4**
pH, control of protein function by 92–93
induced conformational switching 93, **F3-5**
phase problem 170
PH domain 88, **F3-2,** 91, **F4-6**
phenylalanine 4, **F1-3**
phi torsion angle 9, **F1-9**
phosphate-binding loop *see* P-loop
phosphoester hydrolysis 72–73, **F2-33**
phosphoglucomutases 79, **F2-41**
phosphoglucose isomerase (PGI) 156–157, **F4-47**
phosphoglycerate kinase (PKA) 68, **F2-29**
phospholipase Cγ 88
phosphorylation 28, 29, 87
in control of protein function 110–111, **F3-21**
glycosylation and 123
in protein targeting 91, **F3-4**
two-component signal carriers 116–117, **F3-29**
phosphotriesterase **F1-47**
phylogenetic tree 133, **F4-5**
Pi-helix **F1-14**
pK_a values 74–75, **F2-35**
plants 8, 116, 136
plasminogen 143, **F4-24**
activation **F3-32**
pleated sheets *see* beta sheets
pleckstrin homology domain *see* PH domain
P-loop **F3-12,** 99, **F4-7,** 137, **F4-12**
pockets, binding 56–57
poly(Lys), circular dichroism spectrum **F1-35**
polynucleotides 73
polypeptide chains
assemblies *see* quaternary structure
properties in water 9, **F1-8, F1-9**
polypeptides 8
polyproline helices 15, **F1-14**
pores, membrane 24, 25, **F1-33**
position-specific score matrix (PSSM) 135, **F4-8**
post-translational modifications 28–29, **F1-40**
in control of protein function 87, 110
glycosylation 122–123
lipid modification 124–125
other mechanisms 126–127
in protein targeting 91
see also phosphorylation; proteolysis
potassium (K^+) **F2-38**
potassium channel 25, **F1-33**
PQQ 28, **F1-39**
pre-albumin 37, **F1-58, F1-69**
precursors, activation of 118–119
pregnenolone biosynthesis 71, **F2-31**
prenylation 124–125
"prepro" hormones 118–119
primary structure (sequence) 3, **F1-2,** 6
tertiary structure determination from 20
see also sequence(s)
primary transcript 6–7, **F1-5**
prions 160–161, **F4-52**
probes, binding site **F4-32,** 149
profile-based threading 144, **F4-25**
profiles, protein 134, 135, **F4-8**
prokaryotes (bacteria)
cell structure 90, **F3-3**
genes 6, **F1-5**
glycoprotein binding 122
post-translational modifications 29
protein degradation 109
protein kinases 110–111
two-component signaling systems 98, 116–117
proline
Cdk2 activation and 115
conformational preferences 18
in helical structures 15, **F1-14**
structure and side chain 4, **F1-3**
pro-opiomelanocortin 118–119, **F3-33, F4-16**
propinquity (proximity) factor 66
PROSITE database 135
protease inhibitors **F2-5**
proteases 34, 72, **F2-33,** 146
endosomal 92–93, **F3-5**
hormone-processing 118–119, **F3-33**
reaction steps 78–79, **F2-40**
see also serine proteases

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 208

proteasome **F1-74,** 108–109, **F3-19**
Protein Data Bank 142, **F4-28**
protein G, *Staphylococcus aureus* 159
protein kinase domains 112, **F3-24,** 153
protein kinases 87, 110–111
  activation mechanism 112–113, **F3-25**
  activation loop (segment) 112, **F3-24, F3-25, F3-26,** 114, **F3-27,** 115, **F3-28**
  autoregulation 88
  localization 90
  multiple alignment **F4-4**
  regulation 112–115
protein microarrays 138
protein phosphatases 110, 115
  two-component signaling systems 116–117
proteolysis
  limited 28, 29, 110, 118–119
  self (autoproteolysis) 120–121
  ubiquitin-dependent 109, **F3-20**
  see also degradation, protein; peptide bonds, hydrolysis
proteolytic cascade 118, 119, **F3-34**
protomer 44
protons 92
proton transfer 52–53, 74–75
proximity 66–67, **F2-25**
proximity factor 66
Pseudomonas aeruginosa 90, **F4-10**
Pseudomonas ovalis 152
pseudosymmetric proteins 44
PSI-BLAST 135
PSIPRED **F4-25**
psi torsion angle 9, **F1-9**
PTB domain 88, **F3-2**
PUA domain **F4-6**
purification, protein 139
PutA **F4-47**
pyridoxal phosphate (PLP) 73, **F2-37**
  -dependent transamination 154, **F4-43**
  in unknown gene product 164, **F4-59**
pyruvate decarboxylase 53, 73, **F2-34,** 141, **F4-20**

quantity, protein 87
quaternary structure 3, **F1-2,** 40–45
  in genetic diseases 42–43
  geometry 44, **F1-75, F1-74**
  intermolecular interfaces 42–43
  intracellular environment and 92
  ligand-induced conformational changes 47
  prediction from sequence 143
  role of complementarity 40–41

Rab 124, 125
Ramachandran plot 12, **F1-11**
RanGAP1 127
Rap-Raf oligomerization **F1-70**
Ras **F1-1,** 100–101, **F3-21**
  lipid modification 125
  mutations 101
  switching cycle 101, **F3-14**
  targeting to membrane 91
reaction sub-site, enzyme active site 64, 65
recessive mutations 43
recognition, molecular 52
recognition modules (interaction domains) 86, 88, **F3-1**
redox environment, control by 92–93
redox reactions 70–71, **F2-31**
reducing environment 10, 92
reduction 70
refinement 170
regulation of protein function
  by *N*-acetylation 126–127
  by degradation 108–109
  by location 90–91
  mechanisms 86–87
  by methylation 126
  by nitrosylation 127
  by pH and redox environment 92–93
  by phosphorylation 110–111
  by proteolysis 118–121
  by sumoylation 127
  by switching *see* switches, molecular
regulator of G-protein signaling (RGS) proteins 102–103
residue 4
resolution 168, **F5-1,** 169, 172
resonance 8
  peptide bonds 9, **F1-8**
response regulator protein (RR) 116–117, **F3-29, F3-30**
retinol-binding protein **F1-19,** 37
reverse transcriptase 147, **F4-30**
reverse turns *see* beta turns
R-factors 172
RGS proteins 102–103
rhinovirus 44, **F1-74,** 97
Rhizobium meliloti DctB protein 25, **F1-31**
rhodopsin 103, **F3-15, F4-23,** 143
ribbon diagram 172, **F5-5**
ribosome 60, **F2-16,** 104
ribozymes 62
RING finger domain **F3-2**
RNA
  binding sites 56
  catalytic (ribozymes) 62
  coenzymes and 77
  degradation 87
  editing 6, 7
  interference (RNAi) 138, 139
  messenger (mRNA) 6, **F1-5**
  regulation 87
  ribosomal 62
  transfer (tRNA) 104, **F3-16**
RNA polymerases 147
root-mean-square deviation (RMSD) 169, 172
Rosetta method 144–145, **F4-26, F4-27**
Rossmann fold 34

Saccharomyces cerevisiae (yeast) **F4-10**
  MATα2 158–159, **F4-51**
  protein functions 137, **F4-13**
  Tem1 90
  two-hybrid system 138, 139, **F4-18,** 165
  YBL036c **F4-60,** 164–165, **F4-59**
S-acylation 124, 125
SAGE (serial analysis of gene expression) 138
Salmonella typhimurium 82
salt bridges 10, **F1-10,** 42
  in phosphorylated proteins 111
Sam68 126
SAM domain **F3-2**
scaffold proteins 60, 61, **F2-18**
  protein binding to 91, **F3-4**
Schizosaccharomyces pombe 90
scorpion toxin 39, **F1-63**
scrapie 160
scytalone dehydratase 141, **F4-21**
secondary structure 3, **F1-2,** 12–13
  classifying types of folds from 12
  elements 12
    membrane proteins 18, 24–25
    see also alpha helix; beta sheet; beta turn
  prediction 18, **F1-21,** 134
  steric constraints 12, **F1-11**
selenium (Se) **F2-38**
"semi-liquid" nature of proteins 46
septin 127
sequence(s) 3, **F1-2,** 6
  alignment 132–133
    active-site residues 162
    local 134–135, 137
    multiple 132, 133, **F4-4**
    pairwise 132–133, **F4-2**
  analysis 132–135
  chameleon 158–159, **F4-49**
  comparison 132, 134
  conserved 132, 133
  deriving function from 136–137, **F4-11**
  divergent and convergent evolution 140–141, **F4-19**
  growth of information on 136, **F4-9**
  identity/similarity 132–133, **F4-3,** 142
    40% rule 136–137, **F4-11**
    percent 132, 136–137
  insertions and deletions 33, 132
  secondary structure prediction from 18, **F1-21,** 134
  signal 86, 91, **F3-4**
  structure determination from 142–145
  see also primary structure
Ser1 125
serine
  phosphorylation 110, 111, 112
  structure and side chain 4, **F1-3**
serine proteases
  activation 118, **F3-32**
  catalytic triad 34–35, **F1-52,** 78, 146–147
    detection methods 150, **F4-35**
  convergent evolution 140, 146–147, **F4-29,** 155
  multi-step reactions 78–79, **F2-40**
  precursors, homology modeling 143, **F4-24**
serpins (serine protease inhibitors) 161, **F4-54**
seven-transmembrane helix fold 153
SH2 domain 88, **F3-2,** 153
  in analysis of protein function **F4-6,** 135, 137
  in protein kinase regulation 112–113, **F3-26**
  in STAT signaling 58
SH3 domain 88, **F3-2**
  in analysis of protein function 134, **F4-6,** 135, 137
  in protein kinase regulation 112–113, **F3-26**
  in protein targeting 91
sickle-cell anemia 42–43, **F1-71**
side chains 4
  amino acid 4, **F1-3**
  movements 47
signal sequences 86, 91, **F3-4**
signal transduction pathways 87, 88
  protein switches 98
  regulation by phosphorylation 110, **F3-21**
  regulation by protein localization 90
signal transduction proteins
  domain arrangements 32, **F1-46**
  partner swapping 58–59, **F2-13**
  scaffold proteins 61
silk **F1-1,** 16, 37, 60
single-nucleotide polymorphism (SNP) 6, 7
site-directed mutagenesis 150
Smt3 127, **F3-49**
SNARE domain **F3-2**
sodium dodecyl sulfate (SDS) 55
solubility, glycosylation and 123
soluble proteins
  folding 20–21, **F1-23**
  hydration shell 23, **F1-26**
Sos **F3-14, F3-21**
Sp100b **F4-27**
space-filling image 172, **F5-5**
specificity
  enzyme-mediated reactions 64, **F2-24**
  ligand binding 52, 59
  sub-site, enzyme active site 64, 65
splicing, protein 120–121, **F3-35, F3-38**
sporulation, yeast 165
Src kinases 112–113, **F3-26**
Src-Lck 32, **F1-46**
S–S bridges *see* disulfide bridges
stability, protein 26–29, 108

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 209

agents disrupting 27
covalent bonds in 28
post-translational modifications and 28–29
secondary structure role 12
weak interactions in 10–11, **F1-10,** 26–27
Staphylococcus aureus, protein G 159
STAT (signal transducer and activator of transcription) molecules 58–59, **F2-13**
Ste5p 61, **F2-18**
steroid-delta-isomerase 141, **F4-21**
steroid hormone biosynthesis 71, **F2-31**
stop codon 6, 7
stress-activated protein kinases (SAPK) **F4-5**
stress-response proteins 108, 109
structural genomics 146
structural proteins **F1-1,** 3, 50, 60–61
catalytic proteins as 61
forming assemblies 60–61
as frameworks and connectors 60
as scaffolds 61
structure, protein
accuracy and precision 168–169
convergent evolution 140–141, **F4-19,** 154–155
determination 168–173
divergent evolution 140–141, **F4-19,** 154, 155
diversity 32–33
four levels 3, **F1-2**
and function 33
functional diversity of specific 152–153
growth of information 146, **F4-28**
interpretation of information 168–169
membrane protein 24–25
metastable 160–161
modular nature 32–33
prediction from sequence 134
prediction of function from 130, 134–135
presentation methods 172, **F5-5**
primary 6
quality of model produced 172
quaternary 40–45
resolution 168, **F5-1,** 169
secondary 12–19
tertiary 22–23
substrates 50, 62, 63
chirality 155
finding candidate 151, 162–163
subtilisin **F1-52, F4-29**
binding sites **F4-33**
coordinate covalent bonds **F1-38**
subunits 40
binding sites at interface between 57, **F2-11**
succinate dehydrogenase 70, **F2-31**
sugar isomerases 155
SUMO 127, **F3-49**
sumoylation 126, 127, **F3-49**
superfamilies, protein 134, 135, 140, 146–147
evolution of function 146–147
functional variation within 147
TIM-barrel proteins 152–153
serine proteases 146–147
DNA polymerases 147
superoxide dismutase 36, **F2-22**
surface plasmon resonance 139
surfaces, protein
amphipathic beta sheets 17
beta turns 13
binding sites for macromolecules 56, **F2-8, F2-9**
hydrophobic side chains at 21
loops 22–23, **F1-25**
movements of atoms at 46
surface topography image 172, **F5-5**
switch I regions **F3-12,** 99, **F3-17,** 107
switch II regions **F3-12,** 99, **F3-17,** 107
switches, molecular **F1-1,** 3, 50, 51, 98–107
based on nucleotide hydrolysis 98–105
see also G proteins; motor proteins; two-component signaling systems
symmetric protein complexes 44, **F1-75**
synthesis, protein 73
GTPase switches in 104–105, **F3-16**
see also transcription; translation

T4 lysozyme **F1-79**
Taq DNA polymerase 27, **F4-30**
targeting, protein 86
to degradative machinery 87, 108–109, **F3-20**
by lipid modification 124–125, **F3-44**
mechanisms 91, **F3-4**
TATA binding protein **F1-1, F1-62**
taxol **F2-1**
Tem1 90
temperature
dependence, enzyme function **F2-6,** 55
high, protein denaturation 27
protein stability and 108
sensitivity 26, 27
temperature-sensitive (ts) mutations 26, 27, 108
tertiary structure 3, **F1-2,** 22–23
determinants 20
flexibility 27, **F1-36**
molecular recognition and 52
packing of atoms in interior 23, **F1-27**
post-translational modifications and 28–29
prediction from sequence 134, 142–145
stability *see* stability, protein
water molecules at surface 23, **F1-26**
tetrahedron 44
tetrahydrofolate **F2-37**
tetramers 40, 41, 44, **F1-74**
theoretical microscopic titration curves (THEMATICS) **F4-36,** 151
thermodynamics, folded proteins 26–27
thermolysin 79, 149, **F4-34**
thermophiles 62, 108
thermophilic proteins 26, 27, 55, **F2-6**
Thermotoga maritima **F2-6**
thiamine pyrophosphate (TPP, thiamine diphosphate, TDP) 73, **F2-34, F2-37**
thioesterase, *Escherichia coli* 31, **F1-43**
thioester dehydrase 31, **F1-43**
thioredoxin **F4-47**
threonine
phosphorylation 110, 112, 114–115, **F3-27**
structure and side chain 4, **F1-3**
thrombin **F4-47**
thymidylate synthase 81, 83, **F4-47**
TIM *see* triosephosphate isomerase
TIM-barrel proteins 35, **F1-53**
functional versatility 152–153
of unknown function 164
TIM barrels 12, 33, 38, **F1-61**
torsion angles 9, **F1-9**
toxins 39
tox operator DNA sequence **F2-9**
TPP *see* thiamine pyrophosphate
transamination 154
transcription 6, 7, **F1-5**
regulation 87
transducin 103
transfer RNA (tRNA) 104, **F3-16**
transition state 62, 63, **F2-21**
stabilization 68, **F2-27, F2-28**
translation 6, 7, **F1-5**
GTPase switches in 104–105
transthyretin 160
TRBD domain **F4-6**
trimers 40, 41, 44, **F1-74**
triosephosphate isomerase (TIM) **F1-24, F1-53**
catalytic residues **F4-36, F4-37**
ligand-induced conformational change **F1-77**
phosphoglucose isomerase and 157
reaction rate **F2-20**
tropomyosin **F1-67**
troponin C 95
trypsin 118
trypsinogen 143
tryptophan 4, **F1-3**
tryptophan synthase **F1-45,** 82, **F2-44**
Tu *see* EF-Tu
tubulin 50, **F2-1**
two-component signaling systems 98, 116–117
conserved features of RR regulatory domains 117, **F3-30**
mechanisms 116–117, **F3-29**
two-dimensional gel electrophoresis 138, **F4-15,** 165
two-hybrid system, yeast 138, 139, **F4-18,** 165
tyrosine
phosphorylation 110, 111, 112
structure and side chain 4, **F1-3**
tyrosine protein kinases 110, 111, 112–113

ubiquitin 108, 109
ubiquitination 108, 109
up-and-down structural motif 36, 37, **F1-57**
urea 55
urease **F2-20**

valine **F1-3**
van der Waals interactions 4, 10, **F1-10**
in enzyme binding 64
vesicular transport, control of 124, 125, **F3-46**
viruses
coat proteins 59, **F2-14**
oligosaccharide binding 122, 123
see also specific viruses
vitamins, cofactors derived from **F2-37**

Walker motif **F4-7,** 137, **F4-12**
water
displacement, in binding interactions 59
exclusion, in enzyme reactions 68–69, **F2-30**
hydrogen-bonding properties 11
in peptide bond formation **F1-7**
properties of polypeptide chains in 9, **F1-8, F1-9**
protein folding in 20–21, **F1-23**
in protein interior 23
protein stability and 26, **F1-34**
in quaternary structure 42, **F1-69**
removal/addition reactions 72–73
on surface of folded protein 23, **F1-26**
WD40 domain 88, **F3-1, F4-6**
weak noncovalent interactions
in binding interactions 52, 58–59
in protein assemblies 60
stabilizing folded proteins 10–11, **F1-10,** 26–27
stabilizing quaternary structure 42
wire diagram 172, **F5-5**
WW domain 88, **F3-2**

X-ray crystallography
binding-site studies 149
interpretation of information produced 168
quality of model produced 172
resolution 169
structure determination by 170, **F5-3**

YBL036c 164–165, **F4-59, F4-60**
yeast *see Saccharomyces cerevisiae*

zinc (Zn^{2+}) 28, 77, **F2-38**
zinc finger motif 38, 39
functional diversity 153
in protein characterization **F4-7,** 137
structure 34, **F1-49, F1-64**
z-score 144, **F4-25**

©2004 New Science Press Ltd

Petitioner Merck, Ex. 1209, p. 210

Protein Structure and Function

Gregory A Petsko & Dagmar Ringe

What the expert reviewers said:

...about Chapter 1

"...this chapter covers truly a lot of ground with some insightful depth and cohesion. It is written by two experts in the field who are not afraid to introduce the details of elements of the field in depth, yet cover the paradigms well. In many ways it is a powerful successor to the several books on protein structure... It has certainly moved well beyond the scope of other books."

Robert Stroud
University of California, San Francisco

"This [text] focuses on structure and stability, rather than on biological categories, and it does an excellent job of outlining the fundamental themes. It also ties in physical chemistry (through forces and energetics)....We could therefore use it as a basic text in the protein structure plus physical chemistry course we teach for juniors at Harvard."

Stephen Harrison
Harvard

...about Chapter 2

"A really good book provides a wealth of information in a manner that the reader remembers what he/she has read. This book does just that. The contents just pop into mind from time to time. This book will be a winner."

Tom Bruice
University of California, Santa Barbara

...about Chapter 3

"I've reviewed the materials that you sent regarding the book from Petsko and Ringe. I'm very excited about what they are putting together. This treatment and focus is much needed. It will be essential reading by students and faculty alike, and will serve as a crucial reference book for everyone in the biosciences."

James Spudich
Stanford University School of Medicine

"I very much liked the Petsko-Ringe signaling section—it covered all the major bases in a thoughtful and coherent way; I'm amazed they managed to make the complexities of signal transduction so intelligible."

Tony Pawson
Samuel Lunenfeld Research Institute

...about Chapter 4

"It is well-written, makes for interesting reading and contains a number of excellent examples. The whole book appears to be an important, useful and timely one."

Roman Laskowski
European Bioinformatics Institute

ISBN 978-0-19-955684-7

Petitioner Merck, Ex. 1209, p. 211