1-20 Quaternary Structure: Intermolecular Interfaces

Figure 1-69 Water molecules at a subunit interface Pre-albumin is a dimeric plasma protein that binds iodinated hormones. The structure of the dimer clearly shows a network of water molecules (red spheres) trapped between the two subunits. (PDB 1bm7)

All types of protein-stabilizing interactions contribute to the formation of intermolecular interfaces

The weak bonds that hold subunits together are the same as those that stabilize the folded structures of monomeric proteins (see section 1-4). Thus, hydrophobic interactions, hydrogen bonds and salt bridges are all observed at the interfaces of protein–protein and protein–peptide complexes. Cross-linking interactions, such as disulfide interactions and metal-ion ligation, also occur at some interfaces. Of these interactions, the hydrophobic effect deserves special mention. The portion of the surface area of a subunit that becomes buried when an oligomer forms is, in general, less polar than is typically the case for surface regions of soluble, monomeric proteins. The amount of surface area that is actually buried at an interface varies greatly. Salt bridges, which would usually be found at the exposed surface as they involve charged residues, are surprisingly common at interfaces. Examples are found in virus structures, in coiled-coil interactions and in the recognition of phosphate groups by signaling proteins, among others. Hemoglobin has several intersubunit salt bridges, which can break in response to a change in pH, altering the relative orientations of the subunits and the affinity of the protein for oxygen. A rough correlation exists between the stability of the oligomer and the type of interaction that predominates at the interface. Very stable oligomers tend to bury a large hydrophobic surface area between subunits, whereas subunits that assemble and disassemble more easily as part of their function seem to employ more polar interactions.

The atomic-packing density at the interface between subunits usually approaches that of the interior of a monomeric protein, but water molecules are present more often at subunit interfaces than they are in protein interiors. Some water molecules may be trapped when individual folded subunits associate, but others are likely to have more important roles. The subunits of many oligomeric proteins are only partially folded before oligomerization (as we have seen in the case of the leucine zipper proteins: see Figure 1-68), and only in the context of their neighbors do they assume their final, correct, tertiary structure. Thus, it may be that water molecules found at interfaces are essential for preserving the structures of the partially folded monomer units before aggregation (Figure 1-69).

Hydrogen-bonding potentials at the interface between subunits must be satisfied just as they are throughout any folded protein; many of the hydrogen bonds are between subunits, and the rest are with interfacial waters. Because hydrogen bonds are highly directional, they orient interactions between subunits, and provide much of the specificity for complex oligomerizations. Sometimes, intersubunit hydrogen bonding can be part of the secondary structure of a protein; there are many cases where a beta sheet continues smoothly from one subunit to the next (Figure 1-70).

Figure 1-70 Oligomerization by beta sheet formation The signal transduction proteins Rap (left) and Raf (right) both contain beta sheets with exposed edge strands. These proteins form a heterodimer by using the edge strands to complete a continuous extended beta sheet that traverses both molecules. (PDB 1gua)

Inappropriate quaternary interactions can have dramatic functional consequences

A number of genetic diseases, or striking phenotypes, originate from hydrophobic surfaces that are inappropriately created as a result of mutation. Sickle-cell anemia is one example. The mutation of a glutamate to valine on the surface of the beta subunit of hemoglobin creates a hydrophobic patch that causes hemoglobin tetramers to polymerize into long fibrils (Figure 1-71): the polymer is formed by burial of the new patch at the tetramer–tetramer interface.

In the case of the sickle-cell mutation, abnormal polymerization occurs through the generation of an inappropriate hydrophobic interaction. But the normal polymerization interactions of

Definitions

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.

©2004 New Science Press Ltd

Halozyme EX2164 (Part 2 of 4)
Merck v. Halozyme
PGR2025-00030

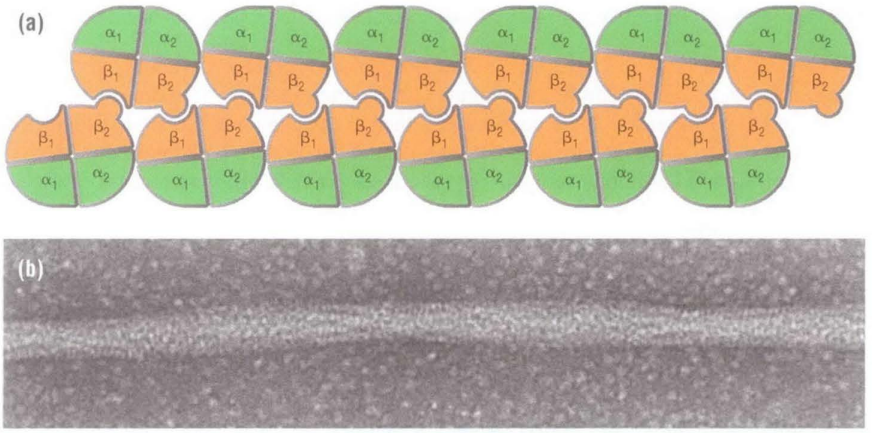


Figure 1-71 Sickle-cell hemoglobin Hemoglobin molecules form long polymers when they carry the sickle-cell mutation, in which a hydrophobic patch is created on the surface of the tetramer by the substitution of a hydrophobic valine for a hydrophilic glutamine in the beta subunit. **(a)** The hydrophobic patch created by the mutant valine is represented by a bump in the beta2 subunit, which binds in a hydrophobic pocket in the beta1 subunit of another hemoglobin molecule. Because the hydrophobic pocket into which the mutant valine binds is present only in the deoxy form of hemoglobin, the formation of the fibers, which constrains the molecule in the deoxy state, also functionally disables it. **(b)** Polymers of sickle-cell hemoglobin aggregate to form thicker fibers. These rigid fibers distort the hemoglobin-carrying red blood cells, causing them to rupture or to block blood vessels, with painful and sometimes fatal consequences. Severe anemia is thus not the only pathological consequence of the sickle-cell mutation. (Courtesy of Stuart J. Edelstein.)

oligomeric proteins can also make them more susceptible to disruption by mutation than are monomeric proteins. In the case of monomeric proteins, loss of function by mutation usually has to occur in both copies of the gene in question before the individual is affected: that is, the individual must be homozygous, and the effect is said to be recessive. In oligomeric proteins, however, mutant subunits produced by one copy of the gene may disrupt the function of normal subunits produced by the other, unmutated, copy so that the effect of the mutation is seen even in the presence of one normal gene: that is, when the individual is heterozygous. This is known as the **dominant-negative** effect, and can occur when the interaction surface is intact in the mutant subunit but the active site is abnormal or missing. Thus, for example, the introduction of a premature STOP codon may result in the expression of a protein fragment that can bind to a normal subunit but lacks the functional domain that contributes to the active site of the protein (Figure 1-72). The fragment will thus act as an inhibitor for the function of the intact subunit. This is the basis, for example, of an immune deficiency in the response to mycobacteria, in which anti-mycobacterial mechanisms normally induced in cells by interferon are abrogated by a mutant subunit of the heterotetrameric interferon receptor.

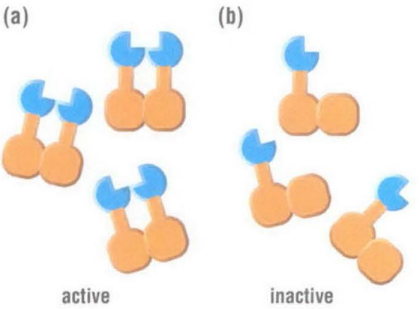


Figure 1-72 Dominant-negative phenotype resulting from hydrophobic interactions between mutant and normal subunits of a dimeric protein **(a)** Schematic representation of a protein, active only as a dimer, in which one domain (yellow) is necessary for dimerization while the other (blue) forms an active site on dimerization. **(b)** A truncated protein comprising the dimerization domain only and produced from a mutant copy of the gene can associate with normal protein to form an inactive dimer lacking an active site. Some active dimers will also form between normal monomers, so this will produce a dominant lack-of-function phenotype only if either normal function requires normal levels of the protein, or the truncated mutant is produced in excess, so that almost all of the dimers are inactive.

References

Bunn, H.F. and Forget, B.G.: *Hemoglobin: Molecular, Genetic and Clinical Aspects* Chapter 11 (Saunders, New York, 1986).

Harrington, D.J. *et al.*: **The high resolution structure of deoxyhemoglobin S.** *J. Mol. Biol.* 1997, **272**:398–407.

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

Jouanguy, E. *et al.*: **A human IFNGR1 small deletion hotspot associated with dominant susceptibility to mycobacterial infection.** *Nat. Genet.* 1999, **21**:370–378.

©2004 New Science Press Ltd

1-21 Quaternary Structure: Geometry

Figure 1-73 The human growth hormone–receptor complex Structure of the human growth hormone (yellow) complexed with two identical molecules of its receptor (orange and green). The receptor is a membrane protein, but only the extracellular hormone-binding portion is shown. The plane of the membrane is indicated by the slanted line. A molecule of the monomeric hormone binds to two identical receptor molecules. Similar regions of the two receptor molecules are used to bind two distinct regions of the hormone; the conformational flexibility of these regions allows for this versatility. (PDB 3hhr)

Protein assemblies built of identical subunits are usually symmetric

Protein complexes are built up through interactions across complementary binding surfaces. If one subunit has binding region A, the subunit it binds to must have the complementary region A′. If the interacting subunits are not identical, then nothing definite can be said about the spatial relationship of the monomers in the complex and the complex is said to be asymmetric. The human growth hormone–receptor complex is an example of an asymmetric complex (Figure 1-73).

If the subunits are identical, however, interactions across complementary surfaces nearly always produce symmetric complexes, in which the subunits are related to one another with one of a few kinds of geometry (Figure 1-74). Identical subunits form symmetric complexes because, in order to interact, each subunit must possess binding region A and its complement A′. (This is in contrast to non-identical subunits, each of which has only one or the other.) Depending on the location of A and A′ on the surface, subunits can associate to form closed structures (Figure 1-75a and b), with dimers and trimers being most common, or, much more rarely, open-ended chains, with helical arrangements being most common.

The repeating unit from which a symmetric complex is built can be either a monomer or an association of unlike polypeptide chains. For example, hemoglobin, which is constructed from four polypeptide chains, (a2b2), is a symmetric dimer of two (ab) units. The asymmetric unit from which a symmetric complex is built is referred to as the **protomer**.

If the subunit has a second set of complementary binding regions, B and B′, in addition to A and A′, it can associate to form more elaborate complexes (Figure 1-75c). A second binding region can allow symmetric rings to pair, with pairs of dimers that form tetramers and pairs of trimers that form hexamers being the most common. Insulin is an example of a hexameric protein that is built in this way (Figure 1-74f). In a similar way, open-ended chains can associate side-by-side to form multistranded helices. This is what happens in sickle-cell hemoglobin, when an additional binding site is created by mutation (see Figure 1-71). Subunits with two sets of complementary binding regions can also associate into more complex structures, usually described by reference to geometric figures—tetrahedra, octahedra, icosahedra—with the same symmetry. Type II 3-dehydroquinate dehydratase, for example, crystallizes as a dodecamer in which a tetramer of trimers forms a tetrahedron (Figure 1-74j); and the rhinovirus that causes the common cold is a large multisubunit icosahedron (Figure 1-74k).

So powerful is the tendency of subunits to form symmetric arrangements that this even influences the structure of oligomers made up of non-identical polypeptide chains. Many of these proteins are **pseudosymmetric**, as we have already seen for hemoglobin. In this protein the alpha and beta subunits are similar in sequence and hence nearly identical in structure, so it is a nearly symmetrical tetramer of four monomers. The giant multisubunit proteolytic complex called the proteasome is another example of a pseudosymmetric structure (Figure 1-74l).

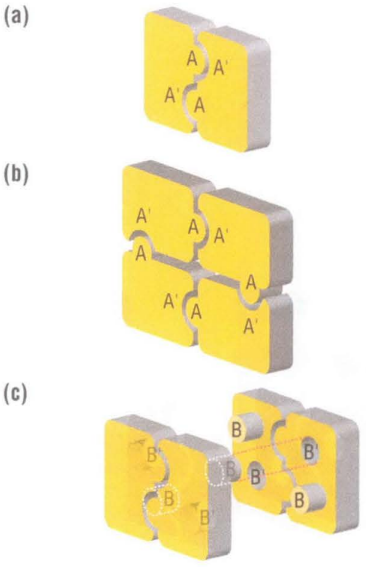


Figure 1-75 Interactions underlying different geometric arrangements of subunits Subunits with a pair of complementary binding sites A and A′ may form symmetric dimers **(a)** or tetramers **(b)** depending on the positions of the two binding sites. More complex assemblies may be formed by subunits with a second pair of complementary binding sites B and B′ that could for example allow the formation **(c)** of a tetrameric complex of two dimers.

Figure 1-74 Examples of quaternary arrangements observed for oligomeric proteins The structures shown in **a-k** are homo-oligomers. The proteasome **(l)** is a pseudo-symmetric structure, in which the subunits are not identical. **(a)** D-amino acid aminotransferase (PDB 3daa); **(b)** KDGP aldolase (PDB 1fq0); **(c)** neuraminidase (PDB 1a4q); **(d)** lactate dehydrogenase (PDB 1ldn); **(e)** cholera toxin (PDB 1chp); **(f)** insulin (PDB 4ins); **(g)** molybdenum cofactor biosynthesis protein C (PDB 1ekr); **(h)** GroES co-chaperonin (PDB 1g31); **(i)** galactonate dehydratase; **(j)** 3-dehydroquinate dehydratase (PDB 2dhq); **(k)** rhinovirus (PDB 1aym): this multisubunit protein has the same geometry as a soccer ball; **(l)** proteasome (PDB 1g65).

Definitions

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

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

References

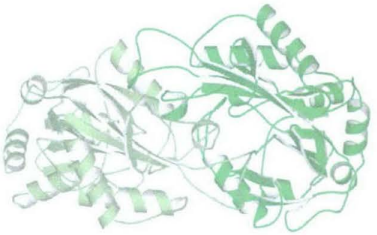
Goodsell, D.S. and Olson, A.J.: **Structural symmetry and protein function.** *Annu. Rev. Biophys. Biomol. Struct.* 2000, **29**:105–153.

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

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

©2004 New Science Press Ltd

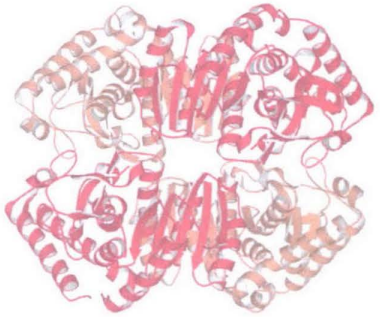
(a) dimer



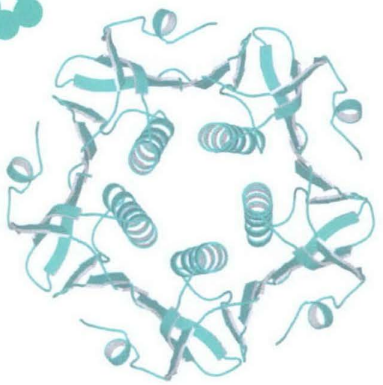
(b) trimer

(c) planar tetramer

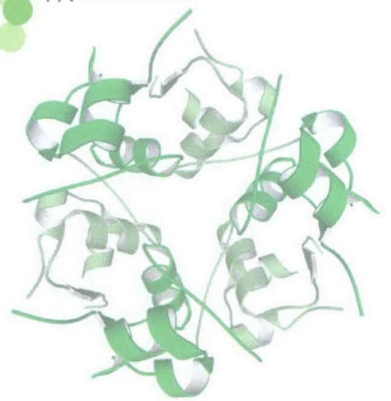
(d) tetramer



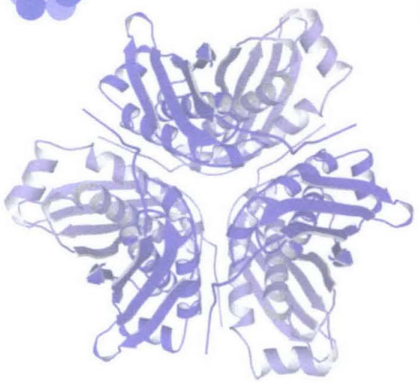
(e) pentamer



(f) planar hexamer

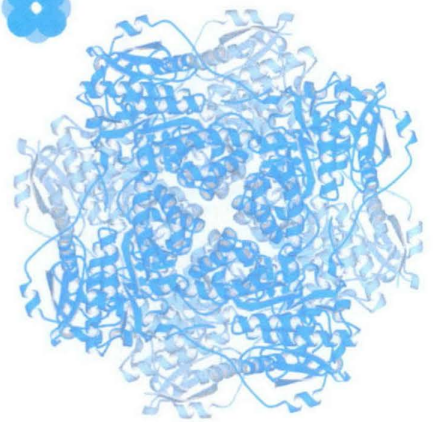


(g) hexamer (trimer of dimers)

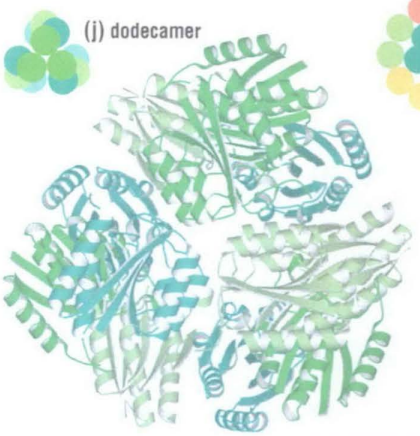


(h) heptamer

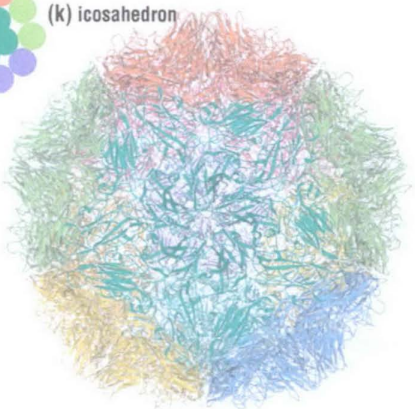
(i) octamer



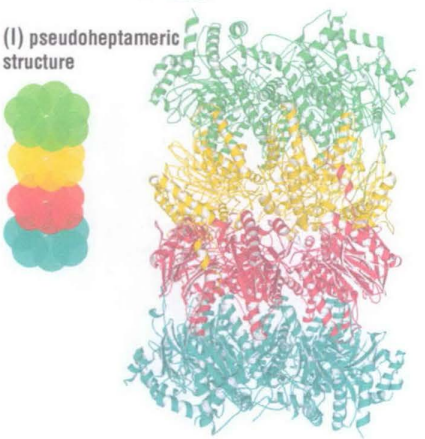
(j) dodecamer



(k) icosahedron



(l) pseudoheptameric structure



©2004 New Science Press Ltd

1-22 Protein Flexibility

Types of Motion Found in Proteins (all values approximate)

Motion	Spatial displacement (Å)	Characteristic time (s)	Energy source
Fluctuations (e.g., atomic vibrations)	0.01 to1	10^{-15} to 10^{-11}	k_bT
Collective motions (A) fast, infrequent (e.g., Tyr, Phe ring flips) (B) slow (e.g., domain movement; hinge-bending)	0.01 to > 5	10^{-12} to 10^{-3}	k_bT
Triggered conformational changes	0.5 to > 10	10^{-9} to 10^{3}	Binding interactions

Figure 1-76 Table of protein motions

Proteins are flexible molecules

The pictures of protein structures that emerge from X-ray crystallography and NMR seem rigid and static; in reality, proteins are highly flexible. Because the forces that maintain the secondary and tertiary folds are weak, there is enough energy available at body temperature to break any particular interaction. When existing weak interactions are broken, the groups that are released can make new interactions of comparable energy. These rearrangements can occur on a time scale that is faster than the time required to determine the structure by tools such as X-ray crystallography. Thus, the three-dimensional structures of proteins determined by physical techniques are average structures.

Protein motions can be classified in terms of their relationship to the average structure (Figure 1-76). The fastest motions are atomic fluctuations such as interatomic vibrations and the rotations of methyl groups. Next come collective motions of bonded and non-bonded neighboring groups of atoms, such as the wig-wag motions of long side chains or the flip-flopping of short peptide loops. The slowest motions are large-scale, ligand-induced conformational changes of whole domains.

Conformational fluctuations in domain structure tend to be local

It is almost as important to understand what types of conformational change are not observed in proteins as it is to realize that they are flexible in the first place. Whole folded domains never undergo large, thermally driven distortions at ordinary temperatures. Transitions from one type of folding motif to another are rarely seen except in pathological cases; an all alpha-helical protein will not normally refold to an all beta-sheet protein, except, for example, in the cases of amyloid and prion diseases. Smaller-scale refolding does occur in some proteins, however. Ligand binding may induce disordered polypeptide segments to become ordered. Ligands can also induce the disordering of a previously ordered strand, although this is less common. Association and dissociation of subunits can also be triggered by ligand binding, and the ligand can be as small as a proton if it changes the charge of a crucial residue.

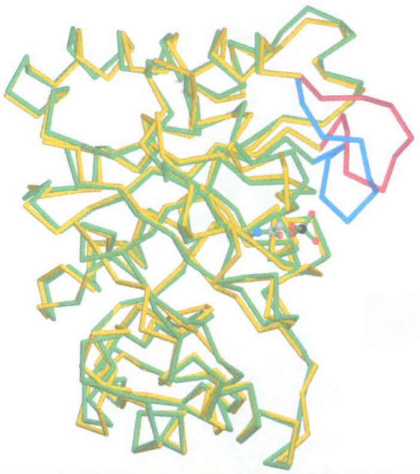


Figure 1-77 Triosephosphate isomerase Binding of substrate or inhibitor to the active site of the enzyme triosephosphate isomerase induces a 10 Å rigid-body movement in an eight-residue loop (red; open) which closes down over the active site (blue; closed) and shields the substrate from solvent. The inhibitor can be seen just below the loop.

Perhaps the most common ligand-induced conformational change is the lid-like movement of a polypeptide segment to cover a ligand-binding site (Figure 1-77). When the lid is open, there is free access to the ligand-binding site. Once the site is occupied, the loop interacts with the ligand to stabilize the closed conformation, and closure isolates the bound ligand from the surrounding solvent. Most loop closures involve rigid-body movement of the loop on two hinges. The internal conformation of the loop does not change appreciably because its side chains are packed closely together, making it function like a solid lid for the ligand-binding site. Mobile loops both act as gates for ligand binding and can make interactions that stabilize the complex. They play an important part in many enzymes.

Protein motions involve groups of non-bonded as well as covalently bonded atoms

At body temperature, the atoms in most protein molecules fluctuate around their average positions by up to an Ångstrom or occasionally even more, depending on their position in the protein (Figure 1-78). In the tightly packed interior, atomic motions are restricted to less than an Ångstrom. The closer to the surface of the molecule, the greater the increase in mobility until, for surface groups that are not surrounded by other atoms, the mean fluctuation may be several Ångstroms. Proteins have been called "semi-liquid" because the movements of their atoms are larger than those found in solids such as NaCl, but smaller than those observed in a liquid like water.

In a protein, the covalent structure of the polymer sets limits on the motions of atoms and groups of atoms. Chemical groups such as methyl groups or aromatic side chains display collective motions. Methyl groups rotate on a picosecond time-scale; aromatic rings, even those in the interior of the protein, flip at average rates of several thousand per second. The actual ring flip takes only about a picosecond, but it happens only about once every 10^9 picoseconds.

Figure 1-78 Protein shaded according to flexibility Space-filling model of sperm whale myoglobin in which each atom is shaded according to its average motion as determined by X-ray crystallography. The darker the atom, the more rigid it is. Note that the surface is not uniform in its flexibility. (PDB 1a6k)

©2004 New Science Press Ltd

Flipping an aromatic ring inside a protein, where the packing density is high, requires that surrounding atoms move out of the way. The probability that they will all move in the right direction at the same time is very low: hence the relatively long interval between flips. In the interior of proteins, close atomic packing couples the motions of non-bonded neighboring atoms. If a methyl group in the center of a protein is next to another methyl group, the motions of both will be correlated by virtue of their tendency to collide. Thus both the extent of motion of every group and its preferred directions depend on non-bonded as well as bonded contacts. Only for surface side chains and protruding loops are non-bonded interactions of little importance, and residues in such unrestrained positions are always the most flexible parts of a protein structure.

At biological temperatures, some proteins alternate between well-defined, distinct conformations (Figure 1-79). In order for two conformational states to be distinct, there must be a free-energy barrier separating them. The motions involved to get from one state to the other are usually much more complex than the oscillation of atoms and groups about their average positions. It is often the case that only one of the alternative conformations of a protein is biologically active.

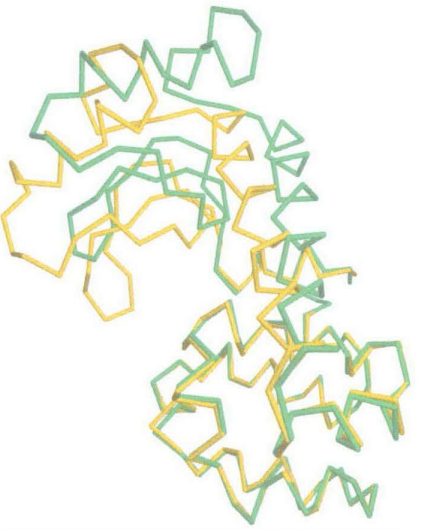


Figure 1-79 T4 lysozyme The enzyme T4 lysozyme contains two domains connected by a hinge. In different crystal forms of the protein, an open and closed state have been observed, related to each other by a hinge-bending motion. It is presumed that the protein in solution can exist in an equilibrium between both states at physiological temperature. (PDB 1l96 and 1l97)

Triggered conformational changes can cause large movements of side chains, loops, or domains

Of most importance for protein function are those motions that occur in response to the binding of another molecule. Ligand-induced conformational changes can be as modest as the rearrangement of a single side chain, or as complex as the movement of an entire domain. In all cases, the driving force is provided by ligand–protein interactions.

Often, the motion enables some part of the structure to make contact with a ligand. For example, the binding of aspartate to a large domain of the enzyme aspartate aminotransferase causes a smaller domain to rotate by 10°. This rotation moves the small domain by more than 5 Å, bringing it into closer contact with the rest of the protein (Figure 1-80). When the binding of a specific ligand causes a protein to change from an inactive to an active conformation, the process is described as **induced fit**. The driving force for induced fit in aspartate aminotransferase appears to be the formation of a salt bridge between an arginine residue in the mobile domain and the alpha-carboxylate of the bound aspartate. Mutant enzymes that are unable to carry out this triggered conformational change are inactive. We discuss the use of conformational changes to regulate enzymes in more detail in Chapter 2.

Ligand-induced conformational changes can also change the quaternary structure of proteins. This usually involves repacking of the interfaces between subunits so that the relative positions of the monomers are altered; this happens when oxygen binds to the tetramer hemoglobin. Sometimes, however, the stoichiometry of the oligomer changes on ligand binding. One example is the polymerization of actin monomers, driven by the binding of ATP, into linear helical polymers called thin filaments or microfilaments. Regulated polymerization of actin is essential for the formation and disassembly of cytoskeletal components needed for cell movement. Changing the oligomeric state of actin is a mechanism for controlling what it does.

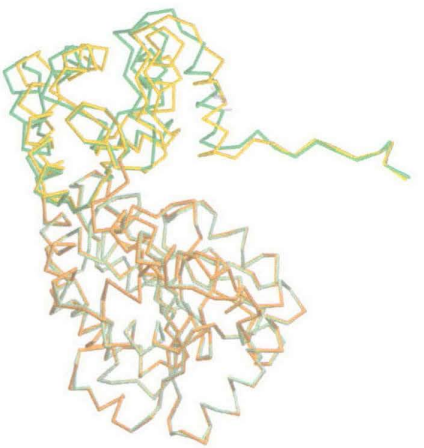


Figure 1-80 Aspartate aminotransferase, open and closed forms The enzyme L-aspartate aminotransferase contains two domains with the active site lying between them. Substrate binding induces a movement of the small domain (green) to a new position (yellow) in which the active site is more enclosed. This movement is essential to position some of the residues important for catalysis, and only the specific substrates of the enzyme induce it. (PDB 1ars and 1art)

Definitions

induced fit: a change in the conformation of a protein induced by the binding of a ligand. In the case of an enzyme, this may result in catalytic activation.

References

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.

Daggett, V.: **Long timescale simulations.** *Curr. Opin. Struct. Biol.* 2000, **10**:160–164.

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

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

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

Ringe, D. and Petsko, G.A.: **Mapping protein dynamics by X-ray diffraction.** *Prog. Biophys. Mol. Biol.* 1985, **45**:197–235.

Wall, M.E. *et al.*: **Large-scale shape changes in proteins and macromolecular complexes.** *Annu. Rev. Phys. Chem.* 2000, **51**:355–380.

©2004 New Science Press Ltd

2

From Structure to Function

There are many levels of protein function, ranging from atomic reorganizations to changes in the development of an organism, but all of them involve binding to other molecules, large and small. Sometimes this specific molecular recognition is the sole biochemical function of a protein, but in other cases the protein also promotes a chemical transformation in the molecule that it binds. This chapter looks first at how the structural features described in Chapter 1 dictate the ability of proteins to recognize specifically and bind a wide variety of ligands. The second part of the chapter looks at how these structural features dictate the ability of proteins to catalyze the wide variety of chemical transformations on which life depends.

2-0 Overview: The Structural Basis of Protein Function
2-1 Recognition, Complementarity and Active Sites
2-2 Flexibility and Protein Function
2-3 Location of Binding Sites
2-4 Nature of Binding Sites
2-5 Functional Properties of Structural Proteins
2-6 Catalysis: Overview
2-7 Active-Site Geometry
2-8 Proximity and Ground-State Destabilization
2-9 Stabilization of Transition States and Exclusion of Water
2-10 Redox Reactions
2-11 Addition/Elimination, Hydrolysis and Decarboxylation
2-12 Active-Site Chemistry
2-13 Cofactors
2-14 Multi-Step Reactions
2-15 Multifunctional Enzymes
2-16 Multifunctional Enzymes with Tunnels

2-0 Overview: The Structural Basis of Protein Function

There are many levels of protein function

It is a fundamental axiom of biology that the three-dimensional structure of a protein determines its function. Understanding function through structure is a primary goal of structural biology. But this is not always simple, partly because a biologically useful definition of the function of a protein requires a description at several different levels. To the biochemist, function means the biochemical role of an individual protein: if it is an enzyme, function refers to the reaction catalyzed; if it is a signaling protein or a transport protein, function refers to the interactions of the protein with other molecules in the signaling or transport pathway. To the geneticist or cell biologist, function includes these roles but will also encompass the cellular roles of the protein, as judged by the phenotype of its deletion, for example, or the pathway in which it operates. A physiologist or developmental biologist may have an even broader view of function.

We can take as an example tubulin, which not only has several cellular functions, but also has more than one biochemical function: it is an enzyme that hydrolyzes GTP, and also a structural protein that polymerizes to form stiff hollow tubes. In the cell, it forms a network of microtubules growing out of the centrosome (Figure 2-1), creating a system of tracks along which proteins, vesicles and organelles can be moved from one part of the cell to another. In a dividing cell it forms the mitotic spindle that segregates the chromosomes equally into the two daughter cells. In certain motile eukaryotic cells it forms the cilia and flagella that provide propulsion or sweep fluid over the cell surface. Tubulin is a protein whose functions cannot be condensed into a single sentence. This is likely to be the case for most gene products, especially in higher organisms.

Not surprisingly, biochemical function is generally the easiest to deduce from sequence and structure, although in some cases it is possible to go further. In the age of genomics, function will be derived in a partly empirical way from many different techniques employed together, augmented by comparative sequence analysis across genomes and the recognition of functional motifs in both the primary and tertiary structure. We shall illustrate in Chapter 4 how this operates in some selected cases. In this chapter and the next, we outline the general principles that have been experimentally established about the relationship of structure to the biochemical function of proteins.

There are four fundamental biochemical functions of proteins

We illustrated in Chapter 1 four biochemical functions of proteins: binding, catalysis, switching, and as structural elements (see Figure 1-1). The most fundamental of these is binding, which underlies all the other biochemical functions of proteins. Enzymes must bind *substrates*, as well as cofactors that contribute to catalysis and regulatory molecules that either activate or inhibit them. Structural proteins are, at their simplest, assemblages of a single type of protein molecule bound together for strength or toughness; in more complex cases they bind to other types of molecules to form specialized structures such as the actin-based intestinal microvilli or the spectrin-based mesh that underlies the red blood cell membrane and helps maintain its integrity as the cells are swept round the body. Protein switches such as the *GTPases* (see Figure 2-1) depend on both binding and catalytic functions of proteins: their switching properties rely fundamentally on the binding and the hydrolysis of GTP, which they catalyze. They must also bind the molecules with which they interact when GTP is hydrolyzed plus the regulatory molecules that activate GTP hydrolysis and that exchange GDP for GTP to enable the cycle to start again.

References

Alberts, B. *et al.*: *Molecular Biology of the Cell* 4th ed. Chapter 3 (Garland Publishing, New York, 2002).

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

Nogales, E.: **Structural insights into microtubule function.** *Annu. Rev. Biochem.* 2000, **69**:277–302.

©2004 New Science Press Ltd

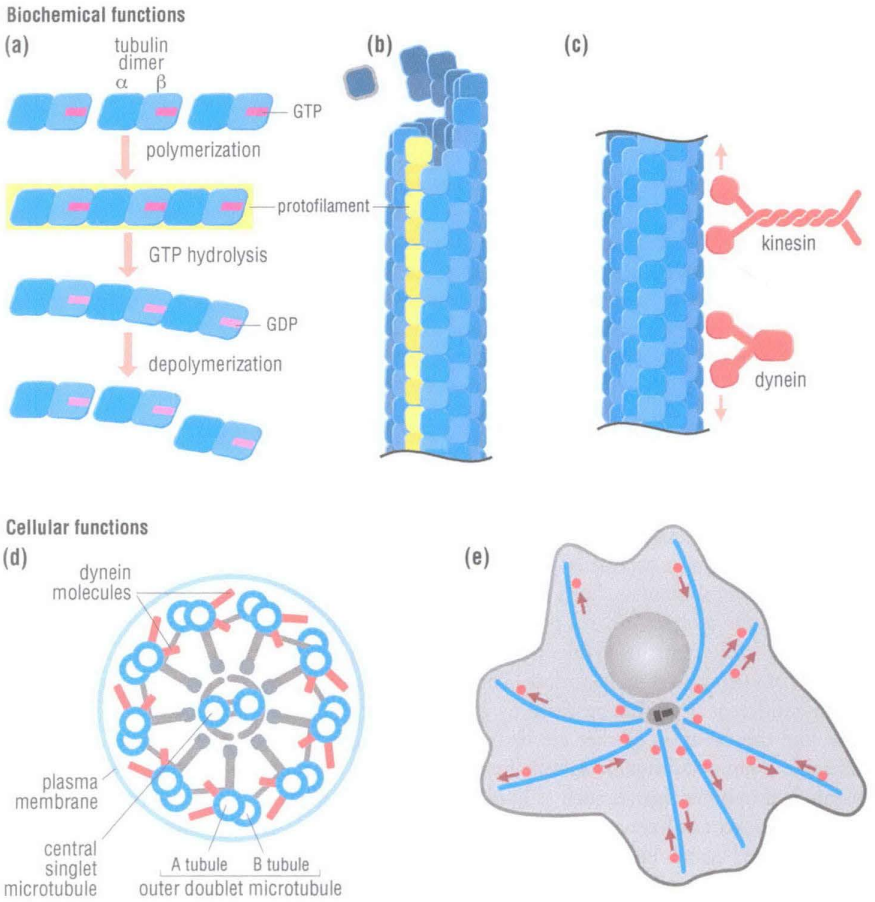


Figure 2-1 The functions of tubulin (a) The biochemical functions of tubulin include binding of tubulin monomers to each other to form a polymeric protofilament, a process that is reversed by the hydrolysis of bound GTP to GDP. Tubulin-catalyzed hydrolysis of GTP acts as a switching mechanism, in that protofilaments in the GDP form rapidly depolymerize unless the concentration of free tubulin is very high or other proteins stabilize them. **(b)** This nucleotide-dependent mechanism is used by the cell to control the assembly and disassembly of the protofilaments and the more complex structures built from them. These structures include microtubules, which consist of 13 protofilaments arranged as a hollow tube (here shown in growing phase). **(c)** Binding to motor proteins such as kinesin or dynein allows the microtubules to form molecular machines in which these motor proteins "walk" along microtubules in a particular direction, powered by ATP hydrolysis. **(d, e)** The functions of these machines are defined at the cellular level. Assemblies of microtubules, motor proteins and other microtubule-associated proteins form the flagella that propel sperm, for example (d); microtubules and associated motor proteins also form a network of "tracks" on which vesicles are moved around in cells (e). The "role" of tubulin thus encompasses both biochemical and cellular functions. The individual functions of proteins work in concert to produce the exquisite machinery that allows a cell, and ultimately a multicellular organism, to grow and survive. The anti-cancer drug taxol blocks one essential cellular function of tubulin. It binds to the polymerized protein, preventing the disassembly of microtubules that must occur during cell division.

We start this chapter, therefore, with the surface properties of proteins that determine where and how they bind to the other molecules with which they interact, whether these are small-molecule *ligands* or other macromolecules, and whether they bind stably, as oligomeric complexes, or dynamically, as in intracellular signaling pathways. And we shall see how they create the specialized microenvironments that promote the specific interactions required for catalysis.

We then explore in detail the structural basis for enzyme catalysis, and how the special chemical properties of some of the charged amino acids described in Chapter 1 (see section 1-1) make a critical contribution to enzyme action. We shall see that protein flexibility plays an essential part in the binding interactions of proteins and their catalytic actions.

Protein switches, and the more dramatic conformational changes that proteins can undergo in response to ligand binding, are described in Chapter 3, where we discuss the ways in which the activities of proteins are regulated.

©2004 New Science Press Ltd

2-1 Recognition, Complementarity and Active Sites

Protein functions such as molecular recognition and catalysis depend on complementarity

The functions of all proteins, whether signaling or transport or catalysis, depend on the ability to bind other molecules, or **ligands**. The ligand that is bound may be a small molecule or a macromolecule, and binding is usually very specific. Ligand binding involves the formation of noncovalent interactions between ligand and protein surface; these are the same types of bonds that are involved in stabilizing folded proteins (see section 1-4) and in interactions between protein subunits (see section 1-19). Specificity arises from the complementarity of shape and charge distribution between the ligand and its binding site on the protein surface (Figure 2-2), and from the distribution of donors and acceptors of hydrogen bonds. Changes in the conformation of a protein may accompany binding or be necessary for binding to occur. Alternatively, even a small change in the structure of a ligand or protein can abolish binding.

Molecular recognition depends on specialized microenvironments that result from protein tertiary structure

Specific binding occurs at sites on the protein that provide the complementarity for the ligand. These are called **ligand-binding sites** if their sole function is molecular recognition (the ligand may be as small as a proton or as large as another macromolecule) or **active sites** if they promote chemical catalysis. Such sites are formed as a consequence of the three-dimensional structure of the protein. When a polypeptide sequence folds into a compact three-dimensional structure, it creates internal cavities where the side-chain packing is not perfect, and also pockets or clefts of various sizes on its surface. These regions can have a microscopic environment that is quite different from that provided by the bulk solution around the protein. If the residues that line the cavity or pocket are hydrophobic, for example, the environment inside can resemble a nonpolar organic solvent more than it does water, enabling the protein to bind highly hydrophobic ligands such as lipids. If the residues all have, say, a negative charge, the cavity or pocket can have a very strong local electrostatic field, which would enable it to bind highly charged ligands like a calcium ion, as occurs in ion-transport proteins. Such arrangements can occur even if at first sight they seem to be unfavorable energetically—for instance, crowding a number of like charges together—because the enormous number of other, favorable interactions throughout the rest of the structure can more than make up for a small number of unfavorable ones in one place.

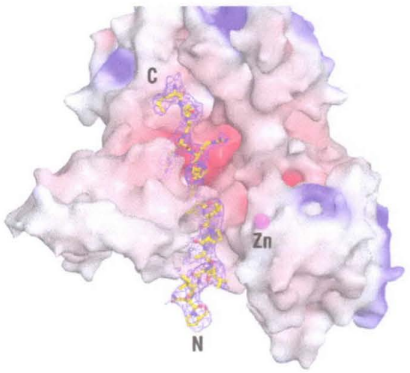


Figure 2-2 Substrate binding to anthrax toxin lethal factor Lethal factor (LF) is a component of anthrax toxin that acts as a protease to cut mitogen-activated protein kinase kinase (MAPKK-2), thereby blocking the cell cycle. This figure shows part of the surface of LF colored by charge (red, negative; blue, positive), with the model of the MAPKK-2 amino-terminal peptide shown in ball-and-stick representation. Where the model or map would be hidden by the protein surface, the surface is rendered as translucent. The active-site cleft of LF is complementary in shape and charge distribution to the substrate. Taken from Pannifer, A.D. *et al.*: *Nature* 2001, **414**:229–233. Graphic kindly provided by Robert Liddington.

Specialized microenvironments at binding sites contribute to catalysis

Most enzymes operate, in part, through *general acid-base catalysis*, in which protons are transferred between donating or accepting atoms on the substrate and key basic and acidic side chains in the enzyme active site (we discuss this in detail later, in section 2-12). Proton transfer can be promoted in two ways by specialized microenvironments on the surface of proteins.

In one case, a strong electrostatic field is produced in which acids and bases are close to each other but cannot react with each other. For instance, the close juxtaposition of an acidic side chain such as glutamic acid and a basic one such as lysine is possible. In aqueous solution at neutral pH a carboxylic acid would give up its proton to the more basic amino group, forming a carboxylate–ammonium ion charge pair. Indeed, such interactions are used to stabilize protein structure (see Figure 1-10, salt bridge). But a folded protein can also position these two

Definitions

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

ligand: small molecule or macromolecule that recognizes and binds to a specific site on a macromolecule.

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

©2004 New Science Press Ltd

residues in reasonable proximity, yet provide microenvironments around each one that make it unfavorable for them to exchange a proton. This occurs in many enzyme active sites; for example, in aspartate aminotransferase, a lysine and an aspartic acid are both involved in binding the pyridoxal phosphate cofactor, but in the environment of the active site their proton affinities are adjusted so that they do not transfer a proton between them. Thus, sites can be created that have both reasonably strong acids and reasonably strong bases in them, which is very difficult to achieve in free solution but very useful for general acid-base catalysis.

The other type of environment is one in which the affinity of a functional group for protons has been altered dramatically. Placing two lysine side chains close to one another will lower the proton affinity of both of them, producing a stronger acid. At physiological pH the two side chains will exist as a mixture of protonated and unprotonated states, enabling one of them to function as an acid-base catalyst (Figure 2-3). Analogously, burying a single lysine side chain in a hydrophobic pocket or cavity will lower its affinity for protons so that it will tend to be unprotonated at neutral pH, a state in which it is more reactive to certain chemical processes. One example of this effect is the enzyme pyruvate decarboxylase, where the active site is hydrophobic. Model compounds that mimic the substrates of this enzyme react thousands of times faster in non-polar solvents than they do in water. The enzyme favors the reaction by providing a nonpolar environment that destabilizes the substrate, which is negatively charged, and stabilizes the product, which is uncharged.

Once again, it is important to note that both of these situations are energetically unfavorable, but the energy cost is offset by favorable interactions elsewhere in the protein. In addition, long-range electrostatic effects in proteins can create a microenvironment in which the affinity of an isolated side chain for a proton can be perturbed. The significance of this effect is to allow a particular ionization state to predominate so that it can be exploited in the chemical reaction.

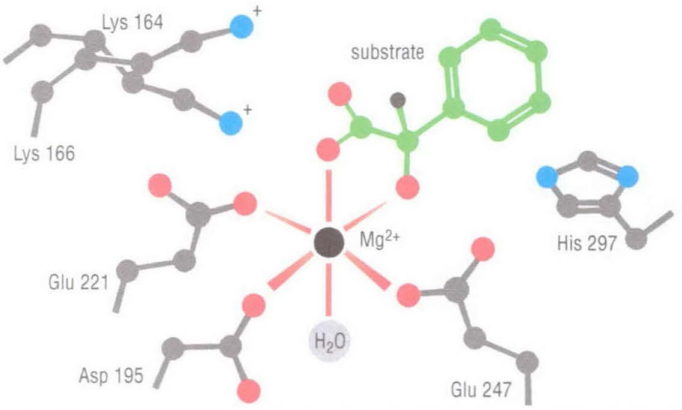


Figure 2-3 Schematic of the active site of mandelate racemase showing substrate bound Lysine 164 is located very close to the catalytically important residue lysine 166, shown at the top left of the figure. The proximity of these two positive charges lowers the proton affinity of both of them, making lysine 166 a better proton shuttle for the metal-bound substrate. Landro, J.A. *et al.*: *Biochemistry* 1994, **33**:635–643.

References

Badger, J. *et al.*: **Structural analysis of antiviral agents that interact with the capsid of human rhinoviruses.** *Proteins* 1989, **6**:1–19.

Highbarger, L.A. *et al.*: **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.

Landro, J.A. *et al.*: **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.

Pannifer, A.D. *et al.*: **Crystal structure of the anthrax lethal factor.** *Nature* 2001, **414**:229-233.

Ringe, D.: **What makes a binding site a binding site?** *Curr. Opin. Struct. Biol.* 1995, **5**:825–829.

©2004 New Science Press Ltd

2-2 Flexibility and Protein Function

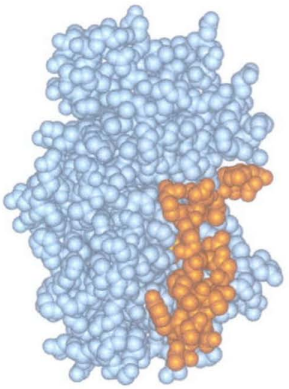


Figure 2-4 Tight fit between a protein and its ligand A space-filling representation of the catalytic domain of protein kinase A (blue) bound to a peptide analog (orange) of its natural substrate shows the snug fit between protein and ligand, achieved by mutual adjustments made by the two molecules. (PDB 1atp)

The flexibility of tertiary structure allows proteins to adapt to their ligands

In order to bind a ligand specifically a protein must have, or be able to form, a binding site whose stereochemistry, charge configuration, and potential hydrogen-bond-forming groups are complementary to those of the ligand. The classic analogy is that of a key fitting into a lock. This analogy holds for many proteins and ligands: more than one key will sometimes fit into a lock but most other keys will not, and usually only the right key will open the lock. Similarly, more than one ligand will sometimes fit into a binding site on the surface of a protein, but most other ligands will not, and usually only the right ligand will produce the "right" biological function. However, the lock-and-key analogy implies rigidity of the protein (the lock) and of the ligand (the key). In reality, both proteins and the ligands that bind to them are naturally flexible, so the classic view has been augmented by a model of **induced fit**: during binding, each can adjust its structure to the presence of the other and the protein can be said to "snuggle" around the ligand, optimizing interactions between them (Figure 2-4). More fundamentally, this give and take on the part of protein and ligand is essential for the biochemical activities of proteins.

Such conformational changes are allowed because of the inherent flexibility of proteins. In undergoing conformational changes, the protein is responding to changes in the balance between the forces that hold the tertiary structure together and the new interaction forces provided by association with a ligand. It is because both the ligand and the molecule to which it binds are flexible that many drugs that do not obviously resemble the biological ligand are nonetheless able to bind tightly to a ligand-binding site (Figure 2-5). Both the drug and the site can adjust somewhat to accommodate the difference in shape of the small molecule, provided that enough favorable interactions can be made to overcome any energetic cost of the adjustments.

Protein flexibility is essential for biochemical function

Although some proteins serve only to bind ligand, in most cases binding is followed by some action. That action may be the chemical transformation of the ligand (catalysis), a conformational change in the protein, translocation of the protein to another part of the cell, transport

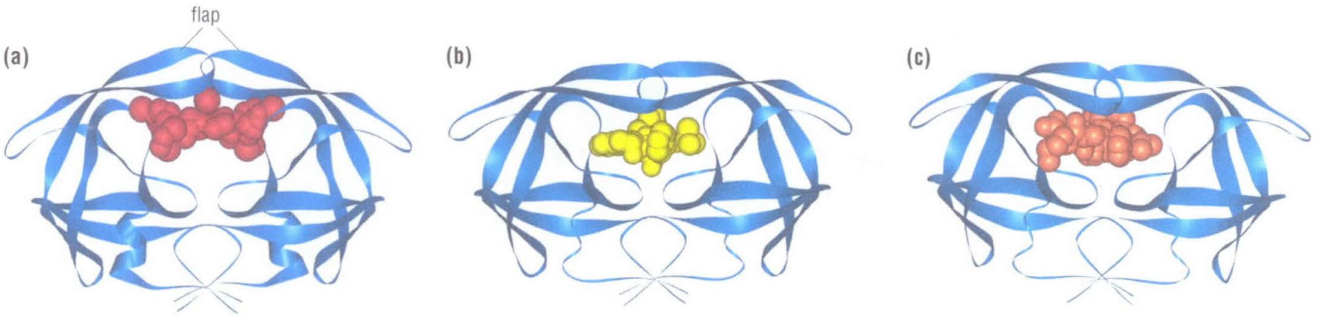


Figure 2-5 HIV protease, an enzyme from the virus that causes AIDS, bound to three different inhibitors The anti-viral action of some drugs used in AIDS therapy is based on their ability to bind to the active site of viral protease and inhibit the enzyme. The protease inhibitors haloperidol **(a)** and crixivan **(b)** are shown, with a peptide analog **(c)** of the natural substrate also shown bound to the enzyme. Each inhibitor clearly has a quite different structure and two of them (a, b) are not peptides, yet all bind tightly to the active site and induce closure of a flap that covers it, a conformational change that also occurs with the natural substrate. (PDB 1aid, 1hsg, 1a8k)

Definitions

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.

References

Ding, X., *et al.*: **Direct structural observation of an acyl-enzyme intermediate in the hydrolysis of an ester substrate by elastase.** *Biochemistry* 1994, **33**:9285–9293.

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

Koshland, D.E. Jr. *et al.*: **Comparison of experimental binding data and theoretical models in proteins containing subunits.** *Biochemistry* 1966, **5**:365–385.

Rasmussen, B.F. *et al.*: **Crystalline ribonuclease A loses function below the dynamical transition at 220 K.** *Nature* 1992, **357**:423–424.

Rutenber, E. *et al.*: **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.

Tsou, C.L.: **Active site flexibility in enzyme catalysis.** *Ann. NY Acad. Sci.* 1998, **864**:1–8.

Wrba, A. *et al.*: **Extremely thermostable D-glyceralde-**

©2004 New Science Press Ltd

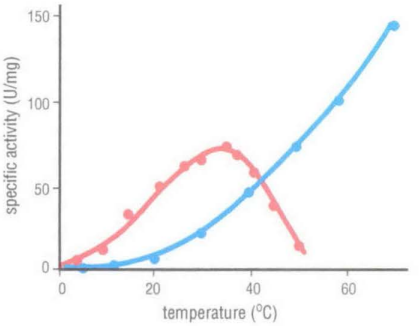


Figure 2-6 Differences in the temperature dependence of the specific activity of D-glyceraldehyde-3-phosphate dehydrogenase (GAPDH) from two organisms The red line shows the increase in enzyme activity as the temperature rises for the enzyme from yeast, a mesophile (with optimal growth rate at about 35 °C). This is as expected because most reaction rates increase with increasing temperature. Above 35 °C, however, the activity drops precipitously as the protein becomes thermally denatured. The blue line shows the behavior of GAPDH from an extreme thermophile, *Thermotoga maritima* (whose optimal growth rate is at about 85 °C). At room temperature the activity of the thermophilic enzyme is less than that of the mesophilic enzyme. One reason for this could be that the enzyme is more rigid at this temperature than the mesophilic enzyme. At temperatures where the mesophilic enzyme begins to denature, the thermophilic enzyme is still active and continues to be so beyond 70 °C. The denaturation temperature for this enzyme is above 85 °C. Wrba, A. *et al.*: *Biochemistry* 1990, **29**:7584–7592.

of the ligand, alteration of the properties of the ligand if it is a macromolecule, or some combination of these. In all cases, the structure of the protein must be flexible enough that the net free energy released by binding and/or chemical transformation of the ligand can drive the required changes in the protein's structure or properties. Binding can range from weak (dissociation constant $K_d \sim 10^{-3}$ M) to extremely strong ($K_d \sim 10^{-12}$ M or even tighter). In the case of an enzyme, the protein must undergo a series of adjustments in order to allow binding, response to changes in ligand structure and subsequent release of ligand. Yet, if the protein is too flexible, neither specific recognition nor specific action can occur. For example, adding small quantities of a denaturant such as urea or the detergent sodium dodecyl sulfate (SDS) to a protein will often cause it to become less specific for the ligands it binds; adding a bit more will greatly reduce or abolish biochemical activity, and these effects can occur long before the structure completely unfolds.

Conversely, observations on mutant enzymes that are stable at higher temperatures than normal and have been shown to be more rigid, have shown that, at least for some such proteins, rigidifying the structure abolishes function. Comparisons of enzymes from organisms that live at normal temperatures and from extreme thermophiles also show that the thermophilic enzymes are less active than the normal ones at lower temperatures (Figure 2-6). The exact mechanism of these effects is still under investigation, but it is possible that stability at higher temperatures has been achieved at the expense of flexibility at lower ones, and that the proper balance between flexibility and rigidity is necessary for many, if not most, proteins.

Protein flexibility is a natural consequence of the weak forces that hold the tertiary structure together; because the free energies of these interactions are close to the kinetic energy available at ordinary temperatures (this energy is sometimes denoted as kT), such interactions are frequently breaking and reforming under physiological conditions. The marginally stable nature of proteins also helps make them easier to degrade, and we discuss in Chapter 3 how degradation enables the protein complement of a cell to be regulated.

The degree of flexibility varies in proteins with different functions

Not all proteins are equally flexible. Although some very small rearrangements in atomic positions always occur on ligand binding, a number of proteins behave as though they were relatively rigid. Many of these are extracellular proteins, and their rigidity may help them survive in the more hostile environment outside the cell. Other proteins undergo very large shape changes when the correct ligand binds (Figure 2-7): we discuss these large ligand-induced conformational changes in Chapter 3, in the context of ligands that regulate protein function.

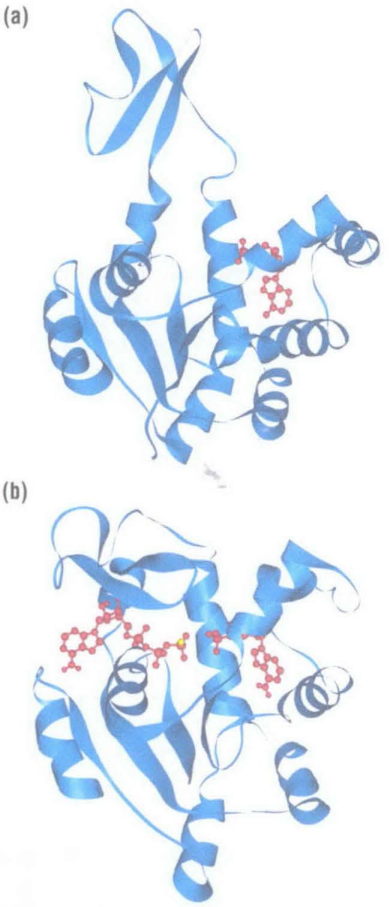


Figure 2-7 Example of a large conformational change The enzyme adenylate kinase can adopt either an open or closed conformation depending on which substrates are bound. **(a)** In the presence of AMP alone, no conformational change occurs. **(b)** On binding of the *cosubstrate* ATP, here in the form of the analog AMPPNP, a large rearrangement occurs that closes much of the active site. (PDB 2ak3, 1ank)

hyde-3-phosphate dehydrogenase from the eubacterium *Thermotoga maritima*. *Biochemistry* 1990, **29**:7584–7592.

©2004 New Science Press Ltd

2-3 Location of Binding Sites

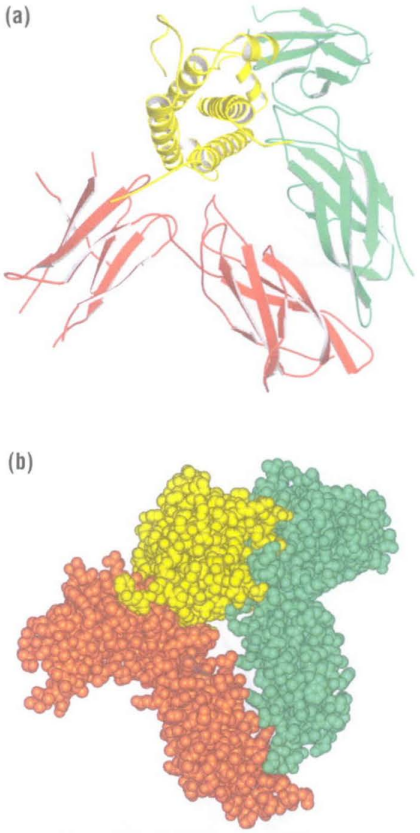


Figure 2-8 The complex between human growth hormone and two molecules of its receptor (a) Ribbon diagram of the complex. Two different protein–protein interfaces can be made by one molecule of growth hormone (yellow) with two identical receptor molecules (orange and green). **(b)** A space-filling model of the complex shows the tight fit at both interfaces. Kossiakoff, A.A. and De Vos, A.M.: *Adv. Protein Chem.* 1998, **52**:67–108.

Binding sites for macromolecules on a protein's surface can be concave, convex, or flat

The specific recognition of a macromolecule by a protein usually involves interactions over a large contiguous surface area (hundreds of square Ångstroms) or over several discrete binding regions (Figure 2-8). A macromolecule will make many points of contact with the protein's surface; these add up to provide a great deal of binding energy, so a binding site for a macromolecule can occur, in theory, anywhere on a protein's surface. Consequently, accurate prediction of a macromolecule-binding site on the surface of an experimentally determined or modeled protein structure is difficult. The most frequently observed sites are protruding loops or large cavities because these provide specific shape complementarity, but relatively flat binding sites are also found. Many binding sites for RNA or DNA on proteins are protruding loops or alpha helices that fit into the major and minor grooves of the nucleic acid (Figures 2-9a, b). These protrusions do not have any obvious common features that make them identifiable from a simple examination of the protein's structure, and most do not even have the same structure in the unbound protein and when ligand is bound.

Binding sites for small ligands are clefts, pockets or cavities

Many important biological ligands are small molecules: examples are the substrates for enzyme catalysis; cofactors that bind to the active sites of enzymes and contribute to catalysis; and *allosteric effectors*, which bind at sites remote from the active site yet modulate enzyme activity. (These will be discussed later, in sections 2-6 to 2-16.) Such small-molecule ligands bind at depressions on the protein surface, except in certain cases when they are buried within the protein's interior. Deep binding pockets allow the protein to envelop the ligand and thus use complementarity of shape to provide specificity (see Figure 2-2). They hinder access of water

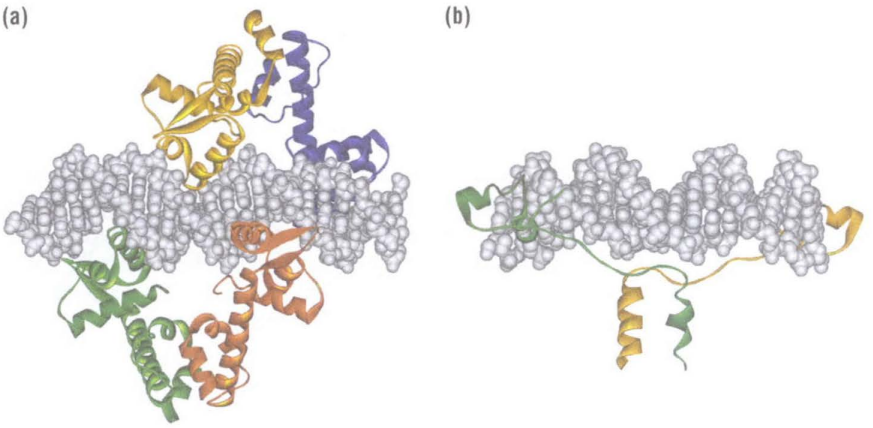


Figure 2-9 Two protein–DNA complexes (a) The complex between the bacterial diphtheria toxin gene repressor protein and the *tox* operator DNA sequence to which it binds. The repressor is a natural homodimer and two dimers bind to the pseudo-symmetrical operator sequence. The DNA sequence is recognized in the major groove by a helix in a helix-turn-helix motif, a feature often seen in DNA-binding proteins. **(b)** Structure of the complex between the eukaryotic Gal4 transcription factor and DNA. The interaction occurs in the major groove again, but this time the recognition unit contains a loop of chain that is stabilized by a cluster of zinc ions (not shown). Marmorstein, R. *et al.*: *Nature* 1992, **356**:408–414.

Definitions

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.

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.

©2004 New Science Press Ltd

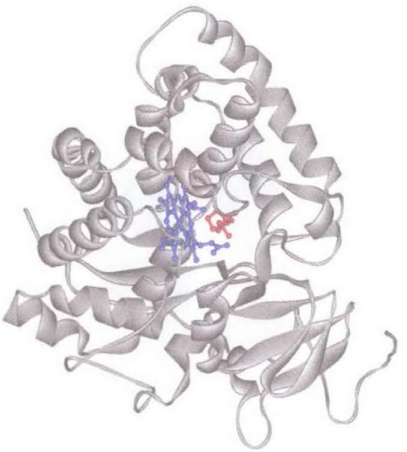


Figure 2-10 Structure of bacterial cytochrome P450 with its substrate camphor bound The active site of this enzyme contains a catalytic heme group (purple) most of which is completely buried inside the protein. There is no obvious route from the exterior to the active-site pocket in the average structure. Fluctuations in the structure must open a transient path or paths for the substrate camphor (red) to bind. Poulos, T.L. *et al.*: *J. Mol. Biol.* 1987, **195**:687–700.

to the bound ligand, which can be important for many enzyme reactions. Clefts or cavities can easily provide unusual microenvironments. And they enable even a small molecule to have enough contact points to bind strongly if that is needed. This characteristic of ligand-binding sites means that they can often be identified even in the structure of an unliganded protein: one looks for a large cleft or pocket on the protein surface or, if none is obvious, an internal **cavity** large enough to accommodate the ligand.

Binding in an interior cavity requires that the ligand diffuses through the protein structure within a reasonable time frame. Protein structural flexibility can allow such penetration, even for large substrates (Figure 2-10). It is not known for certain whether there are multiple pathways by which a ligand can "worm" its way through to a buried cavity, but computer simulations suggest that there can be. Another possibility is that the enzyme could exist in an open and closed state, with substrate binding to the open form triggering a conformational change (domain or flap closure) that sequesters the active site.

Catalytic sites often occur at domain and subunit interfaces

Enzyme active sites and most receptor binding sites for small ligands are found at generally predictable locations on protein surfaces. If a protein has more than one **structural domain**, then the catalytic site will nearly always be found at the interface between two of them, or all of them. If the protein is composed of more than one subunit, then the active site will often be found at an intersubunit interface. And if both conditions apply, then the active site will usually be located at a site corresponding to both an interdomain and an intersubunit interface (Figure 2-11).

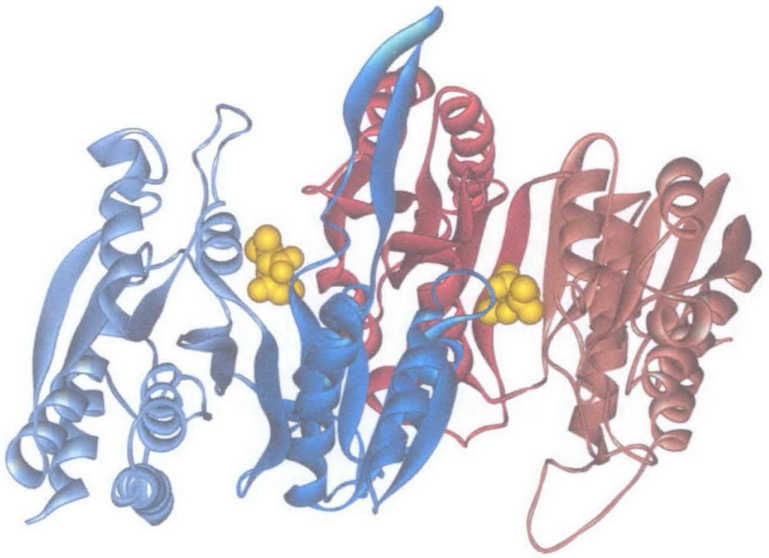


Figure 2-11 Structure of the dimeric bacterial enzyme 3-isopropylmalate dehydrogenase The two active sites are indicated by the presence of the bound cofactor NADPH (yellow). Each site occurs at an interface between the two subunits (blue/light-blue and red/brown) of the enzyme and also at an interface between the two domains of each subunit (blue and light-blue; red and brown). Imada, K. *et al.*: *J. Mol. Biol.* 1991, **222**:725–738.

References

Imada, K. *et al.*: **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.

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.

Marmorstein, R. *et al.*: **DNA recognition by GAL4: structure of a protein-DNA complex.** *Nature* 1992, **356**:408–414.

Poulos, T.L. *et al.*: **High-resolution crystal structure of cytochrome p450 cam.** *J. Mol. Biol.* 1987, **195**:687–700.

Ringe, D.: **What makes a binding site a binding site?** *Curr. Opin. Struct. Biol.* 1995, **5**:825–829.

©2004 New Science Press Ltd

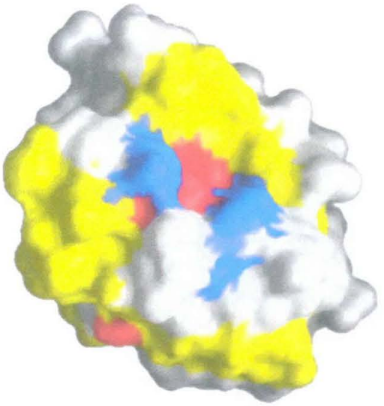


Figure 2-12 Surface view of the heme-binding pocket of cytochrome c6, with hydrophobic residues indicated in yellow The area around the heme (red) is very nonpolar because this protein must bind to another protein via this site to form an electron-transport complex involving the heme. The blue area indicates the presence of two positively charged residues important for heme binding. Graphic kindly provided by P. Roesch, Beissinger, M. *et al.*: *EMBO J.* 1998, **2**:27–36.

Binding sites generally have a higher than average amount of exposed hydrophobic surface

Protein surfaces are never completely polar as there are always isolated nonpolar groups in contact with the solvent. Ligand-binding sites, however, are generally distinguished by a much higher than average amount of exposed hydrophobic surface area: they are sites where nonpolar groups tend to be clustered on the protein surface, and this physical-chemical characteristic can sometimes be used to recognize them (Figure 2-12).

Binding sites for small molecules are usually concave and partly hydrophobic

Large hydrophobic areas on the surface of a protein lead to self-association and indeed they are the basis for oligomerization (see section 1-20). Hydrophobic ligand-binding sites do not, however, lead to oligomerization of the protein because they are usually too small and too concave to allow the protein to self-associate. But they will readily associate with a small-molecule ligand that they can interact with more favorably than with the water that covers them. Thus the combination of a high degree of concavity plus considerable exposed hydrophobic surface is reasonably diagnostic of a binding site for a small molecule on the surface of the protein.

Weak interactions can lead to an easy exchange of partners

As well as the large hydrophobic areas that lead to stable oligomerization, many proteins have smaller hydrophobic patches that are important in more transient protein–protein interactions. These occur, for example, in associations between the components of signal transduction pathways which must form and dissociate according to need. These hydrophobic patches are generally not only smaller but also less hydrophobic than those involved in oligomerization, reflecting the need for the two partners to exist independently in the aqueous environment of the cell. Because the favorable interactions at the interface between such proteins are often relatively few, and each individual interaction at the interface is weak, some of them, perhaps many, may be broken at any given time. Thus, given an existing protein–protein complex, and other proteins that can make weak interactions of a similar total energy with the partners in the complex, the dissociated components of the complex can combine with the other proteins to form new associations. This process is termed **partner swapping** and is the basis for many of the dynamic protein associations needed for signal transduction pathways.

In signal transduction, proteins such as kinases, phosphatases, and G-protein effectors are targeted to other proteins by specialized modules such as the SH2 domain, which recognizes phosphotyrosine-containing peptides on other proteins. For example, many hormones, growth factors, and immune regulators signal to the cell nucleus through STAT (signal transducer and activator of transcription) molecules, which bind to phosphotyrosine on the receptor tails via their SH2 domains. The STAT molecules are then themselves phosphorylated, and their SH2 domains dissociate from the receptor and bind to the phosphotyrosine on the other STAT

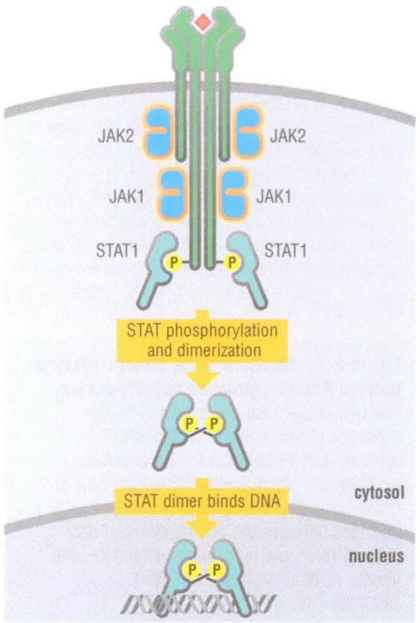


Figure 2-13 Partner swapping in a signaling pathway Binding of a signaling molecule to its receptor induces tyrosine phosphorylation by JAK kinase molecules of the cytoplasmic domains of the receptor, which then bind STAT molecules. This leads to tyrosine phosphorylation of the STAT molecules which dissociate from the receptor and instead bind through the same domains to the phosphotyrosine on the other STAT molecule, thereby forming dimers which migrate to the nucleus and bind to DNA, activating genes with various cellular effects.

Definitions

affinity: the tightness of a protein–ligand complex.

anisotropic: behaving differently in different directions; dependent on geometry and direction.

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.

partner swapping: exchange of one protein for another in multiprotein complexes.

References

Beissinger, M. *et al.*: **Solution structure of cytochrome c6 from the thermophilic cyanobacterium *Synechococcus elongatus*.** *EMBO J.* 1998, **2**:27–36.

Bennett, M.J. *et al.*: **Domain swapping: entangling alliances between proteins.** *Proc. Natl Acad. Sci. USA* 1994, **91**:3127–3131.

Bourne, Y. *et al.*: **Crystal structure of the cell cycle-regulatory protein suc1 reveals a beta-hinge conformational switch.** *Proc. Natl Acad. Sci. USA* 1995,

©2004 New Science Press Ltd

molecule to form an active signaling dimer (Figure 2-13). Such partner swapping is easy because the interactions between the STAT and the receptor are weak and each possible complex has about the same energy. However, because the STAT molecules become bound to one another by two SH2–phosphotyrosine interactions instead of one, as in the initial STAT–receptor interaction, the STAT dimer once formed is relatively stable and this association is unlikely to be reversed.

Domain swapping has been found in viral coat proteins (Figure 2-14) and signal transduction proteins. The inactive form of the enzyme PAK1 protein kinase is a domain-swapped dimer in which a regulatory domain from each monomer inhibits the active site of the other monomer. When an activator protein, such as Cdc42 in its GTP-bound form, binds to this regulatory domain, it relieves the domain swapping and frees the active site.

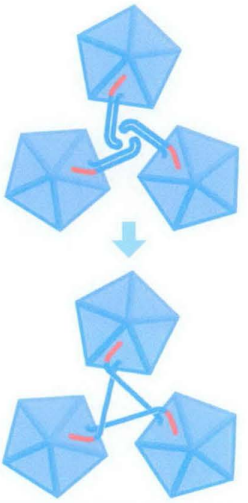


Figure 2-14 Domain swapping in the papilloma virus capsid protein Schematic representation of the conformational rearrangement of the carboxy-terminal arm of the papillomavirus capsid protein required for the stable assembly of the virus coat. On the structure of isolated capsid protein (top), the carboxy-terminal region of the polypeptide (red) is folded back and interacts with the rest of the protein. On the full-size virus particle, the arm "invades" the adjacent subunit on the surface of the virus, and makes a similar interaction with the other molecule. The domain swapping stabilizes trimers of the coat protein.

Displacement of water also drives binding events

A protein in solution is completely surrounded by water (see Figure 1-25). Some of these water molecules will interact more or less tightly with the protein surface. In fact, it is generally accepted that at least a single layer of bound water molecules should be considered an integral part of a protein structure. However, in order for a ligand which is itself surrounded by water molecules to bind to a solvated protein, both water layers must be disrupted and, at least partially, displaced. Thus, the protein and the ligand would exchange a layer of waters for favorable interactions with each other, and the enthalpic cost of releasing the surface waters can be balanced by the favorable enthalpy of the new interactions as well as by the hydrogen bonds the water can make with other solvent molecules.

The relationship between these energetic contributions is not simple. Although it might seem that the free energy of a water hydrogen bond to a protein group would be comparable to that of a hydrogen bond with another water molecule, the difference in enthalpy could be either positive or negative, depending on the microenvironment on the protein surface. For example, in an environment of reduced polarity, a hydrogen bond between a water molecule and a serine side chain could be stronger than in aqueous solution because the electrostatic attraction is greater in a medium of low polarity. Further, it is likely that some protein-bound waters gain entropy when they are displaced. Thus, the free energy of ligand binding will depend on the tightness of the water interactions with the protein surface. Although there are many potential binding sites on the irregular, largely polar protein surface, the sites where ligands actually bind will be those where favorable interactions can occur and where bound solvent can also be displaced.

Contributions to binding affinity can sometimes be distinguished from contributions to binding specificity

Although the displacement of water may provide some of the energetic driving force for ligand binding, from a practical point of view one often needs to distinguish between contributions to the **affinity**, or strength, of such binding and contributions to ligand specificity. It is generally accepted that the affinity between a protein and its ligand is chiefly due to hydrophobic interactions, which are non-directional, whereas specificity of binding is chiefly due to **anisotropic**, or directional, forces such as hydrogen bonding (Figure 2-15). The evidence for this view comes largely from measurements of ligand affinities between mutant proteins and altered ligands. However, the relative contributions of the different types of interactions to specificity and affinity will clearly vary from case to case, so conclusions drawn from mutational analysis may be misleading in the absence of a thorough investigation of the structure of the mutant with bound ligand and a thermodynamic analysis of the changes in binding energy.

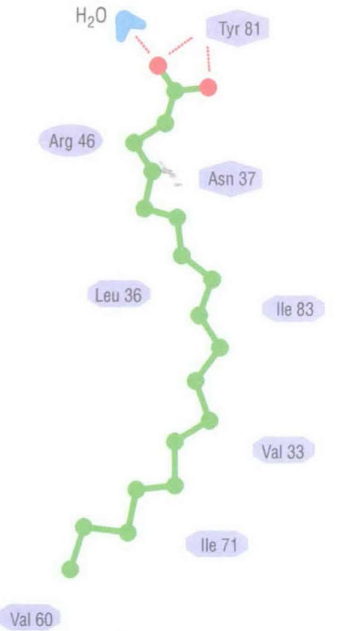


Figure 2-15 Ligand binding involving hydrophobic and hydrogen-bond interactions The binding of the lipid oleate (green) to the maize lipid-transport protein nsLTP involves mainly hydrophobic interactions of uncharged polar or nonpolar residues with the lipid tail, and hydrogen bonds (red dotted lines) to the charged head group.

92:10232–10236.

Darnell, J.E. Jr.: **STATs and gene regulation.** *Science* 1997, **277**:1630–1635.

Han, G.W. *et al.*: **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.

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

Liu, Y. *et al.*: **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.

Modis, Y. *et al.*: **Atomic model of the papillomavirus capsid.** *EMBO J.* 2002, **21**:4754–4762.

Ringe, D.: **What makes a binding site a binding site?** *Curr. Opin. Struct. Biol.* 1995, **5**:825–829.

Szwajkajzer, D. and Carey, J.: **Molecular constraints on ligand-binding affinity and specificity.** *Biopolymers* 1997, **44**:181–198.

©2004 New Science Press Ltd

2-5 Functional Properties of Structural Proteins

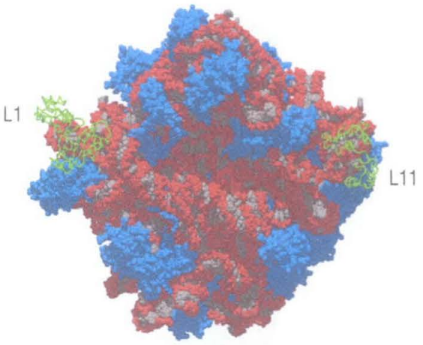


Figure 2-16 Structure of the 50S (large) subunit of the bacterial ribosome The ribosomal RNA is shown in red and grey, while most of the structural proteins are shown as blue space-filling models. Two of the proteins (L1 and L11) are shown as backbone diagrams in green because their positions are not known and have been arrived at by modeling. Graphic kindly provided by Poul Nissen and Thomas Steitz.

Proteins as frameworks, connectors and scaffolds

Cells are not just structureless bags filled with freely floating molecules. All cells are surrounded by a protein-reinforced membrane; some have a cell wall that is primarily protein and carbohydrate. Internal structures within the cell also are made up of particular structural proteins that confer shape, strength and flexibility on these cellular structures. In some cases, structural proteins are assisted by DNA, RNA, lipid and carbohydrate molecules; in other cases the structure is built up from a large number of different proteins. The ribosome, for example, has over a hundred different protein components that stabilize the folded form of the ribosomal RNA, which provides the catalytic function (Figure 2-16).

There is a dynamic character to many of these subcellular structures. In some, for example muscle, the structure itself can change shape in response to external stimuli; in other cases, the structural proteins provide a framework for dynamic processes to occur, driven by other types of proteins. The actin filaments that provide the tracks along which many proteins and protein complexes run within the cell are an example of such a framework. Some structural proteins form temporary structures that are then destroyed when no longer needed. Fibrinogen, the primary component of blood clots, is such a protein: it polymerizes to form a dense fibrous mass but dissolves when the wound is healed. The inappropriate formation of such structures can underlie serious human diseases. Unregulated clotting can lead to fatal thromboses; aggregation of beta-amyloid protein, which is not normally a structural protein but becomes one after proteolytic cleavage, produces the amyloid fibrils associated with Alzheimer's disease.

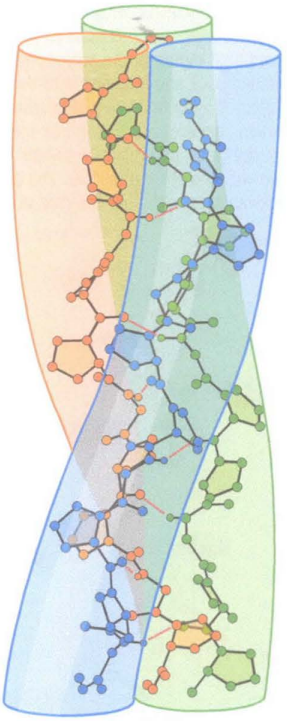


Figure 2-17 Structure of collagen The collagen triple helix is a coiled-coil structure in which each of the three protein chains is made up of repeating GlyXY sequences, where X is often proline (in the example shown here, Y is also proline). The hydrophobic nature of this repeat results in a set of regularly spaced hydrophobic sites along each chain; these complementary sites plus interchain hydrogen bonds (red dotted lines) hold the triple helix together. In collagen fibers, multiple triple helices are aligned end-to-end and side-by-side in a regular fashion, producing the light and dark bands observed when collagen fibers are imaged in an electron microscope.

Some structural proteins only form stable assemblies

Many of the structural components of cells and organisms are designed to be permanent: they are neither altered nor destroyed during the lifetime of the organism. Such assemblies can be constructed from proteins alone, as in the case of silk, collagen, elastin or keratin, or the coat proteins of a virus (see Figure 1-74k), or from protein plus some other component, as in the case of cartilage, which is composed of protein plus carbohydrate.

There are two ways in which these structures can be stabilized. One is by protein–protein interactions alone. Although such interactions are noncovalent and thus relatively weak in energetic terms, there can be enormous numbers of them in any given structural assembly, which produces an overall stabilization equal to many covalent bonds. To make the assembly take on a particular shape, the interactions also need to be specific. One way to achieve a large number of specific weak interactions is to place the complementary surfaces on simple repeating secondary structure elements such as alpha helices and beta strands. Stable assemblies are thus often coiled coils of long helices or stacks of beta sheets, and to achieve regularity the sequences of the component proteins are often made up of simple repeating motifs, which can be recognized easily in sequence analysis. Collagen, the fibrous component of tendons, is one example of such a structure: the basic component of collagen is a triple helix of three protein chains made up of repeating GlyXY sequences (Figure 2-17). Variability at the third position in the repeat can impart special local properties. Silk is an example of a stack of beta sheets (see Figure 1-1). The other way of stabilizing structural assemblies is by covalent cross-linking of their protein components. For example, the structure of collagen is stabilized further by covalent cross-linking. This is initiated by the enzyme lysyl oxidase which converts the terminal amino groups in the side chains of lysine residues to peptidyl aldehydes. These then undergo a number of uncatalyzed reactions that cross-link the chains. Elastin, the protein that gives lung tissues their elastic properties, is stabilized by the same cross-linking reaction: the cross-links hold the

Definitions

scaffold protein: a protein that serves as a platform onto which other proteins assemble to form functional complexes.

©2004 New Science Press Ltd

individual polypeptide chains in a rubber-like network. Deficiency in lysyl oxidase activity is associated with the genetic diseases Type IX Ehlers-Danlos syndrome and Menkes syndrome, which are characterized by loss of stability and elasticity in the connective tissues.

Some catalytic proteins can also have a structural role

The regular assemblies formed by structural proteins very often have cellular functions that require time-dependent changes in shape or conformation or some other property of the assembly. These changes are often brought about by changes in the structure of a single component of the multicomponent assembly, which in turn derive from the energy released by a chemical process. This may be merely the binding of another protein or small molecule—as small as a proton if the change is pH-driven as in the conformational changes required for fusion of viruses to cell membranes—but is usually a protein-catalyzed chemical transformation such as the hydrolysis of ATP.

Muscle is an example of such a multicomponent assembly, in which the dynamic component is the motor protein myosin II. Muscle is composed chiefly of interdigitating actin filaments and myosin filaments, held together by other structural proteins. Myosin II is a homodimer of two main subunits, each with a long helical "tail" and a head domain that contains both a binding site for an actin filament and a catalytic site. The two tails form an extended coiled coil which associates with those of other myosin molecules to form the thick myosin filament, from which the heads protrude at regular intervals. Hydrolysis of ATP at the catalytic site produces a conformational change in the head domain that results in a change in the position of the head on the actin filament; this motion causes the myosin and actin filaments to slide against each other, thus causing the muscle fiber to contract. Myosin II is thus a structural protein (it forms filaments), a catalytic protein (an ATPase), and a motor protein involved in cell motility.

Some structural proteins serve as scaffolds

When signals external to the cell are transmitted inside the cell, the kinases, phosphatases and transcription factors that make up the intracellular signal transduction pathways must find each other in order to carry out the sequence of reactions that transmit the signal. Diffusion of a protein across a typical eukaryotic cell can occur in a few milliseconds, but for two randomly distributed proteins the time to interact by diffusion will be much longer, implying that specific recruitment is needed. Sometimes this recruitment occurs by localizing one or more of the components to the cell membrane, but in other cases specific structural proteins serve as **scaffold proteins** onto which the other members of the pathway assemble, forming a signaling complex.

Mitogen-activated protein kinase (MAPK) signal transduction cascades form such molecular assemblies within cells. The spatial organization for these is provided by scaffold proteins. Yeast Ste5p was the first MAPK cascade scaffold protein to be described. Ste5p selectively tethers the MAP kinases MAPKKK, MAPKK and MAPK, which act sequentially in the pheromone-stimulated yeast mating pathway. Recent work indicates that Ste5p is not a passive scaffold but plays a direct part in the activation of the MAPKKK by interacting with a heterotrimeric G protein and another kinase. This activation event requires the formation of an active Ste5p oligomer and proper recruitment of Ste5p to the cell cortex, forming a "signalosome" linked to a G protein (Figure 2-18). Many other such scaffold proteins have been identified. They can often be detected in sequence analysis because they frequently contain repeats of known protein–protein interaction domains such as RING-H2, WW and WD40 domains. These will be discussed in detail in Chapter 3.

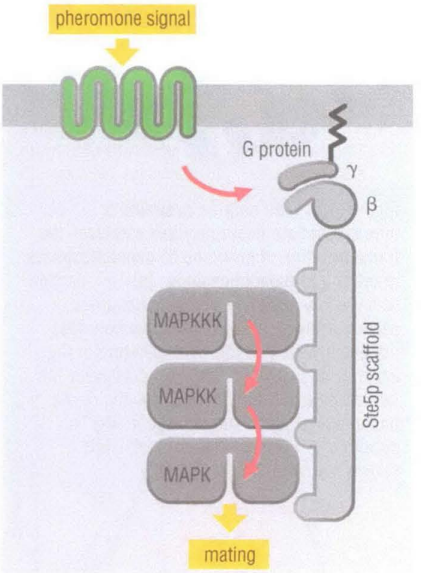


Figure 2-18 The Ste5p scaffold Ste5p is essential for the yeast response to mating pheromone and binds the sequential kinase components of a *mitogen-activated protein kinase* (MAPK) cascade: these are the MAPKKK, the MAPKK and the MAPK. Pheromone stimulation releases the *G-protein* components Gβγ, which recruit Ste5p and a cytoplasmic kinase (not shown) to the plasma membrane, activating the MAPK cascade. Pathway activation is thought to be coordinated with the conversion of a less active closed form of Ste5p containing a protected RING-H2 domain into an active Ste5p dimer that can bind to Gβγ and form a multimeric scaffold lattice upon which the kinase components of the MAPK cascade can assemble.

References

Ban, N. *et al.*: **The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution.** *Science* 2000, **289**:905–920.

Berisio, R., *et al.*: **Recent progress on collagen triple helix structure, stability and assembly.** *Curr. Pharm. Des.* 2002, **9**:107–116.

Elion, E.A.: **The Ste5p scaffold.** *J. Cell Sci.* 2001, **114**:3967–3978.

Houdusse, A. *et al.*: **Three conformational states of scallop myosin S1.** *Proc. Natl Acad. Sci. USA* 2000, **97**:11238–11243.

Rochet, J.C. and Lansbury, P.T. Jr.: **Amyloid fibrillogenesis: themes and variations.** *Curr. Opin. Struct. Biol.* 2000 **10**:60–68.

Sette, C. *et al.*: **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.

©2004 New Science Press Ltd

2-6 Catalysis: Overview

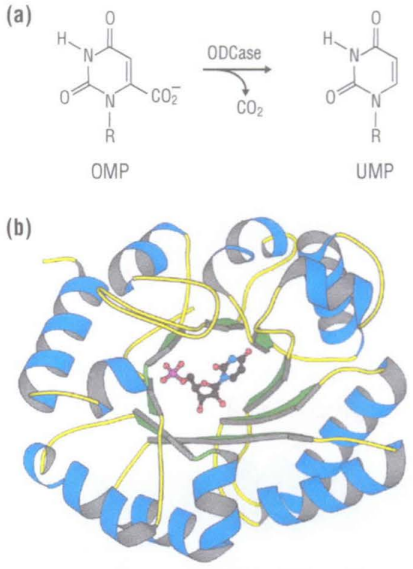


Figure 2-19 The enzyme orotidine 5′-monophosphate decarboxylase catalyzes the transformation of orotidine 5′-monophosphate to uridine 5′-monophosphate (a) The reaction catalyzed by orotidine 5′-monophosphate decarboxylase (ODCase). This reaction does not occur readily at room temperature in the absence of a catalyst. ODCase accelerates the rate of the reaction 10^{17}-fold. R = the ribose phosphate group of the nucleotide. **(b)** The structure of ODCase. The product, UMP, is shown bound in the active site.

Catalysts accelerate the rate of a chemical reaction without changing its overall equilibrium

In the ligand-binding events we have discussed so far, the ligand is unchanged on binding the protein. We now turn to the actions of enzymes—biological catalysts whose ligands undergo chemical transformation during metabolism, biosynthesis and in many of the signaling and motor activities of cells. A **catalyst** is any substance that accelerates the rate of a chemical reaction without itself becoming permanently altered in the process. The word permanently is important, because, as we shall see, enzymes often undergo transient alteration in both their conformation and covalent structure while catalyzing reactions; but by the end of the overall reaction they are always restored to their original condition. Although most enzymes are proteins, a number of RNA molecules also function as catalysts (termed *ribozymes*), including the ribosomal RNA on the large subunit of the ribosome, which has been shown to be the catalyst of peptide-bond formation during protein synthesis.

Enzymes can be extraordinarily efficient catalysts. The transformation of orotidine 5′-monophosphate (OMP) to uridine 5′-monophosphate (UMP) (Figure 2-19a) is a key decarboxylation reaction in purine biosynthesis. In free solution, the uncatalyzed reaction is estimated to take approximately 78 million years to go halfway to completion (obviously, this is an estimated, not measured, rate). But when the same reaction is catalyzed by the enzyme orotidine 5′-monophosphate decarboxylase (OMP decarboxylase or ODCase; Figure 2-19b), it is completed in less than a second—an increase in reaction rate of 10^{17}-fold. OMP decarboxylase is not the only enzyme that is known to accelerate a reaction by more than a billion fold; catalytic accelerations of this magnitude or more are common in biology (Figure 2-20), and are the reason that living organisms can exist at moderate temperatures. In the absence of an efficient catalyst, a reaction such as OMP decarboxylation would require very high temperatures to proceed at a measurable rate. Since primordial enzymes are unlikely to have been very efficient—the assumption being that catalytic proficiency requires billions of years of evolution by trial-and-error through random mutations—it is thought that the earliest living organisms were probably extreme thermophiles, unicellular microorganisms that live at high temperatures. Such organisms can still be found today, in hot springs such as those in America's Yellowstone National Park.

All chemical transformations, even those that are not obviously reversible, actually represent equilibria that involve forward and backward reactions. Each has a characteristic **equilibrium constant**, which is the ratio of the concentrations of the products (multiplied together) to the concentrations of the reactants (multiplied together) that are present when **equilibrium** is attained. The equilibrium constant can also be determined by the relative rates of the forward and reverse processes. A reaction whose forward rate is faster than its reverse will have more product than reactant at equilibrium, whereas one where the back reaction is faster will have a preponderance of reactant. As the chemical transformations in the back reaction just follow the same path but in the opposite direction, any catalyst, including an enzyme, will speed up the forward and the back reaction by the same amount, leaving the relative concentrations of product(s) and reactant(s) at equilibrium unchanged. Thus, catalysis accelerates the rate at which equilibrium is reached, but not the final equilibrium concentrations (or the equilibrium constant).

Catalysis usually requires more than one factor

For over a century biochemists have been fascinated by the enormous catalytic power of enzymes. Rate accelerations of up to 10^{17}, as with ODCase, are so far beyond those achievable

Definitions

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.

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

catalyst: a substance that accelerates the rate of a reaction without itself being permanently altered.

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.

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.

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

intermediate: a species that forms transiently along the path from substrate to product.

substrate: the molecule that is transformed in a reaction.

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**.

©2004 New Science Press Ltd

in ordinary laboratory reactions that it was long thought that there must be some special, undreamed-of catalytic principle that enzymes use. But after decades of intense research, including the determination of the atomic structure of many enzyme–substrate complexes, it has become apparent that there is no unique secret to enzymatic catalysis. Enzymes have at their disposal a variety of simple contributory factors to help them attain huge rate accelerations; each enzyme uses a combination of several of these, but the particular combination, and especially the relative importance of each one, vary from enzyme to enzyme.

Some of these factors are physical in nature: they depend on the structure and physical properties of the enzyme and on the ability to orient the ligand very precisely relative to catalytic residues in the active site. Others are chemical: they involve the chemical properties of the amino acids (and cofactors, if any) that make up the enzyme, including their ability to stabilize unstable chemical species by weak interactions, their ability to polarize bonds, and their ability to form covalent adducts. There are some enzymes in which a single catalytic factor predominates, but these are the exception. Similar enzymes from different organisms may also have subtle differences in the balance of contributions of the different factors. It is the net contribution of many simple effects, not any one special feature unique to living organisms, that accounts for the extraordinary power of enzymes to speed up reactions, thereby providing the chemistry necessary to sustain life at ordinary temperatures and pressures.

Catalysis is reducing the activation-energy barrier to a reaction

For a chemical reaction to occur, the reactants (which in enzymology are generally called **substrates**) must undergo rearrangements in stereochemistry, charge configuration, and covalent structure. If the reactant is a stable compound, there will be a free-energy barrier to such transformations, even if the product of the reaction is more stable. The higher this barrier, the slower the reaction and the more difficult the chemical step is to achieve. The energy required to overcome this barrier is known as the **activation energy**, and the barrier is called the **activation-energy barrier**, because, in order to react, the substrate must attain a higher free-energy state and in this state is said to be activated. We shall see examples of how activation is achieved in later sections. The **transition state** is the highest point in free energy on the reaction pathway from substrate to product; it is the top of the activation-energy barrier (see TS_u in Figure 2-21). Chemically, it is a species that exists for about the time required for a single atomic vibration to occur (about 10^{-15} s). In the transition state, the making or breaking of chemical bonds in the reaction is not yet complete: the atoms are "in flight". The stereochemistry and charge configuration of the transition state is thus likely to be quite different from that of either the substrate or the product, although it may resemble one more than it does the other.

The activation free-energy barrier must be overcome for the reaction to proceed. One way to speed up the reaction is therefore to lower the barrier, either by raising the free energy, or **ground state**, of the reactant—or of the product in the reverse direction (all relative to some standard state). An alternative way of lowering the barrier is to lower the free energy of the transition state: this would correspond to stabilizing it, as molecules of lower energy are more stable. Still another is to cause the reaction to take a different path, one in which there may well be more free-energy "hills" than for the uncatalyzed reaction, but along which every hill is smaller. In such cases there will be local "valleys" of free energy between the hills; these metastable molecules, usually of relatively high free energy, are called the **intermediates** in the reaction. This third option is shown in Figure 2-21. Enzymes, as we shall see, employ all these catalytic tricks.

Comparison of Uncatalyzed and Catalyzed Rates for Some Enzymatic Reactions

Enzyme	Nonenzymatic rate k_{non} (s^{-1})	Enzymatic rate k_{cat} (s^{-1})	Rate acceleration k_{cat}/k_{non}
Cyclophilin	2.8×10^{-2}	1.3×10^{4}	4.6×10^{5}
Carbonic anhydrase	1.3×10^{-1}	10^{6}	7.7×10^{6}
Chymotrypsin	4×10^{-9}	4×10^{-2}	10^{7}
Triosephosphate isomerase	6×10^{-7}	2×10^{3}	3×10^{9}
Fumarase	2×10^{-8}	2×10^{3}	10^{11}
Adenosine deaminase	1.8×10^{10}	370	2.1×10^{12}
Urease	3×10^{-10}	3×10^{4}	10^{14}
Alkaline phosphatase	10^{-15}	10^{2}	10^{17}
ODCase	2.8×10^{-16}	39	1.4×10^{17}

Figure 2-20 Table of the uncatalyzed and catalyzed rates for some representative enzymatic reactions k is the rate constant for the reaction. Adapted from Radzicka, A. and Wolfenden, R.: *Science* 1995, **267**:90–93.

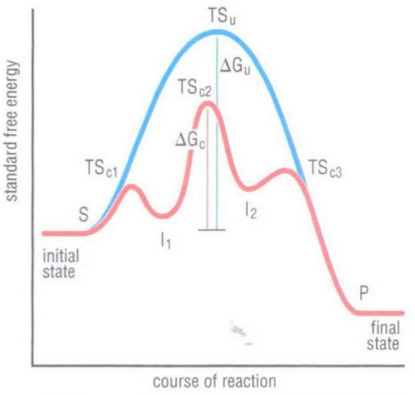


Figure 2-21 Energetics of catalysis The energy course of a hypothetical reaction from substrate S to product P can be described in terms of transition states and intermediates. For the uncatalyzed reaction (blue curve) a single transition-state barrier determines the rate at which product is formed. In the presence of a catalyst (red curve), which in this case is acting by changing the pathway of the reaction and introducing additional smaller activation-energy barriers, intermediate I_1, formed by crossing transition-state barrier TS_{c1}, leads to transition-state barrier TS_{c2}. Its free energy (ΔGc), although the highest point in the reaction, is considerably lower than the free energy (ΔGu) of the uncatalyzed transition state, TS_u. After formation of a second intermediate, I_2, a third transition state, TS_{c3}, leads to product. Because TS_{c2} is the highest transition state in the catalyzed reaction, the rate at which the reactants pass over this barrier determines the overall rate and thus it is said to be the rate-determining transition state of the catalyzed reaction. The rate-determining step of this reaction is thus the conversion of I_1 to I_2.

References

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

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

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

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

Radzicka, A. and Wolfenden, R.: **A proficient enzyme.** *Science* 1995, **267**:90–93.

Silverman, R.B.: *The Organic Chemistry of Enzyme-Catalyzed Reactions* (Academic Press, New York, 2000).

Walsh, C.: *Enzymatic Reaction Mechanisms* (Freeman, San Francisco, 1979).

©2004 New Science Press Ltd

2-7 Active-Site Geometry

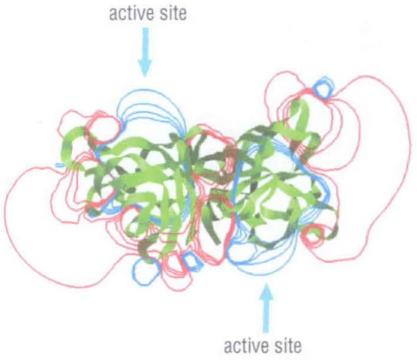


Figure 2-22 The electrostatic potential around the enzyme Cu,Zn-superoxide dismutase Red contour lines indicate net negative electrostatic potential; blue lines net positive potential. The enzyme is shown as a homodimer (green ribbons) and two active sites (one in each subunit) can be seen at the top left and bottom right of the figure where a significant concentration of positive electrostatic potential is indicated by the blue contour lines curving away from the protein surface. The negative potential elsewhere on the protein will repel the negatively charged superoxide substrate ($O_2^{-\bullet}$) and prevent non-productive binding, while the positive potential in the active site will attract it. Graphic kindly provided by Barry Honig and Emil Alexov.

Reactive groups in enzyme active sites are optimally positioned to interact with the substrate

In any enzyme-catalyzed reaction, the first step is the formation of an enzyme–substrate complex in which the substrate or substrates bind to the active site, usually noncovalently. Specificity of binding comes from the close fit of the substrate within the active-site pocket, which is due primarily to van der Waals interactions between the substrate and nonpolar groups on the enzyme, combined with complementary arrangements of polar and charged groups around the bound molecule. This fit is often so specific that even a small change in the chemical composition of the substrate will abolish binding. Enzyme–substrate dissociation constants range from about 10^{-3} M to 10^{-9} M; the lower the value the more tightly the substrate is bound. It is important that an enzyme does not hold onto its substrates or products too tightly because that would reduce its efficiency as a catalyst: the product must dissociate to allow the enzyme to bind to another substrate molecule for a new catalytic cycle.

Formation of a specific complex between the catalyst and its substrates does more than just account for the specificity of most enzymatic transformations: it also increases the probability of productive collisions between two reacting molecules. All chemical reactions face the same problem: the reacting molecules must collide in the correct orientation so that the requisite atomic orbitals can overlap to allow the appropriate bonds to be formed and broken. If we consider, in general terms, that a molecule can have a "reactive side" where the chemical changes take place and an "unreactive side" where they do not (at least immediately; the

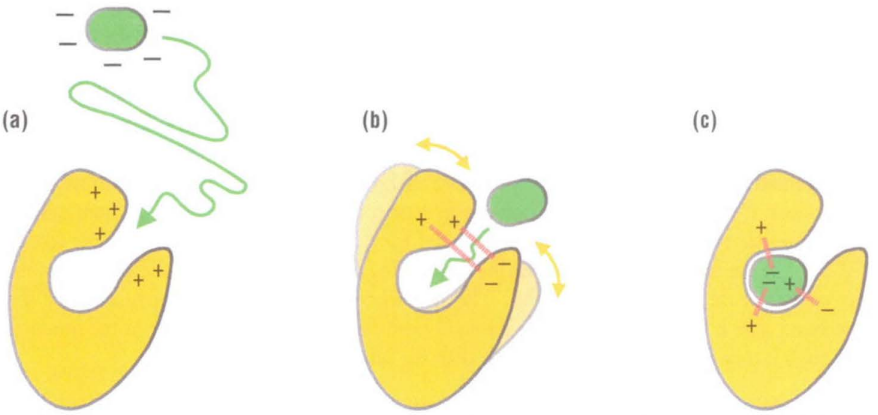


Figure 2-23 Schematic diagram showing some of the ways in which electrostatic interactions can influence the binding of a ligand to a protein **(a)** Electrostatic forces and torques can steer the ligand (green) into its binding site on the protein (shown in yellow). **(b)** Some binding sites are normally shielded from the solvent and can be kept "closed" by salt links between groups on the protein surface. If the correct substrate disrupts these salt links it can gain access to the binding site. This is known as "**gated**" **binding**. Alternatively, the dynamics of the protein may open and close such a site transiently (as indicated by the yellow arrows). **(c)** Electrostatic interactions, particularly salt links and hydrogen bonds, between ligand and protein can contribute to the affinity and specificity of binding and to the orientation of the ligand in the binding site and the structure of the complex formed. All three of these ways of exploiting electrostatic interactions can be used by a single enzyme. Adapted from Wade, R.C. *et al.*: *Proc. Natl Acad. Sci. USA* 1998, **95**:5942–5949.

Definitions

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

reaction sub-site: that part of the active site where chemistry occurs.

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

©2004 New Science Press Ltd

molecule may rearrange during the reaction), then in a simple reaction in which two molecules combine, both of them must collide reactive side-to-reactive side. Any other orientation and the collision will be non-productive. Thus, if both molecules first bind to an enzyme active site, and do so in such a way that their reactive portions are juxtaposed, the probability of a reaction is optimized. In solution, when two molecules collide but do not react they bounce off each other more or less randomly. On the enzyme, however, once the first reactive molecule has bound, it will stay there for some time, waiting for the second to come along. If that molecule does not bind productively, the first one may still remain associated with the enzyme (depending on the affinity constant) long enough for many other collisions to be tried.

In addition, enzyme active sites may even have evolved to attract their substrates so that finding the active site is not a random process. Most biological molecules are charged (so they can be retained within a cell by their insolubility in the hydrophobic membrane that surrounds it). Although active sites have exposed hydrophobic patches, the overall electrostatic field produced by the protein with all its polar and charged groups can yield an electrostatic potential with a net charge in the active-site region (Figure 2-22). It is possible that this net potential may "draw" the substrate into the oppositely charged active site, increasing the probability of productive binding (Figure 2-23a). Other ways in which electrostatics can aid in binding are illustrated in Figure 2-23b and c.

In most cases, there is a further critical factor in the facilitation of reactions by enzyme active sites. Usually, the enzyme itself supplies one or more of the chemical groups that participate in catalysis. Groups that are already part of the active-site structure in the folded protein before the substrate binds are already oriented properly for catalysis, or become so as the enzyme binds its substrate. The folding energy of the protein has already paid most, if not all, of the cost of positioning these groups, so there will be no unproductive collisions because they are in the wrong orientation. Every substrate molecule that forms the enzyme–substrate complex will therefore be exposed to an environment in which the catalytic groups are positioned correctly, relative to the substrate, for the desired reaction to take place.

In fact, the reactive portion of the substrate need not be the part that is used to hold it at the enzyme surface. An enzyme can recognize and interact with the remote parts of the substrate molecule, parts not involved in the chemistry, and use these interactions to hold and orient the substrate. This is a general principle: enzyme active sites consist of a **specificity sub-site** and a **reaction sub-site**, and in these protein groups are positioned around different parts of the substrate. In the specificity sub-site the enzyme uses polar and nonpolar groups to make weak interactions with the substrate; in the reaction sub-site other groups on the enzyme carry out the chemistry (Figure 2-24). In some cases the same amino-acid residue may participate in both specific substrate binding and catalysis. This design feature makes excellent sense. During the catalytic reaction, portions of the substrate molecule will undergo changes in geometry, charge and covalent bonding. If an enzyme had binding interactions with parts of the substrate that had to undergo rearrangement, those interactions would have to be broken before the substrate could change its structure, which could slow the enzyme down. The dual nature of enzyme active sites is being exploited in medicine and industry to design new catalysts. Amino-acid changes can often be made in the specificity sub-site of an enzyme without affecting its catalytic sub-site. So an enzyme that originally catalyzed a reaction involving positively charged substrates, for example, can sometimes be altered to perform exactly the same chemistry on new substrates that are negatively charged.

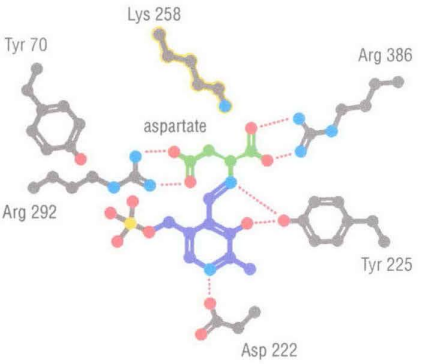


Figure 2-24 Schematic diagram of the active site of *E. coli* aspartate aminotransferase
The enzyme uses a pyridoxal phosphate (PLP) cofactor (purple) and lysine (yellow outline) to carry out chemistry. The substrate amino acid (green) reacts with the cofactor to form an adduct (as shown in this model) which then rearranges to give product. Substrate specificity for the negatively charged aspartic acid substrate is determined by the positively charged guanidino groups of arginine 386 and arginine 292, which have no catalytic role. Mutation of arginine 292 to aspartic acid produces an enzyme that prefers arginine to aspartate as a substrate. Adapted from Cronin, C.N. and Kirsch, J.F.: *Biochemistry* 1988, **27**:4572–4579; Almo, S.C. *et al.*: *Prot. Eng.* 1994, **7**:405–412.

References

Almo, S.C. *et al.*: **The structural basis for the altered substrate specificity of the R292D active site mutant of aspartate aminotransferase from *E. coli*.** *Prot. Eng.* 1994, **7**:405–412.

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.

Wade, R.C. *et al.*: **Electrostatic steering and ionic tethering in enzyme-ligand binding: insights from simulations.** *Proc. Natl Acad. Sci. USA* 1998, **95**:5942–5949.

©2004 New Science Press Ltd

2-8 Proximity and Ground-State Destabilization

Some active sites chiefly promote proximity

Binding of substrates in the correct orientation probably makes a significant contribution to the catalytic efficiency of all enzymes, but in some cases it accounts for virtually all their effectiveness. If the substrate molecules are intrinsically reactive, simply holding them close to each other in the proper orientation may be all that is needed to facilitate the appropriate chemistry. This is often referred to as the **proximity factor,** or sometimes, the **propinquity factor**.

A striking example of this factor in action is provided by the metabolic enzyme aspartate transcarbamoylase (ATCase), which promotes the condensation of carbamoyl phosphate and aspartic acid to yield carbamoyl aspartate (Figure 2-25a). The crystal structure of ATCase with an inhibitor bound to the active site has revealed the potential interactions the enzyme makes with these substrates. The inhibitor, PALA (*N*-phosphonoacetyl-L-aspartate), is a bisubstrate analog: it has features that resemble both substrates and it occupies both binding sites simultaneously (Figure 2-25b). Examination of the structure of the inhibitor–enzyme

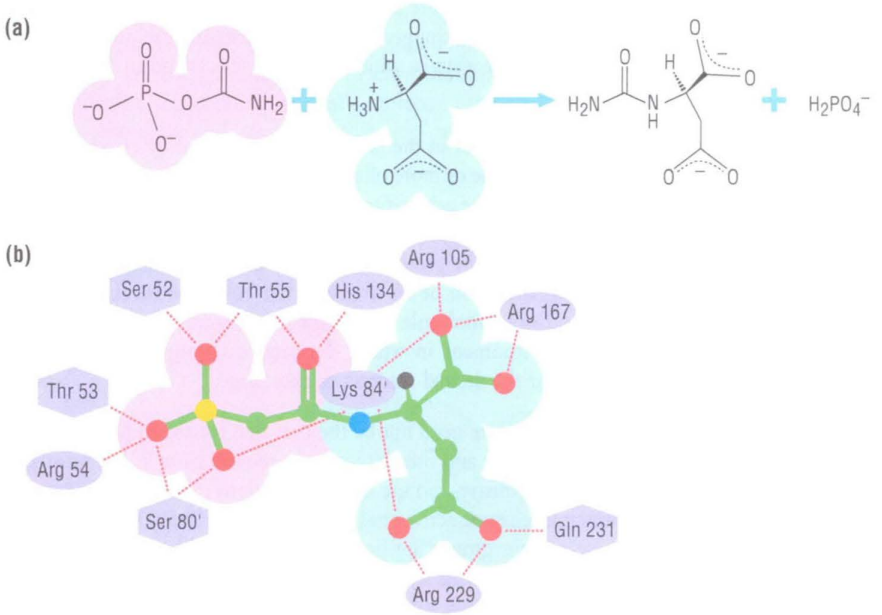


Figure 2-25 Catalysis of the reaction of carbamoyl phosphate and aspartate by the enzyme aspartate transcarbamoylase depends on holding the substrates in close proximity and correct orientation in the active site (a) The reaction catalyzed by aspartate transcarbamoylase (ATCase). Carbamoyl phosphate (shaded in pink) and aspartate (shaded in blue) undergo a condensation reaction to form *N*-carbamoyl aspartate. This is an essential step in pyrimidine biosynthesis. **(b)** Schematic diagram of the active site of ATCase with the inhibitor PALA (*N*-phosphonoacetyl-L-aspartate) (green) bound. The amino acids forming the active site and binding PALA by noncovalent bonds (red dotted lines) are represented by the purple shapes. (Ser 80′ and Lys 84′ come from an adjacent subunit to the other residues shown.) PALA resembles both carbamoyl phosphate and aspartate, as can be seen by comparison with (a), and binds to both the binding sites for these substrates in the active site. Note the absence of catalytic amino acids in a position to interact with the H_3N^+ group of the normal substrate aspartate, the group that reacts to form the bond with carbamoyl phosphate.

Definitions

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.

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

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.

©2004 New Science Press Ltd

complex shows the surprising fact that there are no chemically reactive side chains positioned anywhere near those parts of the substrates where the bond between them would be formed. The active site grips other parts of both substrates, but any chemistry comes chiefly from the reactivities of the molecules themselves. They condense because they are held in close proximity while being oriented so that the atomic orbitals that must form the new chemical bond are positioned to overlap.

Another example of a reaction that is aided primarily by proximity is provided by the enzyme chorismate mutase, which catalyzes a reaction in the pathway for the biosynthesis of aromatic amino acids. The reaction is an internal rearrangement of chorismate, of a type known as a pericyclic rearrangement. Crystal structures of the enzyme with substrate or substrate-analog inhibitors bound show that the enzyme promotes the reaction primarily by binding the substrate in an unusual "chair" conformation, which in solution would be energetically unfavorable. This conformation positions the group to be moved in an orientation that facilitates the internal rearrangement (Figure 2-26). Compounds designed to mimic this conformation are particularly effective inhibitors of the enzyme because the active site has evolved to be complementary to it.

Some active sites destabilize ground states

The bound conformation of chorismate (or of prephenate, the product of the chorismate mutase reaction) in the active site of chorismate mutase is not only a conformation that aids in catalysis by positioning the reacting atoms near one another; it is also a conformation that is much higher in free energy (less stable) than the conformation of substrate or product that normally exists in free solution. By binding the substrate in a less stable conformation (relative to some standard state used for comparison), the enzyme has started it out farther up the free-energy hill towards the transition state (see section 2-6), thereby increasing the reaction rate through **ground-state destabilization**.

The enzyme ODCase, which we encountered in section 2-6, may provide a striking example of the power of ground-state destabilization; it is possible that this factor is the major contributor to the 10^{17}-fold rate acceleration achieved by this enzyme. Crystal structures of ODCase with substrate analogs bound show that the negatively charged carboxylate group of the substrate is not interacting with any positively charged groups in the active site. Instead, it is located in a hydrophobic pocket. This unfavorable environment for an ionized group should greatly increase the free energy of the bound substrate and facilitate the transformation of the carboxylate to the neutral product, CO_2, which will have a much lower free energy in such a site.

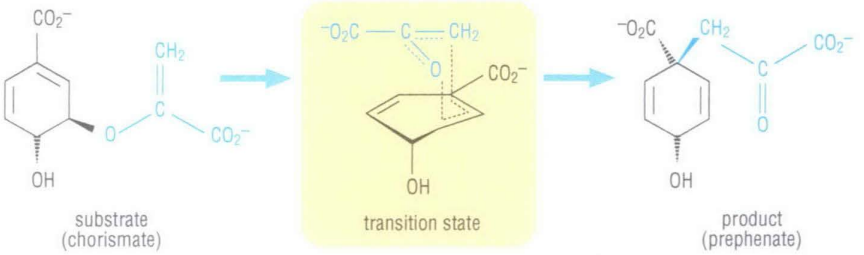


Figure 2-26 The pericyclic rearrangement of chorismate to prephenate via the proposed "chair-like" transition state When chorismate is bound to the active site of the enzyme chorismate mutase, the atoms involved in the internal rearrangement become arranged in a chair configuration. In the active site, the substrate (and, indeed, the product) are held in a conformation that resembles the single transition state for the chemical transformation (highlighted in yellow). The dotted lines indicate partial bond formation that occurs in the presumed transition state. The part of the molecule that is rearranged is indicated in blue.

References

Chook, Y.M. *et al.*: **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.

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

Krause, K.L. *et al.*: **2.5 Å structure of aspartate carbamoyltransferase complexed with the bisubstrate analog *N*-(phosphonoacetyl)-L-aspartate.** *J. Mol. Biol.* 1987, **193**:527–553.

Lau, E.Y. *et al.*: **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.

Mesecar, A.D. *et al.*: **Orbital steering in the catalytic power of enzymes: small structural changes with large catalytic consequences.** *Science* 1997, **277**:202–206.

Radzicka, A. and Wolfenden, R.: **A proficient enzyme.** *Science* 1995, **267**:90–93.

Strater, N. *et al.*: **Mechanisms of catalysis and allosteric regulation of yeast chorismate mutase from crystal structures.** *Structure* 1997, **5**:1437–1452.

Wu, N. *et al.*: **Electrostatic stress in catalysis: structure and mechanism of the enzyme orotidine monophosphate decarboxylase.** *Proc. Natl Acad. Sci. USA* 2000, **97**:2017–2022.

©2004 New Science Press Ltd

2-9 Stabilization of Transition States and Exclusion of Water

Some active sites primarily stabilize transition states

Many enzyme active sites are complementary to the transition states of the reactions they catalyze, both in stereochemistry and charge configuration, or become that way during the reaction. The differences in structure between the transition state and the ground state for a given reaction are important because it is the differential binding of the enzyme for these two states that leads to catalysis. If an enzyme binds both substrate and transition state with equal affinity, the reaction will not be facilitated. But if the transition state can be bound more tightly than the substrate, then the free-energy difference between substrate and transition state will be reduced. The more that difference is reduced, the lower the barrier to reaction, and the more likely reaction becomes (Figure 2-27). This is the definition of catalysis.

The sources of the binding energy required to reduce that free-energy difference are the weak interactions that occur in the active site between enzyme and substrate. Transition-state complementarity often involves the placement in an active site of charged groups at positions where charges of opposite sign will develop when transition states appear. If a transition state has new hydrogen-bond donors or acceptors, or positions existing ones in new places, the enzyme can also stabilize the transition state by having appropriately placed acceptors and donors. Another complementarity that can occur is purely steric. Because bonds are forming and breaking in the transition state, its geometry differs from that of the substrate, and active sites fit transition states sterically better than they do their substrates. Stabilization of the transition state facilitates the electron shifts that are part of the reaction mechanism. Citrate synthase is an example of an enzyme that has certain catalytic groups prearranged so as to stabilize the transition state, which differs considerably in its geometry from the substrate (Figure 2-28).

If an enzyme active site does not start out perfectly complementary to the transition state before substrate binds, the enzyme may undergo conformational changes that increase that complementarity. Phosphoglycerate kinase (PKA) is an example of an enzyme in which a conformational change that occurs when both substrates are bound generates new interactions that stabilize the transition state (Figure 2-29). The change in conformation enables certain amino acids of the active site, together with the enzyme's bound magnesium ion, to stabilize the oxygens of the transferring phosphoryl group to a greater extent than in either the substrate or product. By stabilizing this transition state, PKA catalyzes phosphoryl transfer.

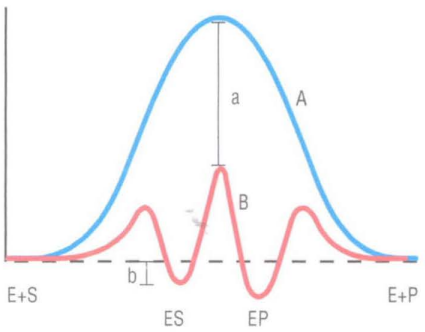


Figure 2-27 Effect of binding energy on enzyme catalysis A hypothetical case in which substrate and product are of equal energy, with a large free-energy barrier between them (A). Binding of the substrate and transition state to the enzyme can lower the free energies of these complexes by the amounts b and a respectively. If the transition state is bound more tightly than the substrate (B; a > b), which may occur, in part, through strain (destabilization) in the enzyme–substrate (ES) complex, or through the development of new interactions with the transition state, or both, the net effect will be catalysis. If, however, the enzyme binds both the substrate and the transition state with equal affinity (a = b), the barrier between the ES complex and the transition state will have the same height as the barrier for the uncatalyzed reaction, and no catalysis will occur. This is still true if the substrate and product are of unequal energies, which is the usual case. As always in discussing stabilization of ground or transition states by enzymes, external reference states must be defined.

Many active sites must protect their substrates from water, but must be accessible at the same time

Water participates in many chemical reactions, but there are other reactions that cannot proceed rapidly, or even at all, in an aqueous environment. The ability of some enzymes to shield their substrates from aqueous solvent by taking advantage of conformational changes that close off the active site from contact with bulk solvent is important in enabling them to accelerate the rates of the reactions they catalyze. There is a problem connected with these conformational shifts, however. When an active site is in its closed conformation, anything already bound is protected from water, but substrates cannot enter it, and products cannot leave it. Similarly, when it is in its open conformation, such an enzyme is not an effective catalyst. This problem is solved by having the resting enzyme exist in an open state to which substrates can bind readily, and then having substrate binding trigger the conformational changes to the closed form. The rate at which these conformational shifts occur could potentially limit the rate at which an enzyme that depends on them could operate.

Definitions

hydride ion: a hydrogen atom with an extra electron.

References

Bernstein, B.E. and Hol, W.G.: **Crystal structures of substrate and products bound to the phopshoglycerate kinase active site reveal the catalytic mechanism.** *Biochemistry* 1995, **37**:4429–4436.

Bernstein, B.E. *et al.*: **Synergistic effects of substrate-induced conformational changes in phosphoglycerate kinase activation.** *Nature* 1997, **385**:275–278.

Karpusas, M. *et al.*: **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.

©2004 New Science Press Ltd

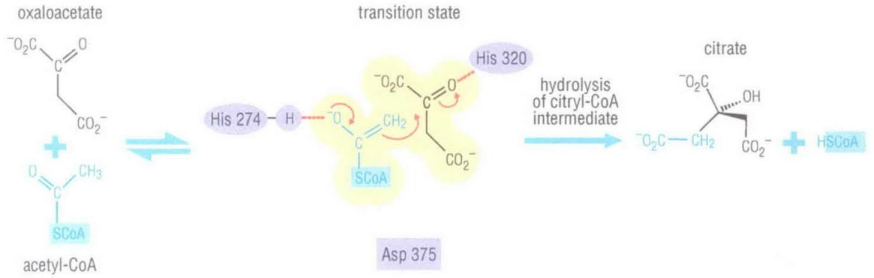


Figure 2-28 The active site of citrate synthase stabilizes a transition state with a different geometry from that of the substrate Citrate is synthesized by an aldol addition reaction between oxaloacetate and the acetyl group of acetyl-CoA. In the active site, histidine 274 and aspartic acid 375 catalyze the formation of the enol of acetyl-CoA, the rate-limiting step in the reaction. The enol is stabilized by a hydrogen bond to histidine 274. The acetyl-CoA enol then attacks the carbonyl carbon of oxaloacetate, resulting in the addition of the elements of the acetyl group at this position. Red arrows show the movement of electron pairs in making and breaking bonds. The resulting intermediate, citryl-CoA, is unstable and is hydrolyzed in the active site. Adapted from Karpusas, M. *et al.*: *Biochemistry* 1990, **29**:2213–2219.

NAD-dependent dehydrogenases transfer a **hydride ion** (H^-) during their catalytic reactions. Hydride ions are unstable in water, so the active site must be shielded from bulk solvent during hydride transfer. These enzymes usually have a flexible loop that closes down over the active site when the cosubstrate NADH or NAD^+ binds. A set of noncovalent interactions makes the closed-lid conformation more favorable energetically than the open-lid conformation when substrate is present (Figure 2-30). For many enzymes that use such a mechanism, the opening and/or closing of the lid is the rate-determining step in the reaction.

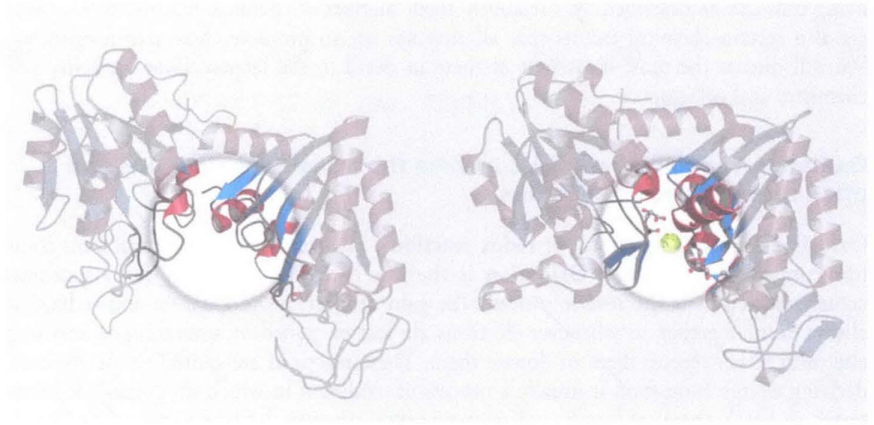


Figure 2-29 Phosphoglycerate kinase (PGK) undergoes a conformational change in its active site after substrate binds PGK catalyzes a reversible reaction in which a phosphoryl group is transferred between 1,3-bisphosphoglycerate and ADP to form 3-phosphoglycerate and ATP. Structures of unliganded PGK (left, from horse) and the ternary complex of PGK and its substrates 1,3-bisphosphoglycerate and ADP (right, from *Trypanosoma brucei*), show the change in conformation that occurs on substrate binding: the two domains shift position to close the active site and a metal-ion-binding site is created by the ligand and close proximity of previously distant residues. (The appearance of elements of secondary structure in the liganded enzyme that are absent from the unliganded enzyme may be an incidental consequence of the crystallization conditions rather than an effect of ligand binding.) Substrates are in ball-and-stick representation. The green sphere represents a magnesium ion. The active-site region is highlighted by the circle. The change in conformation increases the complementarity of the active site for the transition state of the reaction. (PDB 2pgk and 13pk)

There are three further questions one might ask about this system: what drives the motions that open and close the lid, what makes the lid swing open once a bound substrate molecule has been reduced by hydride transfer, and how fast does the lid move? The answer to the first question is straightforward: the movements of the lid are driven by the thermal motions that apply to all molecules in solution. The answer to the second question is not known for any enzyme with a similar moving part. The interactions between enzyme and substrate are not altered in any way that would obviously favor the opening of the lid. The loop of polypeptide chain that forms the lid is only held in place by a few weak interactions with the substrate and the rest of the enzyme, and these interactions will tend to break periodically at ordinary temperatures. When they do, the loop can swing open, and product is released. Whatever the mechanism, the loop does not have to move very fast. A single molecule of the enzyme glyceraldehyde-3-phosphate dehydrogenase can carry out almost 1,000 hydride transfers in a second, so the loop must move at least 10 Å every 0.001 second. That corresponds to a velocity of only 10^{-6} km per hour (a car travels at 10^2 km/h). Peptide conformational transitions in aqueous solution are known to be at least 10^4 times faster than that.

Figure 2-30 NAD-dependent lactate dehydrogenase has a mechanism for excluding water from the active site once substrates are bound Lactate dehydrogenase (LDH) catalyzes the conversion of lactate (orange circle) to pyruvate (red circle), with the accompanying reduction of NAD^+ (light-blue square) to NADH (dark-blue square). The mechanism of action of LDH involves a conformational change that closes a mobile "lid" down over the active site to exclude water when both substrates have bound. Residues in this loop make weak connections to substrate and other parts of the protein when the lid is closed. The products are released when the lid opens, probably spontaneously.

©2004 New Science Press Ltd

2-10 Redox Reactions

A relatively small number of chemical reactions account for most biological transformations

A mammalian cell produces over 10,000 different proteins, of which more than half are enzymes. These thousands of different enzymes would seem to catalyze thousands of different reactions with thousands of different substrates and products. But many of these reactions are actually similar to one another and the substrates and products are often similar as well. In fact, about three-quarters of the reactions involved in metabolism can be described by as few as four general types of chemical transformations: oxidation/reduction; addition/elimination; hydrolysis; and decarboxylation.

In catalyzing these reactions, enzymes face certain common chemical problems. Organic molecules are typically stable under physiological conditions and are replete with –C–H bonds, which are relatively unreactive. One problem, therefore, is how to activate these bonds so that chemistry can occur there under the mild conditions found in living organisms and organic molecules can be degraded to provide energy. Another is how to stabilize thermally or chemically unstable intermediates that form as these reactions proceed. Yet another is the prevention of side-reactions in a milieu containing 55 moles per liter of a reactive substance, namely water. All these tasks require highly specialized and sequestered microenvironments. We have already seen some of the physical properties and mechanisms that enzymes use to create such environments (see section 2-6). Because the chemistry of living cells can be described by a relatively small number of chemical reaction types, there are also certain chemical factors that all enzymes use to promote these transformations. We will discuss the most important of these in detail in the later sections on active-site chemistry and cofactors.

Oxidation/reduction reactions involve the transfer of electrons and often require specific cofactors

Oxidation/reduction reactions, or **redox reactions**, involve the transfer of electrons from one compound to another. **Oxidation** is the loss of electrons, usually from a carbon center; **reduction** is the reverse process, the gain of electrons. Oxidation and reduction always occur together, as whenever electrons are lost or gained by something, something else must either receive them or donate them. These reactions are central to metabolism; deriving energy from food is usually a process of oxidation in which the loss of electrons serves to break chemical bonds and release energy, whereas the biosynthesis of complex molecules usually involves reduction to form new chemical bonds. In general, these reactions require the participation of specialized organic or metal-ion containing groups called cofactors, which we discuss later in the chapter. Certain cofactors have evolved for use in certain sub-types of biological redox reactions: three of these reactions are illustrated in Figure 2-31. Oxidation and reduction at carbon–oxygen centers typically involves the nicotinamide-containing cofactors NAD (nicotinamide adenine dinucleotide) and NADP (nicotinamide adenine dinucleotide phosphate). The enzyme malate dehydrogenase in the tricarboxylic acid (TCA) cycle uses a molecule of NAD to oxidize a –C–OH group in malate to a –C=O group, producing oxaloacetate and reduced NAD (NADH) plus a proton (Figure 2-31a). Redox reactions at carbon–carbon bonds most frequently utilize the flavin-containing cofactors FAD (flavin adenine dinucleotide) or FMN (flavin mononucleotide). Another TCA-cycle enzyme, succinate dehydrogenase, converts succinate to fumarate by an oxidation that requires FAD (Figure 2-31b). Oxidation and reduction at –C–H centers and

Definitions

oxidation: the loss of electrons from an atom or molecule.

redox reactions: reactions in which oxidation and reduction occur.

reduction: the gain of electrons by an atom or molecule.

©2004 New Science Press Ltd

nitrogen-containing centers often involves cofactors that contain metal ions. One example is the enzyme cytochrome P450, which takes specific –C–H bonds in unreactive carbon compounds and inserts molecular oxygen to produce –C–OH groups, as in the biosynthetic pathway for steroid hormones such as pregnenolone (Figure 2-31c); this reaction, the biological equivalent of a propane torch, is carried out at room temperature with the aid of the iron-containing cofactor heme.

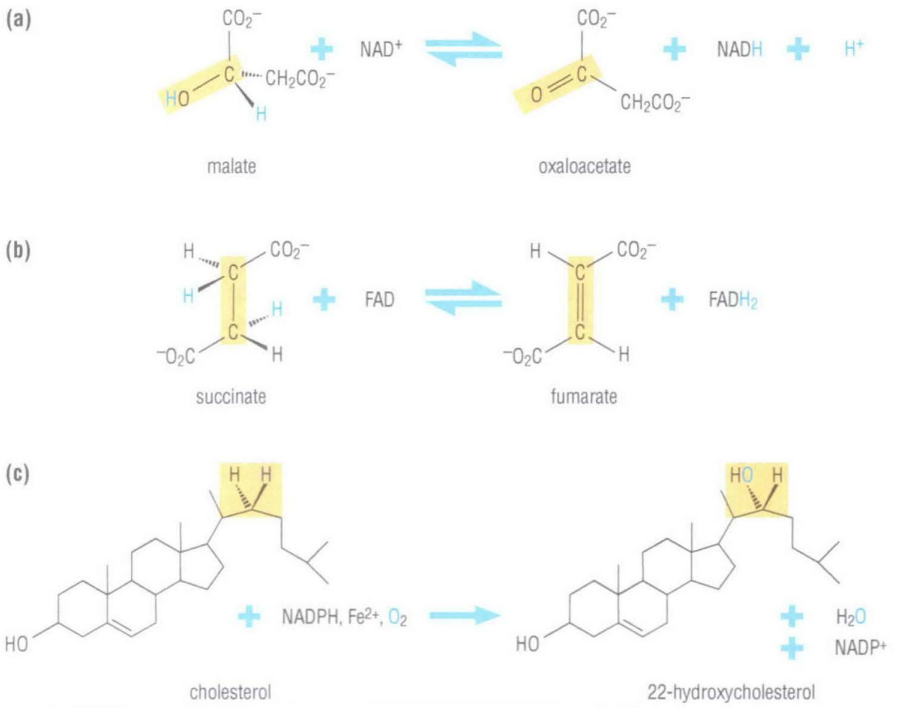


Figure 2-31 Examples of oxidation/reduction reactions (a) The oxidation of an alcohol to a ketone by NAD as illustrated by the reaction catalyzed by malate dehydrogenase. The reaction is driven by the abstraction by NAD of an electron and a hydrogen atom (a hydride ion, H^-) from the carbon center of malate at which oxidation occurs, resulting in the subsequent transformation of the C–OH bond at this center to a C=O bond. In the course of the reaction the NAD is reduced to NADH. **(b)** Oxidation of a saturated carbon–carbon bond to an unsaturated carbon–carbon bond by FAD as in the reaction catalyzed by succinate dehydrogenase. **(c)** In the first step of the pathway for the conversion of cholesterol to pregnenolone—the formation of 22-hydroxycholesterol—the iron atom in the heme cofactor of cytochrome P450 inserts an atom of oxygen into a –C–H bond. Here, oxidation is the insertion of an oxygen atom derived from molecular oxygen. In all these reactions, the sites of bond rearrangement are shaded yellow and the transferred hydrogens or oxygens in blue.

References

Silverman, R.B.: *The Organic Chemistry of Enzyme-Catalyzed Reactions* (Academic Press, New York, 2000).

©2004 New Science Press Ltd

2-11 Addition/Elimination, Hydrolysis and Decarboxylation

Addition reactions add atoms or chemical groups to double bonds, while elimination reactions remove them to form double bonds

Addition reactions transfer atoms or chemical groups to the two ends of a double bond, forming a more highly substituted single bond. Elimination reactions reverse this process, forming a new double bond. In many addition/elimination reactions involving C=C bonds, the transferred species is a molecule of water. In the reaction catalyzed by the TCA-cycle enzyme fumarase a molecule of water is added to the C=C of fumarate (OH^- to one carbon and H^+ to the other) to produce a molecule of malate (Figure 2-32a). Another type of addition reaction, called an aldol condensation (the reverse elimination is called aldol cleavage) is used to make new carbon–carbon bonds. This reaction involves addition of an activated carbon center (for example, the acyl moiety of acyl-coenzyme A) to the C=O carbonyl carbon and is the most common way of making a new C–C bond in biology. In the TCA cycle, such a reaction is catalyzed by the enzyme citrate synthase (Figure 2-32b), which we encountered in section 2-9. Breaking a C–C bond is difficult and is often done by aldol cleavage. In glycolysis, the enzyme aldolase catalyzes an aldol cleavage reaction that splits a six-carbon sugar phosphate into two three-carbon fragments.

Esters, amides and acetals are cleaved by reaction with water; their formation requires removal of water

The breaking of C–N, C–O, S–O, P–O and P–N bonds is accomplished in biology by the reaction of compounds containing such bonds with water. Such reactions are referred to as hydrolysis; hydrolytic reactions of biochemical importance are the hydrolysis of amides (C–N), esters and complex carbohydrates such as acetals (C–O), sulfate esters (S–O), phosphoesters (P–O), and phosphonamides (P–N). Degradation of biopolymers such as proteins in digestion is almost entirely a hydrolytic process. All proteases, such as trypsin, which degrade the amide bonds in proteins use water in this way (Figure 2-33a) as do the nucleases that digest DNA and RNA by hydrolyzing P–O bonds between adjacent nucleotide residues (Figure 2-33b). A P–O phosphoanhydride bond is also hydrolyzed in the conversion of ATP to ADP and

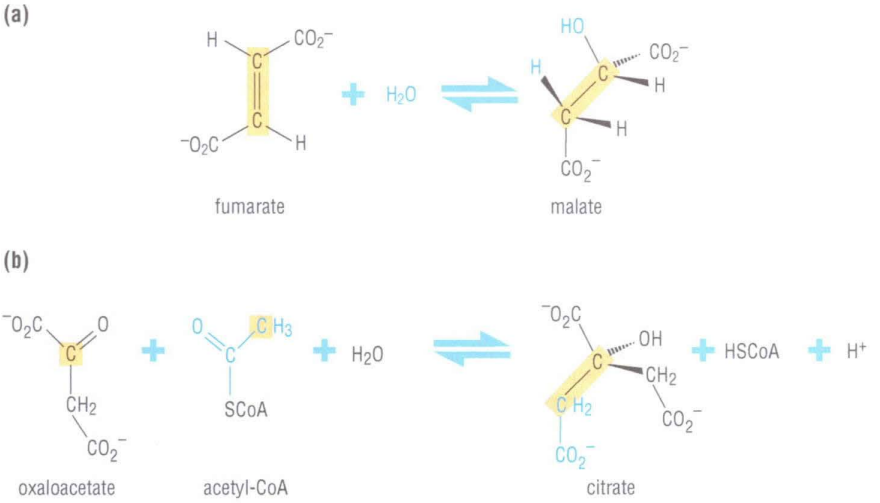


Figure 2-32 Examples of addition/elimination reactions **(a)** Addition of the elements of water across the C=C of fumarate to create the HC–COH group of malate, a reaction catalyzed by fumarase. This reaction is readily reversible, but is driven in the direction of addition rather than elimination by the high concentration of water in the cellular milieu. **(b)** Addition of the elements of acetate to the carbonyl carbon of oxaloacetate in the aldol condensation reaction catalyzed by citrate synthase. Acetate is activated by coupling it to a cofactor called coenzyme A (CoA). The aldol condensation gives an activated *S*-citryl-CoA intermediate (not illustrated here) which is then hydrolyzed to give citrate, regenerated CoA (HSCoA) and H^+.

Definitions

decarboxylation: removal of carbon dioxide from a molecule.

©2004 New Science Press Ltd

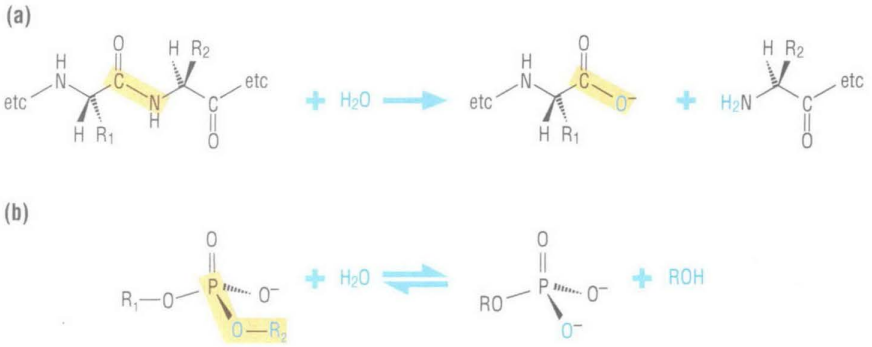


Figure 2-33 Examples of peptide and phosphoester hydrolysis (a) Cleavage of the C–N bond of a peptide involves attack by water on the carbonyl carbon, resulting in formation of a carboxylic acid and an amine. The carboxylic acid carries the –OH portion of water while the amine takes on the proton. This reaction is catalyzed by proteases, which comprise many different families. Formation of amide bonds in proteins is formally the reverse of this process, but the reaction is so unfavorable energetically that it is usually accomplished by a complex process involving several activation steps. **(b)** Breaking the P–O–R bond of a phosphate diester involves attack by water on the phosphorus atom, resulting in formation of a phosphate monoester and an alcohol. When the phosphate diester is part of the backbone of DNA, R_1 is derived from the 3′-hydroxyl of the deoxysugar (ribose) of one nucleotide and R_2 is derived from the 5′-hydroxyl of the deoxysugar of another nucleotide. This reaction is catalyzed by endonucleases such as DNase. In the reverse direction, if the alcohol is on the sugar group of one deoxyribonucleotide, loss of water produces the phosphodiester linkage of the DNA backbone (both of these groups must be activated so the reaction is driven to completion). This biosynthetic reaction is catalyzed by DNA polymerase.

inorganic phosphate, for example, a reaction central to energy metabolism in all cells. Formation of these types of bonds involves the loss of a molecule of water in the reverse of hydrolysis, which is usually called condensation or sometimes dehydration. Reactions in which compounds are formed with the loss of water include the synthesis of proteins (C–N), acylglycerols and oligosaccharides (C–O), and polynucleotides (P–O) from their monomers, and the formation of ATP from ADP and inorganic phosphate during respiration. Since these types of molecules often have higher free energies than their monomers, a common strategy in biology is to first activate the components by, for example, formation of a phosphate ester, which can then be reacted with another activated species via the loss of water. DNA synthesis from activated nucleoside triphosphates is just a series of sequential phosphodiester bond formations in which water is lost, catalyzed by the enzyme DNA polymerase.

Loss of carbon dioxide is a common strategy for removing a single carbon atom from a molecule

Since C–C and C=O bonds are quite stable, shortening a molecule by one carbon atom is not an easy process chemically. In biology, it is usually accomplished by the loss of carbon dioxide, a process that is thermodynamically favored because CO_2 is a very stable molecule. Loss of CO_2 is termed **decarboxylation** and is usually assisted by cofactors. Several different cofactors can participate in decarboxylations. The most common are pyridoxal phosphate (PLP) and thiamine diphosphate (TDP; also referred to as TPP, thiamine pyrophosphate), but transition-metal ions such as manganese are sometimes used instead. Pyruvate decarboxylase, which converts pyruvate to acetaldehyde in the pathway for ethanol biosynthesis, uses TPP as a cofactor in its decarboxylation reaction (Figure 2-34).

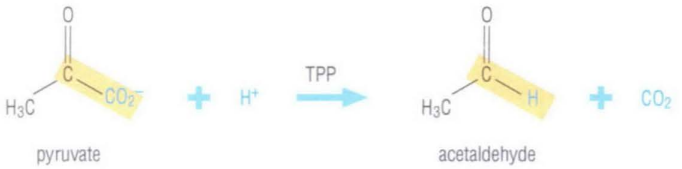


Figure 2-34 Example of the decarboxylation of a carboxylic acid Shortening of the three-carbon unit of pyruvate to the two-carbon unit of acetaldehyde is accomplished by the loss of CO_2, catalyzed by the cofactor TPP bound at the active site of the enzyme pyruvate decarboxylase. The CO_2 molecule is derived from the carboxylate group of the acid. To help break the C–C bond, in this and similar reactions the carbon of the acid group is usually activated in some fashion, often by temporary chemical coupling to the enzyme's cofactor.

References

Silverman, R.B.: *The Organic Chemistry of Enzyme-Catalyzed Reactions* (Academic Press, New York, 2000).

©2004 New Science Press Ltd

Active sites promote acid-base catalysis

In essentially all biological reactions, regardless of the reaction type, there is at least one step that involves the transfer of a proton from one group to another. Groups donating protons are referred to as **acids**; the groups that accept them are called **bases**. Often the gain or loss of a proton will change the chemical reactivity of the group considerably. Catalysis in which a proton is transferred in going to or from the transition state is called **acid-base catalysis**.

The ease with which a proton can be transferred between an acid and a base depends on the relative proton affinities of the two groups. Proton affinity is measured by the **pK_a value**, which over typical ranges of pK_a can be thought of as the pH of an aqueous solution of the acid or base at which half of the molecules are protonated and the other half are deprotonated. Strong acids lose their protons readily to water, forming hydronium ions (H_3O^+). Strong acids have pK_a values of 2 or lower. Strong bases tend to take protons from water, forming the

Table of Typical pKa Values

Acid (proton donor)		Conjugate base (proton acceptor)	pK_a
$HCOOH$ formic acid	⇌	$HCOO^-$ formate ion	3.75
CH_3COOH acetic acid	⇌	CH_3COO^- acetate ion	4.76
$CH_3CH(OH)COOH$ lactic acid	⇌	$CH_3CH(OH)COO^-$ lactate ion	3.86
H_3PO_4 phosphoric acid	⇌	$H_2PO_4^-$ dihydrogen phosphate ion	2.14
$H_2PO_4^-$ dihydrogen phosphate ion	⇌	HPO_4^{2-} monohydrogen phosphate ion	6.86
HPO_4^{2-} monohydrogen phosphate ion	⇌	PO_4^{3-} phosphate ion	12.4
H_2CO_3 carbonic acid	⇌	HCO_3^- bicarbonate ion	6.37
HCO_3^- bicarbonate ion	⇌	CO_3^{2-} carbonate ion	10.25
C_6H_5OH phenol	⇌	$C_6H_5O^-$ phenolate ion	9.89
NH_4^+ ammonium ion	⇌	NH_3 ammonia	9.25
H_2O	⇌	OH^-	15.7

Figure 2-35 Table of pK_a values for some common weak acids in biology Note that for compounds with more than one ionizable group (for example, phosphoric acid), the loss of the second (and third) proton is always more difficult than the loss of the first (and second). This is because, if each loss of a proton creates a negative charge, charge repulsion makes it harder to put more negative charge on the conjugate base.

Definitions

acid: a molecule or chemical group that donates a proton, either to water or to some other base.

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.

base: a molecule or chemical group that accepts a proton, either from water or from some other acid.

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.

©2004 New Science Press Ltd

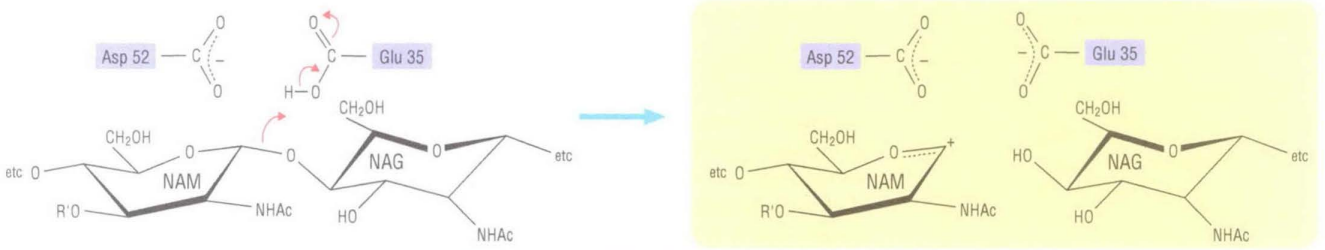


Figure 2-36 Active site of lysozyme The enzyme lysozyme hydrolyzes the acetal links between monomers in certain carbohydrate polymers. The substrate shown here is part of a polymer of *N*-acetyl glucosamine (NAG) and *N*-acetyl muramic acid (NAM). Two carboxylic acid side chains (aspartate and glutamate; purple) are found in the active site of lysozyme. In solution, these residues would be expected to have a pK_a around 4 or 5, just like acetic and lactic acids (see Figure 2-35). But in the microenvironment provided by the protein, their acidities differ considerably. Aspartic acid 52 has the expected pK_a, so at pH 7 it is ionized and can fulfill its function, which is to use its negative charge to stabilize the positive charge that builds up on the sugar during catalysis. Glutamic acid 35, however, is in a hydrophobic pocket, which raises its pK_a to around 7. In its protonated form it acts as a weak acid and donates a proton to the sugar –C–O–R group (where R is the next sugar in the chain), breaking the C–O bond. In its negatively charged form it helps stabilize the positive charge on the transition state. There is recent evidence that the mechanism may also involve a covalent intermediate between aspartic acid 52 and the substrate. The red arrows show the movement of electron pairs as bonds are made and broken.

hydroxide ion (OH^-). Strong bases have pK_a values greater than about 12. In this context, water is a very weak acid and a very weak base. Most biological acids and bases are weak; they only partially give up protons in aqueous solution at physiological pH and exist as an equilibrium between protonated and unprotonated species. If the pK_a of the group is between 4 and 7, it is a weak acid (the higher the pK_a the weaker the acid); if the pK_a is between 7 and 10, the group is a weak base (the lower the pK_a the weaker the base). Proton transfers occur efficiently from groups with low pK_a values to those of higher values. Figure 2-35 shows the pK_a values for some common biological acids and bases.

Missing from the list of pK_a values in Figure 2-35 is the weakest acid of importance in biology, the aliphatic carbon group, –C–H. Carbon has only a vanishingly small tendency to give up a proton in aqueous solution; the pK_a value of the –C–H groups in simple sugars is over 20. Yet the transfer of a proton to and from a carbon center is a common reaction in biology, occurring in almost half of the reactions of intermediary metabolism. That it can occur at all, and occur efficiently, is due to the ability of enzyme active sites to change effective pK_a values.

Enzymes can increase the efficiency of acid-base reactions by changing the intrinsic pK_a values of the groups involved. Thus, the alpha –C–H group in lactic acid can be made more acidic (that is, its pK_a can be lowered) by, for example, making a strong hydrogen bond to the –OH group attached to it. This hydrogen bond will tend to pull electrons away from the oxygen atom, which in turn will pull electrons away from the adjacent –C–H bond, weakening the affinity of the carbon for its hydrogen and thus lowering the pK_a. The pK_a of a weak acid such as the carboxylic acid side chain of lactic acid (pK_a ~ 3.9 in water) can be raised to 7 or higher by, for example, placing the group in a nonpolar environment. With no water molecules around to accept a proton, the carboxylic acid will tend to hang on to its hydrogen rather than lose it, thereby generating a negatively charged carboxylate anion in a hydrophobic region of the protein; thus, its pK_a will be raised and it will become an even weaker acid (and consequently a much stronger base). Figure 2-36 shows just this situation for the carboxylic acid side chain of glutamate in the active site of the enzyme lysozyme, where it is estimated that the pK_a of glutamic acid 35 is raised from about 4 to above 6, and it can donate a proton to catalyze the breaking of the C–O bond in the substrate.

References

Malcolm, B.A. *et al.*: **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.

Vocadlo, D.J. *et al.*: **Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate.** *Nature* 2001, **412**:835–838.

Voet, D. and Voet, J.: *Biochemistry* 2nd ed. Chapter 12 (Wiley, New York, 1995).

©2004 New Science Press Ltd

2-13 Cofactors

Many active sites use cofactors to assist catalysis

Not every biological reaction can be carried out efficiently using only the chemical properties of the 20 naturally occurring amino acids. Oxidation/reduction (redox) reactions, for example, in which electrons are transferred from one group to another, can be promoted to some extent by the cysteine sulfur atom, but most redox reactions need more help than this, as we saw in section 2-10. The creation of unpaired electrons (free radicals), which are useful in a number of chemical reactions, also requires chemical species that are not found in amino-acid side chains. To overcome such limitations, many enzyme active sites contain non-amino-acid **cofactors** that allow specialized chemical functions.

Cofactors can be as small as a metal ion or as large as a heterocyclic organometallic complex such as heme. They are tightly bound—sometimes covalently attached—to the proteins in which they function. Cofactors that are organic compounds and assist catalysis are often referred to as **coenzymes**. Cofactors may be imported into an organism from the food it eats (in humans most cofactors are derived from vitamins and minerals in the diet), or they can be synthesized from simple building blocks. Some common vitamin-derived cofactors are listed in Figure 2-37. In many organisms, specific sets of genes are devoted not only to the synthesis

Figure 2-37 Table of organic cofactors

Some Common Coenzymes

Coenzyme [vitamin from which it is derived]	Entity transferred	Representative enzymes that use coenzyme	Deficiency disease
thiamine pyrophosphate (TPP or TDP) [vitamin B_1, thiamin]	aldehydes	pyruvate dehydrogenase	beri beri
flavin adenine dinucleotide (FAD) [vitamin B_2, riboflavin]	hydrogen atoms	succinate dehydrogenase	(a)
nicotinamide adenine dinucleotide (NAD^+) [niacin]	hydride ion	lactate dehydrogenase	pellagra
nicotinamide adenine dinucleotide phosphate ($NADP^+$) [niacin]	hydride ion	isocitrate dehydrogenase	pellagra
pyridoxal phosphate (PLP) [vitamin B_6, pyridoxal]	amine groups	aspartate aminotransferase	(a)
coenzyme A (CoA) [pantothenic acid]	acyl groups	acetyl-CoA carboxylase	(a)
biotin (biocytin) [biotin]	CO_2	propionyl-CoA carboxylase	(a)
5'-deoxyadenosylcobalamin [vitamin B_{12}]	H atoms and alkyl groups	methylmalonyl-CoA mutase	pernicious anemia
tetrahydrofolate (THF) [folate]	one-carbon units	thymidylate synthase	megaloblastic anemia
lipoamide [lipoic acid]	two-carbon units; R-SH	pyruvate dehydrogenase	(a)
heme [no vitamin]	e^-, O_2, NO, CO_2	cytochrome oxidase	anemia, leukemia

(a) no specific name

Definitions

coenzyme: a cofactor that is an organic or organometallic molecule and that assists catalysis.

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.

©2004 New Science Press Ltd

of some cofactors, but also to their insertion into the active sites of the particular proteins in which they are to function. It is likely that biochemical machinery ensuring that the right cofactor binds to the right protein at the right time is prevalent throughout the living world.

Metal-ion cofactors (Figure 2-38) are usually first-row transition metals and the most common are also among the most abundant in the Earth's crust, suggesting that many enzymes evolved to use whatever chemically reactive species were plentiful at the time. Enzymes are known that use molybdenum, nickel, cobalt, and manganese, but the majority of metalloenzymes use iron, copper, zinc, or magnesium. In some cases, more than one type of metal ion will work (that is, promote catalysis) in experiments carried out on a purified enzyme that has been stripped of its native metal, but it is thought that in the cell one type predominates or is used exclusively.

The structures of typical organic cofactors (coenzymes) reveal a striking pattern: most of them look like pieces of RNA; that is they are either ribonucleotides or derivatives of ribonucleotides. It has been suggested that this is a relic of an earlier "RNA world" that preceded the protein-based world we live in today. In this scenario, the earliest proteins would have looked for help from the most commonly available reactive chemical compounds, and these might have been RNA molecules or fragments thereof.

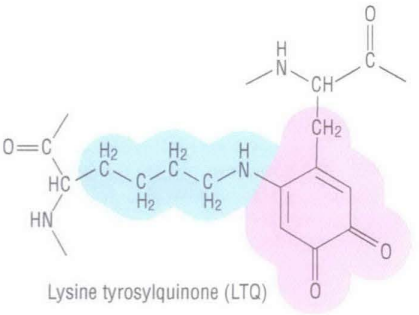
The function of organic cofactors is to transfer specific chemical species to a substrate. The groups transferred range from electrons and hydride ions (H^-) to small carbon-containing fragments up to about two carbons in length. These species are usually either extremely unstable on their own or would be damaging to the protein if they were not contained. The function of the coenzyme is to make them, transfer them, and/or sequester them. If the species transfer causes a change in the chemical structure of the cofactor, as in the case of NAD^+ (or its phosphorylated counterpart $NADP^+$, which does the same chemistry) it is necessary to recycle the cofactor back to its original reactive form before another enzyme turnover can take place. In many cases, specific reactions catalyzed by other enzymes perform this necessary function in a tightly coupled system.

Recently, it has been found that amino-acid side chains in a protein can be modified to produce an organic cofactor *in situ*. One striking example is found not in an enzyme but in the green fluorescent protein (GFP) from marine organisms, which is used in biology as an optical marker. Because the fluorescent chromophore is synthesized by the protein itself from the reaction of a tyrosine side chain with neighboring serine and glycine residues, GFP can be introduced genetically into any organism without the need for the organism to have other genes to make the cofactor. An example of this type of cofactor in enzymes is the unusual coenzyme lysine tyrosylquinone (LTQ) (Figure 2-39) in copper amine oxidase. LTQ is synthesized by the addition of a lysine side chain of the enzyme itself to the aromatic ring of an oxidized tyrosine residue elsewhere on the protein chain. The active enzyme is essential for the proper cross-linking of elastin and collagen in connective tissue because the cross-linking reaction is difficult and involves redox chemistry that the LTQ promotes.

Metal Ions and Some Enzymes Requiring Them

Metal ion	Enzyme
Fe^{2+} or Fe^{3+}	cytochrome oxidase catalase peroxidase
Cu^{2+}	cytochrome oxidase
Zn^{2+}	DNA polymerase carbonic anhydrase alcohol dehydrogenase
Mg^{2+}	hexokinase glucose-6-phosphatase pyruvate kinase
Mn^{2+}	arginase
K^+	pyruvate kinase
Ni^{2+}	urease
Mo	nitrate reductase
Se	glutathione peroxidase

Figure 2-38 Table of metal-ion cofactors



Lysine tyrosylquinone (LTQ)

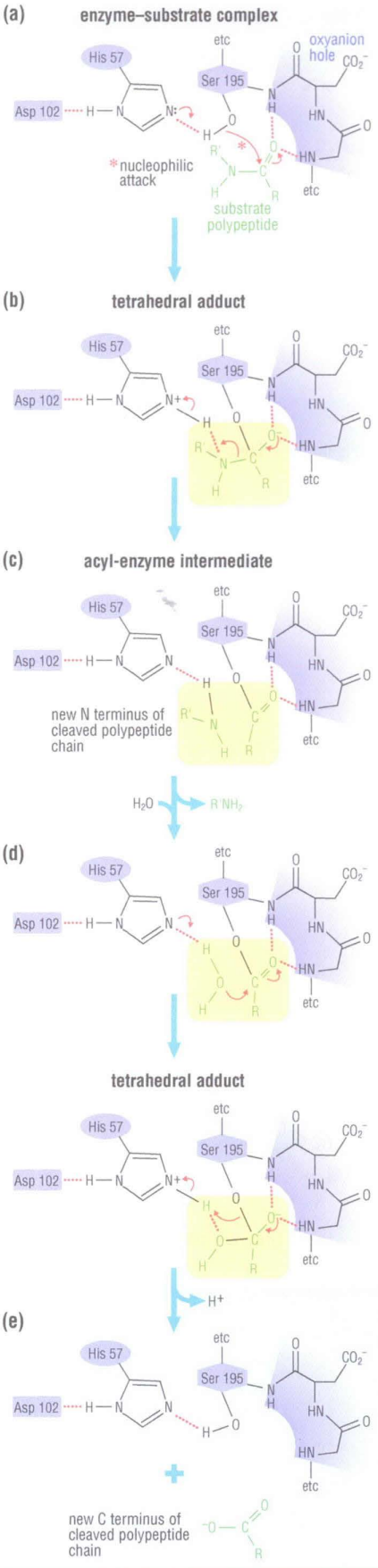
Figure 2-39 The coenzyme lysine tyrosylquinone Lysine tyrosylquinone is formed on the enzyme by the addition of the lysine side chain (blue) to an oxidized form of tyrosine (pink). Formation of the cofactor is catalyzed by the copper ion (not shown) required as part of the overall reaction catalyzed by the enzyme. These types of cofactors are essential in certain enzymes that catalyze oxidation reactions involving a highly reactive radical intermediate. They stabilize the radical and make it available for the reaction.

References

Voet, D. and Voet, J. *Biochemistry* 2nd ed. (Wiley, New York, 1995).

©2004 New Science Press Ltd

2-14 Multi-Step Reactions



Some active sites employ multi-step mechanisms

The chemistry of most biological reactions is very difficult to carry out at ordinary temperatures and pressures in aqueous solution at neutral pH. To catalyze such reactions, enzymes must circumvent the high-energy transition state that prevents the reaction from occurring under moderate conditions (see Figure 2-21). One way of achieving this, as we have seen (section 2-9), is by stabilizing the transition state. Alternatively, enzymes may direct the reaction along a different route from substrates to products, one not accessible to the uncatalyzed reaction in aqueous solution. This often entails breaking the reaction up into a number of steps, each of which has a lower-energy transition state. The "product" of each individual step is a relatively unstable reaction intermediate (see section 2-6). In many instances, these intermediates can be trapped by physical (low temperature) or chemical (reduction) methods, and thus can be characterized completely. In other instances, the existence of these intermediates has only been inferred from kinetic or spectroscopic data, or from chemical logic.

Some intermediates are covalent: they involve the linkage of an amino-acid side chain or part of a cofactor to the substrate. The product, a covalent intermediate, then reacts further, forming other intermediates and/or the final product(s). The classic example of this strategy is the serine protease reaction, in which the peptide bonds between amino acids are broken to degrade proteins, as occurs in digestion. Hydrolysis of a peptide bond is a difficult chemical process: this is why proteins are normally stable in cells. In the uncatalyzed reaction, water attacks the carbonyl carbon of the amide bond. Because of the polarity of this bond, the carbon atom is somewhat electron-deficient and therefore susceptible to attack by a **nucleophile**, a group that is electron-rich. Water is a poor nucleophile. To overcome this problem, two strategies are possible: water can be activated to make it a better nucleophile, and/or the amide group can be activated to make it more susceptible to attack. Enzymes that catalyze the degradation of proteins by hydrolysis of the amide bonds of their backbones solve this problem by dividing the overall reaction into two steps; they activate both the attacking nucleophile (serine OH or H_2O) and the amide group in a reaction that has two steps with an intermediate (Figure 2-40).

In the first step (Figure 2-40a–c), instead of using the –OH group of water as the nucleophile, the enzyme uses the hydroxyl group of a serine side chain, which attacks the carbonyl carbon of the amide bond, producing an acyl-enzyme intermediate and releasing the amino-terminal portion of the peptide substrate. The –OH of a serine is also not a good nucleophile and requires some activation. It is able to break the amide bond in the first step of this reaction for two reasons. One is proximity and orientation: the serine is positioned adjacent to the carbonyl carbon of the substrate in the active site. The second is that the –OH of the serine is activated by hydrogen bonding with a histidine side chain which, in turn, is anchored by interaction with the carboxylate group of an aspartic acid (see Figure 2-40a). The aspartic acid ensures that the histidine is in the right charge state to pull a proton from the –OH of serine, moving the serine towards a serine alkoxide ($-O^-$), which is a much better nucleophile. The Asp-His-Ser signature of the serine proteases is also called the **catalytic triad**. The energy barrier for the formation of the acyl-enzyme intermediate is further lowered by stabilization of the transition state (the tetrahedral adduct) leading to its formation, with concomitant activation of the carbonyl group of the peptide substrate for attack (Figure 2-40b). In the transition state, negative charge develops on the carbonyl oxygen. The protein is configured to stabilize this

Figure 2-40 The chemical steps in peptide hydrolysis catalyzed by the serine protease chymotrypsin
(a) The substrate (green) binds to the active site such that the hydroxyl group of serine 195 is positioned to attack the peptide carbonyl carbon. Simultaneously, the carbonyl oxygen hydrogen bonds to two –NH groups of the protein and the histidine removes a proton from the serine –OH, thus activating it. The two –NH groups are part of the peptide backbone forming a loop called the oxyanion hole (shaded in purple). **(b)** In the transition state for formation of the acyl-enzyme intermediate, negative charge has developed on the carbonyl oxygen; this is also stabilized by hydrogen bonding in the oxyanion hole. The first product is beginning to be released in this step and the bond between the carbonyl carbon of the substrate and the serine oxygen is beginning to form. **(c)** The acyl-enzyme intermediate. In this intermediate, the serine side chain is covalently bound to the carbonyl carbon of the substrate. The histidine gives up the proton it acquired from the serine to the nitrogen of the substrate, making it easier for the first product to split off from the rest of the substrate. **(d)** In the transition state for hydrolysis of the acyl-enzyme intermediate, a water molecule attacks the carbonyl carbon of the acyl-enzyme ester as the histidine accepts a proton from it, analogous to the acylation step (b). The transient negative charge on the oxygen of the acyl enzyme is again stabilized by hydrogen bonding in the oxyanion hole. **(e)** Formation of product. The transition state collapses to release the other portion of the substrate and regenerate the serine –OH.

©2004 New Science Press Ltd

negative charge through the donation of two hydrogen bonds to the carbonyl oxygen from what has been called the **oxyanion hole** in the active site (Figure 2-40).

Step two of the reaction requires hydrolysis of the covalent acyl-enzyme intermediate (Figure 2-40d,e). When the intermediate is hydrolyzed, the carboxy-terminal portion of the peptide substrate is released and the serine on the enzyme is restored to its original state. Attack of water on the acyl enzyme is more facile than attack of water on the original amide substrate, for the hydrolysis of an ester is easier than hydrolysis of an amide. This is because the C–N bond of an amide has more double-bond character than the corresponding C–O bond of an ester. The transition state for hydrolysis of the acyl enzyme is similar to that already described for its formation, and it is stabilized in the same way, by interactions with the oxyanion hole (Figure 2-40d).

An obvious question raised by this example is why not just activate a molecule of water directly by pulling a proton off it, turning the weak nucleophile H_2O into the strong nucleophile OH^-, instead of activating a serine first. Interestingly, this is exactly what happens in another class of proteases, the metalloproteases. Instead of using a covalent enzyme intermediate, these enzymes use a metal-ion cofactor to activate water for direct attack on the substrate carbonyl, while at the same time activating the carbonyl. In contrast to the serine protease mechanism, metalloproteases such as carboxypeptidase and thermolysin do not proceed through a stable intermediate. They bind a molecule of water to a metal ion, usually Zn^{2+}. Binding H_2O to a transition metal ion lowers the pK_a of the bound water, making it a better acid. It is thus much easier to turn metal-bound H_2O into OH^-. The metal-bound OH^- can attack the amide bond of a peptide substrate directly, hydrolyzing it without the need for a covalent enzyme intermediate. The presence of two or more classes of enzymes that carry out the same chemical process by different chemical strategies is a common feature of biology.

The two-step strategy is also often deployed in phosphoryl-group transfer reactions, particularly those catalyzed by kinases and phosphatases, which, as we shall see in Chapter 3, are among the most important mechanisms for regulating protein function. Unlike the serine proteases, in which a serine (or a cysteine in the case of the related cysteine protease family) is always the attacking nucleophile, phosphoryl-group transfer can occur with a wide variety of attacking groups. Serine, threonine and tyrosine –OH groups, the carboxylate groups of aspartate and glutamate, cysteine –SH and the nitrogen of histidine are all known to participate in various phosphoryl-group transfers. In the bacterial enzyme alkaline phosphatase, which catalyzes the transfer of phosphate to water from many organophosphate substrates, a serine –OH on the enzyme first attacks the phosphorus atom, releasing the organic portion of the substrate and leaving behind a phosphoserine-enzyme intermediate, which is then hydrolyzed by water, a mechanism analogous to that of the serine proteases. Although it is difficult to trap acyl-enzyme intermediates such as those formed in the serine and cysteine protease reactions, phosphoenzyme intermediates can be more stable and have been isolated for many enzymes. For example, phosphoglucomutases catalyze the interconversion of D-glucose 1-phosphate and D-glucose 6-phosphate, a reaction central to energy metabolism in all cells and to the synthesis of cell-wall polysaccharides in bacteria. The reaction proceeds through a phosphoaspartyl anhydride intermediate, which is formed when the phosphate on the aspartyl group of the enzyme attacks the unphosphorylated position on the sugar and transfers the phosphate group to it (Figure 2-41). Subsequent reaction of the phosphoryl group at the other position on the sugar with the aspartate regenerates the phosphoaspartyl-enzyme. The structure of the phosphorylated enzyme from *Lactococcus lactis* clearly shows the phosphate bound to the side chain of aspartic acid 8. The absence of a catalytic base near the aspartyl phosphate group to activate a water molecule or another substrate group accounts for the persistence of the phosphorylated enzyme under physiological conditions.

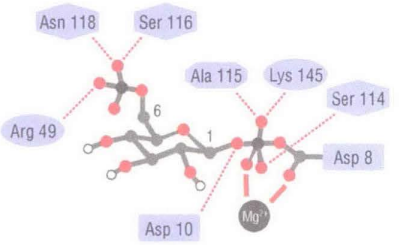


Figure 2-41 The phosphoenzyme–substrate intermediate in the active site of beta-phosphoglucomutase from *Lactococcus lactis*
In the active site of beta-phosphoglucomutase, phosphoryl-group transfer occurs via aspartic acid 8, which can be phosphorylated to form a stable phosphoaspartyl-enzyme. In the reaction glucose 6-phosphate (G6P) to glucose 1-phosphate (G1P), for example, the unphosphorylated C1 OH of the substrate is positioned such that it can become linked to the phosphoaspartyl group as shown. In this intermediate, the phosphate being transferred is pentavalent and is still associated with both aspartic acid 8 (in a mixed anhydride linkage) while becoming covalently linked to the G6P substrate to form the intermediate 1,6-bisphosphoglucose–enzyme complex. Subsequent reaction of the phosphoryl group at the 6 position of the sugar with the aspartic acid 8 is thought to complete the transfer to C1, forming G1P and regenerating the phosphoaspartyl-enzyme.

Definitions

catalytic triad: a set of three amino acids that are hydrogen bonded together and cooperate in catalysis.

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

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 $-O^-$--H–X hydrogen bonds.

References

Lahiri, S.D. *et al.*: **Caught in the act: the structure of phosphorylated beta-phosphoglucomutase from *Lactococcus lactis***. *Biochemistry* 2002, **41**:8351–8359.

Silverman, R.B.: *The Organic Chemistry of Enzyme-Catalyzed Reactions* Revised ed. (Academic Press, New York, 2002).

Walsh, C.: *Enzymatic Reaction Mechanisms* Chapter 3 (Freeman, San Francisco, 1979).

©2004 New Science Press Ltd

2-15 Multifunctional Enzymes

Some enzymes can catalyze more than one reaction

In some cases an enzyme may catalyze more than one chemical transformation. Such enzymes may be composed of a single polypeptide chain with one or more active sites, or may be composed of more than one polypeptide chain, each with an active site. In the latter case, each polypeptide chain represents an independently folded subunit which normally does not, however, exist in the absence of the others. Such enzymes are called **bifunctional** (or **multifunctional** if more than two reactions are involved) and they fall into three classes. In the first class, the two reactions take place consecutively at the same active site. In the second, two separate chemical reactions are catalyzed by two distinct active sites, each located in a different domain some distance apart. In the third, two or more reactions are also catalyzed by two or more distinct active sites, but these are connected by internal channels in the protein, through which the product of the first reaction diffuses to reach the next active site, where it undergoes further reaction. In this section we shall look at examples of the first and second classes. In the following section, we deal with multifunctional enzymes with internal channels, and will also consider enzymes that have additional non-enzymatic functions.

Some bifunctional enzymes have only one active site

In bifunctional enzymes that carry out two different reactions using the same active site, the second reaction is simply an inevitable chemical consequence of the first, because the product of the initial reaction is chemically unstable. Although catalysis of the second step, involving the breakdown of this first product, might seem unnecessary to sustain a rapid overall reaction rate, this second reaction is often catalyzed as well. The active site is the same for the two consecutive transformations, as the first product does not dissociate from the enzyme. An example is the reaction catalyzed by isocitrate dehydrogenase (ICDH), an enzyme in the TCA-cycle pathway. ICDH uses the cofactor NADP to oxidize isocitrate to oxalosuccinate, a compound that is unstable because it is a beta-ketoacid. In the same active site, with the assistance of a manganese ion, oxalosuccinate is catalytically decarboxylated, independent of the presence of NADP, to give the final product, alpha-ketoglutarate (Figure 2-42).

Some bifunctional enzymes contain two active sites

A second class of bifunctional enzyme contains two independently folded domains, each of which has a distinct, non-overlapping active site. Although there are some exceptions, it is usually the case that the product of one active site is the substrate for the other active site.

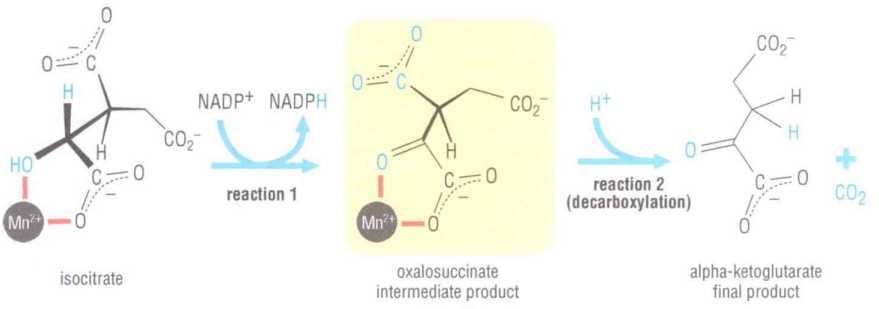


Figure 2-42 The reaction catalyzed by isocitrate dehydrogenase The reaction catalyzed by this enzyme consists of an oxidation reaction and a decarboxylation reaction. In the active site of this TCA-cycle enzyme, isocitrate is oxidized and decarboxylated to form alpha-ketoglutarate. In the first reaction, isocitrate is oxidized to oxalosuccinate and the cofactor NADP is concomitantly reduced to NADPH. Although oxalosuccinate will decarboxylate spontaneously to form the desired product, alpha-ketoglutarate, in the active site of isocitrate dehydrogenase this decarboxylation is catalyzed in a second reaction with the assistance of an enzyme-bound manganese ion. Oxalosuccinate never escapes from the enzyme.

Definitions

bifunctional: having two distinct biochemical functions in one gene product. Bifunctional enzymes catalyze two distinct chemical reactions.

multifunctional: having a number of distinct biochemical functions in one gene product.

References

Greasley, S.E. *et al.*: **Crystal structure of a bifunctional transformylase and cyclohydrolase enzyme in purine biosynthesis.** *Nat. Struct. Biol.* 2001, **8**:402–406.

Hurley, J.H. *et al.*: **Catalytic mechanism of NADP(+)-dependent isocitrate dehydrogenase: implications from the structures of magnesium–isocitrate and NADP+ complexes.** *Biochemistry* 1991, **30**:8671–8678.

Knighton, D.R. *et al.*: **Structure of and kinetic channelling in bifunctional dihydrofolate reductase-thymidylate synthase.** *Nat. Struct. Biol.* 1994, **1**:186–194.

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

Liu, J. *et al.*: **Thymidylate synthase as a translational regulator of cellular gene expression.** *Biochim. Biophys. Acta* 2002, **1587**:174–182.

Wolan, D.W. *et al.*: **Structural insights into the avian AICAR transformylase mechanism.** *Biochemistry* 2002, **41**:15505–15513.

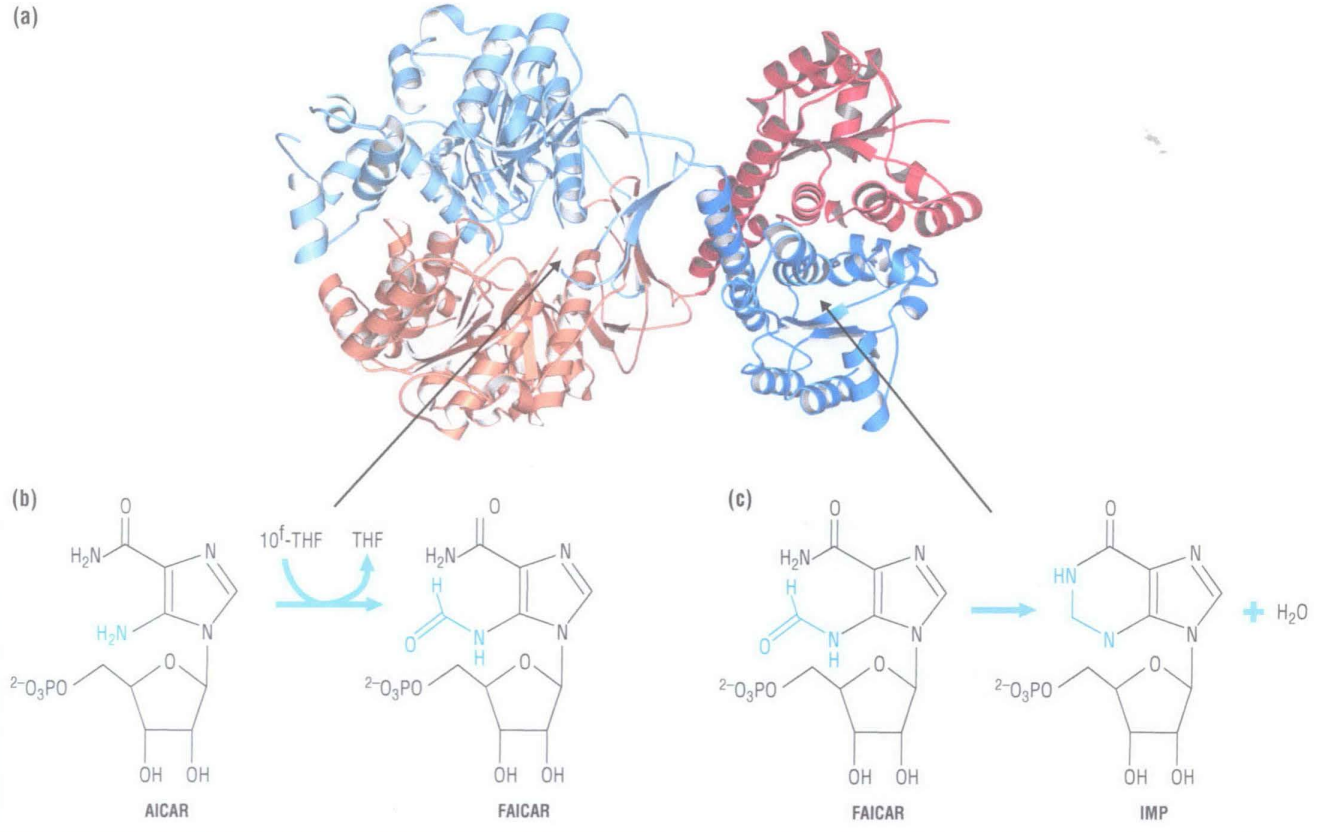
©2004 New Science Press Ltd

Unless these two sites were to face each other (which has never been found) or be physically connected in some fashion, the first product must dissociate from the first domain and diffuse through the cellular medium or along the protein surface in order to find the second active site. There would seem to be no catalytic advantage to having the two activities residing in one fused gene product, but the arrangement does at least allow coordinate regulation of the biosynthesis of both enzyme activities. Genome sequencing has provided evidence that any advantage is small: in nearly all instances, at least one organism can be found in which the two different catalytic activities are encoded by two distinct genes.

One example of this type of bifunctional enzyme is dihydrofolate reductase-thymidylate synthase from the parasite *Leishmania major*, which catalyzes two reactions in the biosynthesis of thymidine. In most other organisms, including humans and *E. coli*, these two enzymatic activities are carried by two separate proteins, encoded by separate genes: a thymidylate synthase and a dihydrofolate reductase. In the first active site of the *Leishmania* enzyme, a molecule of 2′-deoxyUMP is methylated by N^5,N^{10}-methylene-tetrahydrofolate to give thymidylate (TMP) and dihydrofolate. In the second active site, located 40 Å away in a separate folded domain, the dihydrofolate is reduced by NADH to generate tetrahydrofolate. (Cofactor regeneration is completed by another, separate, enzyme.) In *Leishmania*, presumably, two genes encoding the two separate enzyme activities were originally arranged in tandem and at some point during evolution the stop codon for the first one was lost, creating a single gene with a bifunctional protein product.

Another example of this is found in the purine biosynthesis pathway. In most organisms AICAR transformylase-IMP cyclohydrolase (ATIC) is a 64 kDa bifunctional enzyme that possesses the final two activities in *de novo* purine biosynthesis, 5-aminoimidazole-4-carboxamide-ribonucleotide (AICAR) transformylase and IMP cyclohydrolase. ATIC forms an intertwined dimer, with each monomer composed of two separate different functional domains. The amino-terminal domain (up to residue 199) is responsible for the IMP cyclohydrolase activity, whereas the AICAR transformylase activity resides in the carboxy-terminal domain (200–593). The active sites of the two domains are approximately 50 Å apart, with no structural evidence of a tunnel connecting the two active sites (Figure 2-43). In the archaeon *Methanopyrus kandleri*, however, these two activities are carried out by two independent proteins, encoded by two distinct genes.

Figure 2-43 The bifunctional enzyme, AICAR transformylase-IMP cyclohydrolase (ATIC) is a single enzyme with two distinct active sites
(a) This enzyme occurs as an intertwined homodimer, with each monomer composed of two distinct domains (dark and light blue/dark and light red). The two domains are each involved in catalyzing a different reaction in the biosynthesis of inosine monophosphate (IMP), the initial purine derivative that is the precursor to adenosine and guanosine. One site of each type is indicated by an arrow, with the reaction that takes place there described below. In all there are four active sites on the dimer, one on each domain. **(b)** The active sites for the AICAR transformylase reaction on the two monomers are each formed by residues contributed by both monomers. In the first site a transformylase reaction on the substrate 5-aminoimidazole-4-carboxamide ribonucleotide (AICAR) produces a product, formyl-5-AICAR (FAICAR), that is a substrate for the second active site. 10^f-THF is the cofactor N^{10}-formyltetrahydrofolate. **(c)** In the second active site, FAICAR is cyclized to IMP. In most organisms these two reactions are carried out by the bifunctional enzyme, but in some species the transformylase and the cyclohydrolase reactions are carried out by two distinct proteins (encoded by two separate genes). Each of these resembles in structure the corresponding domain in the bifunctional enzyme.



©2004 New Science Press Ltd

2-16 Multifunctional Enzymes with Tunnels

Some bifunctional enzymes shuttle unstable intermediates through a tunnel connecting the active sites

A third small class of multifunctional enzymes has been found to possess a remarkable connection between their separate active sites. A physical channel (or channels) allows the product of one reaction to diffuse through the protein to another active site without diffusing out into contact with the cellular medium. Two reasons for such a feature have been identified: either the first reaction product is an uncharged species that might be lost from the cell altogether by diffusion through the membrane if it were allowed to escape from the protein; and/or the reaction product is so unstable in free solution that it would decompose before it had time to find the second active site.

The first enzyme found to have such a tunnel through its structure was the bifunctional tryptophan synthase from *Salmonella typhimurium*. Unlike the enzymes discussed in the previous section, this is composed of two subunits, encoded by separate genes. The tunnel is 25 Å long and connects the active site of the alpha-subunit, in which indole, an uncharged molecule that might diffuse out of the cell, is generated from indole 3-glycerolphosphate, to the second active site, in which the indole is added to a molecule of acrylate, derived from serine, to produce tryptophan (Figure 2-44). Indole remains enzyme-bound throughout the reaction. The tunnel, as expected, is lined with nonpolar side chains to retain the nonpolar indole molecule.

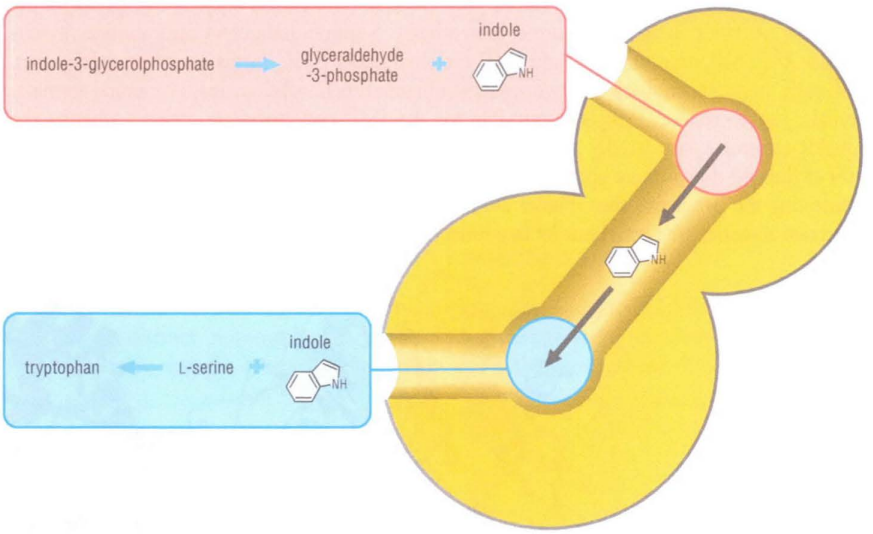


Figure 2-44 The two active sites of the bifunctional enzyme tryptophan synthase are linked by an internal channel In the reaction catalyzed by this enzyme, indole-3-glycerolphosphate (IGP) is first converted to indole and glyceraldehyde-3-phosphate in an active site located in the alpha-subunit. The uncharged indole molecule diffuses through a channel in the enzyme, lined with nonpolar side chains, to the second active site, in the beta-subunit, where a second substrate, serine, has been converted to an acrylate adduct of the cofactor pyridoxal phosphate (PLP). There the indole adds to the acrylate and the product hydrolyzes to give tryptophan and regenerate PLP.

References

Huang, X. *et al.*: **Channeling of substrates and intermediates in enzyme-catalyzed reactions.** *Annu. Rev. Biochem.* 2001, **70**:149–180.

Hyde, C.C. *et al.*:**Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from *Salmonella typhimurium*.** *J. Biol. Chem.* 1988, **263**:17857–17871.

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

Thoden, J.B. *et al.*: **Structure of carbamoyl phosphate synthetase: a journey of 96 Å from substrate to product.** *Biochemistry* 1997, **36**:6305–6316.

©2004 New Science Press Ltd

Trifunctional enzymes can shuttle intermediates over huge distances

The second example is considerably more complicated because it is a trifunctional enzyme, carbamoyl phosphate synthetase, an enzyme involved in the synthesis of 2′-deoxyUMP and in the urea cycle. The single-chain protein has three separate active sites connected by two tunnels through the interior of the protein, and individual genes corresponding to the different activities have never been found in any organism. This design has presumably evolved because the first reaction produces ammonia, a neutral species, which travels along a tunnel to the second active site where it reacts with carboxyphosphate to give a carbamate intermediate that would be too unstable to survive in aqueous solution. Therefore, it is transported through the interior of the protein to the third active site, where it is phosphorylated by ATP to give the final product. The entire journey from first substrate to carbamoyl phosphate covers a distance of nearly 100 Å: this enzyme is illustrated in Figure 2-45. Bifunctional enzymes are known in which the two reactions catalyzed are not consecutive; in this case a separate enzyme catalyzes the intervening reaction and channels substrate between the two active sites.

Some enzymes also have non-enzymatic functions

The genomes of mammals are not that much larger, in the number of genes, than those of plants, fish, flies and worms. It is becoming clear that one reason for this economy of gene number is that, in multicellular eukaryotes at least, many enzymes have at least a second biochemical function that is unrelated to their catalytic activity. This function is usually regulatory: some enzymes double as transcription factors; others act as signaling proteins; some are cofactors for essential reactions in protein synthesis; and yet others are transported out of the cell to serve as cytokines or growth factors. It is possible that this multiplicity of functions, in addition to allowing the genome to remain relatively compact, also connects various processes. Having a metabolic enzyme double as a repressor, for example, couples the expression of some genes to metabolism in a direct way. For example, the folate-dependent enzyme thymidylate synthase (see section 2-15) also functions as an RNA-binding protein. It interacts with its own mRNA to form a ribonucleoprotein complex, and there is also evidence that it can interact with a number of other cellular mRNAs, including transcripts of the *p53* tumor suppressor gene and the *myc* family of transcription factor genes. The functional consequence of such binding is to repress the translation of these genes. Hence, the metabolic enzyme thymidylate synthase may have a critical role in regulating the cell cycle and the process of programmed cell death through its regulatory effects on the expression of cell-cycle-related proteins. It is also a target for several anti-cancer drugs, and its ability to function as a translational regulator may have important consequences for the development of cellular resistance to such drugs. A list of some proteins with more than one function can be found in Figure 4-47.

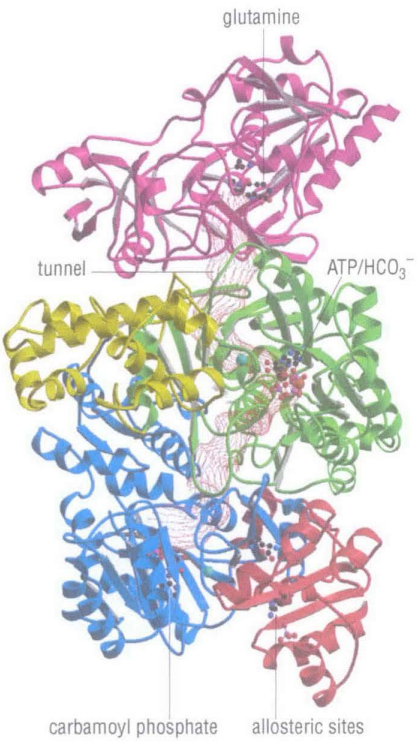


Figure 2-45 Three consecutive reactions are catalyzed by the three active sites of the enzyme carbamoyl phosphate synthetase Carbamoyl phosphate synthetase is a trifunctional enzyme with two long tunnels connecting the three active sites. In the first active site, shown at the top of the figure, glutamine is hydrolyzed to ammonia, which migrates through the first tunnel in the interior of the protein to the second active site, shown in the middle, where it reacts with carboxyphosphate to produce a carbamate intermediate. (The carboxyphosphate is formed by phosphorylation of bicarbonate by ATP.) The carbamate intermediate diffuses through the second tunnel to the third active site, shown at the bottom of the figure, where it is phosphorylated by another ATP molecule to give the final product, carbamoyl phosphate, which escapes from the enzyme. Graphic kindly provided by Frank Raushel and Hazel Holden.

©2004 New Science Press Ltd

3

Control of Protein Function

In the cell, precise regulation of protein function is essential to avoid chaos. This chapter describes the most important molecular mechanisms by which protein function is regulated in cells. These range from control of a protein's location and lifetime within the cell to the binding of regulatory molecules and covalent modifications such as phosphorylation that rapidly switch protein activity on or off. Also covered here are the nucleotide-driven switches in conformation that underlie the action of motor proteins and that regulate many signal transduction pathways.

3-0 Overview: Mechanisms of Regulation
3-1 Protein Interaction Domains
3-2 Regulation by Location
3-3 Control by pH and the Redox Environment
3-4 Effector Ligands: Competitive Binding and Cooperativity
3-5 Effector Ligands: Conformational Change and Allostery
3-6 Protein Switches Based on Nucleotide Hydrolysis
3-7 GTPase Switches: Small Signaling G Proteins
3-8 GTPase Switches: Signal Relay by Heterotrimeric GTPases
3-9 GTPase Switches: Protein Synthesis
3-10 Motor Protein Switches
3-11 Regulation by Degradation
3-12 Control of Protein Function by Phosphorylation
3-13 Regulation of Signaling Protein Kinases: Activation Mechanism
3-14 Regulation of Signaling Protein Kinases: Cdk Activation
3-15 Two-Component Signaling Systems in Bacteria
3-16 Control by Proteolysis: Activation of Precursors
3-17 Protein Splicing: Autoproteolysis by Inteins
3-18 Glycosylation
3-19 Protein Targeting by Lipid Modifications
3-20 Methylation, *N*-acetylation, Sumoylation and Nitrosylation

3-0 Overview: Mechanisms of Regulation

Protein function in living cells is precisely regulated

A typical bacterial cell contains a total of about 250,000 protein molecules (comprising different amounts of each of several thousand different gene products), which are packed into a volume so small that it has been estimated that, on average, they are separated from one another by a distance that would contain only a few molecules of water. It is likely that eukaryotic cells are at least as densely packed. In this crowded environment, precise regulation of protein function is essential to avoid chaos. Regulation of protein function *in vivo* tends to occur by many different mechanisms, which fall into several general classes. Protein function can be controlled by localization of the gene product and/or the species it interacts with, by the covalent or noncovalent binding of **effector** molecules, and by the amount and lifetime of the active protein.

Proteins can be targeted to specific compartments and complexes

Not all proteins are absolutely specific, and many also have more than one function. Consequently, it is often undesirable to have such proteins distributed everywhere in the cell, where they may carry out unwanted reactions. A simple way to regulate their activity is to ensure that the protein is only present in its active form in the specific compartment where it is needed, or when bound in a complex with other macromolecules that participate in its function. There are many ways in which specific localization can be achieved. Proteins can be targeted to cellular compartments by so-called signal sequences that are an intrinsic part of the encoded amino-acid sequence, or by attachment of, for example, a lipid tail that inserts into membranes. They can be directed to a complex of interacting proteins by a structural *interaction domain* that recognizes some covalent modification such as phosphorylation on another protein. Localization is a dynamic process and a given protein may be targeted to different compartments at different stages of the cell cycle: many transcription factors, for example, cycle between the nucleus and the cytosol in response to extracellular signals. When the protein is not in the location where it is needed, very often it is maintained in an inactive conformation.

Protein activity can be regulated by binding of an effector and by covalent modification

Protein activity can also be controlled by the binding of effector molecules, which often work by inducing conformational changes that produce inactive or active forms of the protein. Effectors may be as small as a proton or as large as another macromolecule. Effectors may bind noncovalently or may modify the covalent structure of the protein, reversibly or irreversibly. Effectors that regulate activity by binding to the active site usually take the form of inhibitors that compete with the substrate for binding. Often, the product of an enzyme reaction can act as such a **competitive inhibitor**, allowing the enzyme to regulate itself when too much product might be made. Ligands, including reaction products, may also bind to sites remote from the active site and in so doing either activate or inhibit a protein. Proteins regulated in this way tend to be oligomeric and *allosteric*. Allosteric proteins have multiple ligand-binding sites and these show *cooperativity* of binding: in *positive cooperativity* the first ligand molecule to bind is bound weakly, but its binding alters the conformation of the protein in such a way that binding of the second and subsequent ligand molecules is promoted. Cooperativity may also be negative: the first ligand binding weakens and thereby effectively inhibits subsequent binding to the other sites. Metabolic pathways often employ allosteric effectors as part of a feedback control mechanism: the end product of the pathway acts as an allosteric inhibitor of one of the earlier

Definitions

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

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.

References

Goodsell, D.S.: **Inside a living cell.** *Trends Biochem. Sci.* 1991, **16**:203–206.

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, Cambridge University Press, 2002).

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

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

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

Sato, T.K. *et al.*: **Location, location, location: membrane targeting directed by PX domains.** *Science* 2001, **294**:1881–1885.

©2004 New Science Press Ltd

enzymes in the pathway, so when too much of this product is synthesized it feeds back and shuts off one of the enzymes that help make it, as we shall see later in this chapter.

Binding of effector molecules can be covalent or can lead to covalent changes in a protein. The most common form of post-translational covalent modification is reversible phosphorylation on the hydroxyl group of the side chains of serine, threonine or tyrosine residues, but many other modifications are known, including side-chain methylation, covalent attachment of carbohydrates and lipids, amino-terminal acetylation and limited proteolytic cleavage, in which proteases cut the polypeptide chain in one or more places. Modifications such as phosphorylation or proteolytic cleavage may either activate or inactivate the protein.

Signal amplification is an essential feature in the control of cell function and covalent modification of proteins is the way such amplification is usually achieved. Often, an extracellular stimulus is of short duration and involves only a very low concentration, or a small change in concentration, of a hormone or regulatory molecule. Yet the cellular response must not only be very rapid; in many cases it must be massive, including a change in activities of many enzymes and alteration in the transcription of many genes. Covalent modification of a protein provides a simple mechanism by which a regulatory signal can produce a very large output. A single molecule of an enzyme whose catalytic activity is turned on by a covalent modification can process many thousands of substrate molecules. And if that substrate is another enzyme, the amplification is further magnified. When covalent modification of one enzyme causes it to become active so that it can, in turn, covalently modify and activate another enzyme and so on, a regulatory cascade is set up that leads to enormous, rapid changes in the final output. Blood clotting is an example of such an amplification cascade based on proteolysis.

Protein activity may be regulated by protein quantity and lifetime

The activity of a protein can also be regulated by controlling its amount and lifetime in the cell. This control may be exercised at several places in the flow of information from gene to protein. At its simplest, the amount of protein can be set by the level of transcription, which in turn can be controlled by, for example, the strength of the promoter or the action of a transcription factor, which may be a repressor or activator. The level of mRNA may also be adjusted after transcription by varying the rate of RNA degradation. At the level of the protein, quantities are controlled by the lifetime of the molecule, which is determined by its rate of degradation. The rate of turnover varies considerably from protein to protein; there are several specific mechanisms for targeting protein molecules to degradative machinery in the cell, including covalent attachment of the small protein ubiquitin.

A single protein may be subject to many regulatory influences

These various strategies are not mutually exclusive and any one protein may be subject to several of them. Coordination and integration of regulatory signals is achieved largely through signal transduction networks that set the balance of activities and thereby the balance of metabolism and cell growth and division pathways. Integration of signaling pathways is achieved through proteins, such as protein kinases, whose activity is under the control of several different mechanisms that can be independently regulated by incoming signals, as in the example of the protein tyrosine kinases known as cyclin-dependent kinases (CDKs), which control a eukaryotic cell's progression through the cell cycle (Figure 3-1).

Wyrick, J.J. and Young, R.A.: **Deciphering gene expression regulatory networks.** *Curr. Opin. Genet. Dev.* 2002, **12**:130–136.

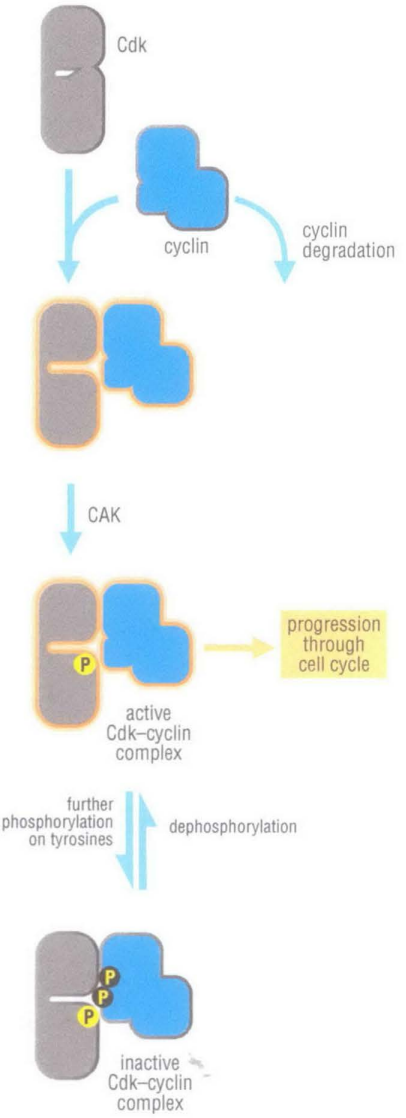


Figure 3-1 The cyclin-dependent protein kinases that control progression through the cell cycle are regulated by a number of different mechanisms Activation of a cyclin-dependent kinase (Cdk) at the appropriate point in the cell cycle requires both binding by its cyclin ligand, which induces a conformational change in the Cdk, phosphorylation of the Cdk in this Cdk–cyclin complex at a particular tyrosine residue, and dephosphorylation at one or two other tyrosines (depending on the particular Cdk). Cyclin binding is determined by the levels of cyclin present in the cell, which are strictly controlled in a temporal fashion by a signaling network that regulates gene transcription and protein degradation. Phosphorylation at the activating site is carried out by a so-called Cdk-activating kinase (CAK) which is present throughout the cell cycle but cannot act until the Cdk has bound cyclin; phosphorylation and dephosphorylation of the inhibitory tyrosines are carried out by another kinase and a phosphatase, respectively, whose activities are also subject to finely tuned regulation.

©2004 New Science Press Ltd

3-1 Protein Interaction Domains

The flow of information within the cell is regulated and integrated by the combinatorial use of small protein domains that recognize specific ligands

Many proteins that regulate cell behavior are constructed in a modular fashion from a number of different small domains with distinct binding specificities and functions (see Figure 1-46). Signal transduction proteins, regulatory proteins of the cell cycle, proteins that carry out targeted proteolysis, and proteins that regulate secretory pathways, cytoskeletal organization or gene expression all come into this category. Many gene regulatory proteins, for example, bind both to DNA and to another protein to carry out their function; they are composed of a domain that binds a specific DNA sequence and a protein-binding domain, which may target another molecule of the same protein (see Figure 1-41) or another gene regulatory protein. In some enzymes, a catalytic domain is attached to one or a number of protein-binding domains. These **interaction domains** (also called recognition modules) target the attached catalytic domain to a particular multiprotein complex or an appropriate subcellular location (such as the nucleus or the plasma membrane). Some interaction domains, such as the calcium-binding EF hand, often act as regulatory domains for enzymes. In signal transduction pathways, active multiprotein complexes are assembled by interaction domains that target the components to the complex. For example, the WD40 domain of the β-subunit of the heterotrimeric G protein targets this subunit to the α and γ subunits (see Figure 3-15). In some cases the interaction domain of a protein may interact with that protein's own catalytic domain in an autoinhibitory manner. Binding of the interaction domain to another protein then relieves the inhibition and activates the enzyme. Some protein kinases are autoregulated this way (see section 3-13).

Interaction domains are independently folded modules, 35–150 residues in length, which can still bind their target ligands if expressed independently of their "host" protein. Their amino and carboxyl termini are close together in space, with the ligand-binding site being located on the opposite face of the module. This allows an interaction domain to be inserted into a loop region of a catalytic domain, for example, as in the case of mammalian phospholipase Cγ, where two SH2 and one SH3 domains are inserted within the catalytic domain without disturbing its fold or having their own binding sites blocked. This type of domain organization also allows different domains to be strung together in combinatorial fashion and still retain their function.

Interaction domains can be divided into distinct families whose members are related by sequence, structure, and ligand-binding properties (Figure 3-2). Different members of a family recognize somewhat different sequences or structures, providing specificity for the proteins into which they are inserted. Examples of interaction domain families are SH3, WW and EVH1, which recognize proline-rich sequences; SH2 and PTB, which recognize phosphotyrosine-containing sequences; and 14-3-3, FHA, PBD and WD40, which bind to phosphoserine and phosphothreonine motifs. These domains are all common in proteins of signal transduction pathways. The PH and FYVE domains that recognize phospholipids are found in several different pathways, including signal transduction pathways and in proteins that control the traffic between internal membrane-bound compartments of the cell. A given protein may contain one or more copies of several different recognition modules, occurring in a different order in different proteins. Among the GTPase-activating proteins, for example, GAP1 has a PH domain after its GAP domain, while p120 GAP has the opposite order. Some interaction domains also form homo- and hetero-oligomers (typically dimers), which creates yet another level of regulatory versatility and specificity.

Figure 3-2 Interaction domains The name of the particular example shown for each family is given below each structure, along with the function and specificity of the domain.

Definitions

interaction domain: a protein domain that recognizes another protein, usually via a specific recognition motif.

References

Elia, A.E. *et al.*: **Proteomic screen finds pSer/pThr-binding domain localizing Plk1 to mitotic substrates.** *Science* 2003, **299**:1228–1231.

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

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

Macias, M.J. *et al.*: **WW and SH3 domains, two different scaffolds to recognize proline-rich ligands.** *FEBS Lett.* 2002, **513**:30–37.

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

Stenmark, H. *et al.*: **The phosphatidylinositol 3-phosphate-binding FYVE finger.** *FEBS Lett.* 2002, **513**:77–84.

Yaffe, M.B.: **Phosphotyrosine-binding domains in signal transduction.** *Nat. Rev. Mol. Cell Biol.* 2002, **3**:177–186.

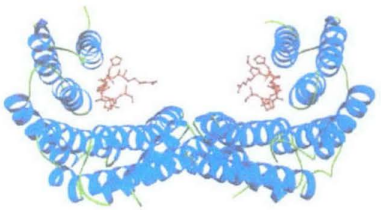
Yaffe, M.B.: **How do 14-3-3 proteins work? Gatekeeper phosphorylation and the molecular anvil hypothesis.** *FEBS Lett.* 2002, **513**:53–57.

Yaffe, M.B. and Elia, A.E.: **Phosphoserine/threonine-binding domains.** *Curr. Opin. Cell Biol.* 2001, **13**:131–138.

For a catalog of known intracellular protein interaction domains, see http://www.cellsignal.com/reference

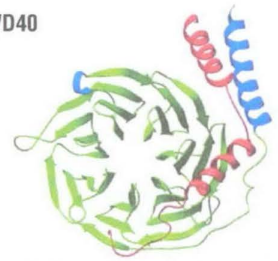
©2004 New Science Press Ltd

14-3-3



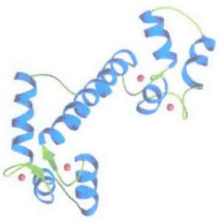
Example: 14-3-3
Function: protein–protein interactions
Specificity: phosphotyrosine

WD40



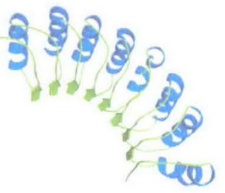
Example: G protein beta subunit
Function: protein–protein interactions; a stable propeller-like platform to which proteins bind either stably or reversibly
Specificity: various

EF-hand



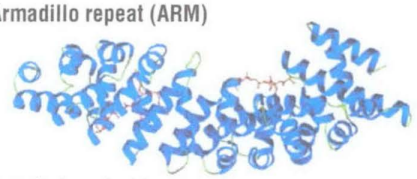
Example: Calmodulin
Function: calcium binding
Specificity: Ca^{2+}

LRR



Example: Rpn1
Function: protein–protein interactions
Specificity: various

Armadillo repeat (ARM)

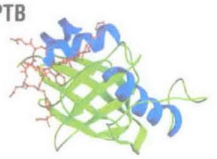


Example: Importin alpha
Function: protein–protein interactions
Specificity: various

SNARE

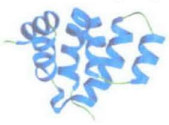
Example: SNAP-25B
Function: protein–protein interactions in intracellular membrane fusion
Specificity: other SNARE domains

PTB



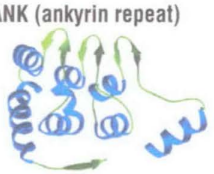
Example: Shc
Function: protein–protein interactions
Specificity: phosphotyrosine

Death domain (DD)



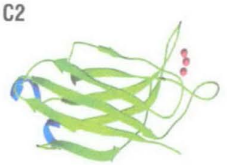
Example: FADD
Function: protein–protein interactions in pathway that triggers apoptosis
Specificity: other DD domains through heterodimers

ANK (ankyrin repeat)



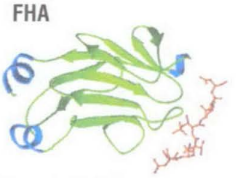
Example: Swi6
Function: protein–protein interactions
Specificity: various

C2



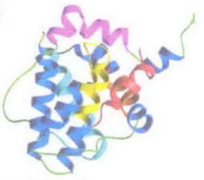
Example: PKC
Function: electrostatic switch
Specificity: phospholipids

FHA



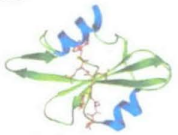
Example: Rad53
Function: protein–protein interactions
Specificity: phosphotyrosine

BH



Example: Bcl-XI
Function: protein–protein interactions
Specificity: Other BH domains through heterodimers

SH2

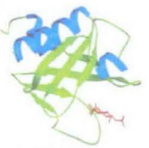


Example: Src
Function: protein–protein interactions
Specificity: phosphotyrosine

SH3

Example: Sem5
Function: protein–protein interactions
Specificity: proline-rich sequences

PH

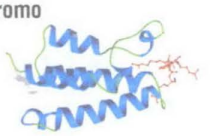


Example: PLC-δ
Function: recruitment of proteins to the membrane
Specificity: phosphoinositides

SAM

Example: EphA4
Function: protein–protein interactions via homo- and heterodimers
Specificity: other SAM domains

Bromo



Example: P/CAF
Function: protein–protein interactions in chromatin remodeling
Specificity: acetylated lysine

PDZ

Example: PSD-95
Function: protein–protein interactions, often involving transmembrane proteins or ion channels
Specificity: -XXXV/I-COOH

GYF

Example: CD2
Function: protein–protein interactions
Specificity: proline-rich sequences

Chromo

Example: Mouse modifier protein 1
Function: protein–protein interactions in chromatin remodeling
Specificity: methylated lysine

FYVE

Example: Vps27p
Function: Regulation of signaling
Specificity: phosphatidyl-inositol-3-phosphate

RING finger

Example: c-Cbl
Function: protein–protein interactions in ubiquitin-dependent degradation and transcription regulation
Specificity: various

WW

Example: Pin1
Function: protein–protein interactions
Specificity: proline-rich sequences

LIM

Example: CRP2
Function: protein–protein interactions, usually in transcription regulation
Specificity: various

F-box

Example: Skp2
Function: protein–protein interactions in ubiquitin-dependent protein degradation
Specificity: various

C1

Example: PKC
Function: recruitment of proteins to the membrane
Specificity: phospholipids

Fibronectin

Example: Fibronectin III
Function: protein–protein interactions in cell adhesion to surfaces
Specificity: RGD motif of integrins

©2004 New Science Press Ltd

Protein function in the cell is context-dependent

Many cellular processes involve the interaction of two or more macromolecules: signal transduction pathways are a good example. But when one estimates the number of gene products apparently involved in such pathways, it frequently appears that there are too few different proteins to account for all the different specific interactions that must be made. Put another way, the genomes of higher organisms seem to contain too few genes to fulfill all the cellular functions required. The logical conclusion is that many proteins participate in more than one cellular process. But if chaos is not to result from all these activities occurring simultaneously, both temporal and spatial control over a protein's activity must be exercised. Temporal control can be achieved partly by regulating gene expression and protein lifetime. However, it is increasingly clear that spatial context, the precise location within the cell at which a gene product exercises its biochemical function, is a major mechanism for regulating function.

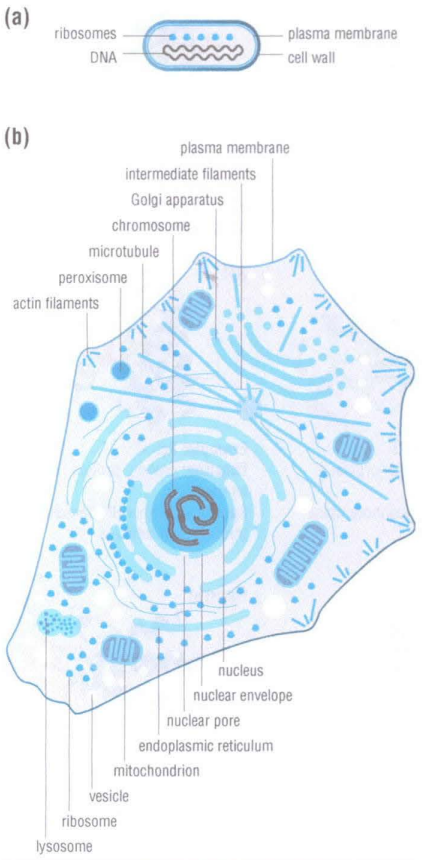


Figure 3-3 The internal structure of cells Schematic diagrams of **(a)** a bacterial cell and **(b)** a typical eukaryotic cell, showing the arrangements of some of the organelles and other internal structures.

It is probable that there is no such thing as a free-floating protein in a eukaryotic cell. Every protein is constrained, whether in a complex with other macromolecules, within a specific organelle, in a cargo vesicle, by attachment to a membrane, or as a passenger on the actin railroads in the cytoskeleton, among others. Moreover, the membrane-bounded vesicles of the cell provide a distinctive environment that can affect protein function, as we discuss in the next section. Prokaryotic cells, which lack organelles and cytoskeletal structures, may be less highly structured, but eukaryotic cells are organized into many compartments (Figure 3-3). In fact, it is increased organization, not increased gene number, that is the real hallmark of eukaryotic cells. For example, the eukaryote fission yeast *Schizosaccharomyces pombe* has fewer genes than the bacterium *Pseudomonas aeruginosa*.

Precise localization of proteins is a central feature of both spatial and temporal organization. In eukaryotic cells, the targeting of proteins to different locations in the cell, such as cytosol versus nucleus, can be regulated according to need, and changes in targeting can modify protein function at the cellular level even when the biochemical function of the protein does not change.

Nowhere is this more evident than in signal transduction pathways. Many of these depend on **protein kinases** which, as discussed later in this chapter, modify protein function by phosphorylating selected amino acids. The number of protein kinase genes in the human genome is large, but is far smaller than the number of potential protein kinase substrates. As we do not appear to possess one kinase for each substrate, kinases must have less than absolute specificity. How then are they prevented from phosphorylating the "wrong" protein at an inappropriate time and place? One answer is to target the kinase to the same location as its "correct" substrate, a location different from that for any other potential substrate; the action of that kinase is then quite specific; and specificity can be altered if required by relocation of kinase and/or substrate.

Another example of control by location is the small monomeric GTPase Tem1 from the budding yeast *Saccharomyces cerevisiae*. Tem1 plays an essential part in terminating the mitotic phase of the cell cycle. The interaction of individual yeast proteins with all the other proteins encoded in the yeast genome has been systematically investigated by two-hybrid analysis (see section 4-4). In such assays Tem1 has been found to interact physically with 24 different yeast gene products. Given its size, no more than about four other proteins could possibly bind to Tem1 at the same time, so what controls which proteins it interacts with at any given time? Differences in the timing of expression of the different target proteins could play some part in determining Tem1's specificity, but binding to specific partners is likely to be mainly controlled by targeting Tem1, and its potential partners, to different locations in the cell at different times.

Definitions

lipid anchor: lipid attached to a protein that inserts into a membrane thereby anchoring the protein to the bilayer.

protein kinase: enzyme that transfers a phosphate group from ATP to the OH group of serines, threonines and tyrosines of target proteins.

References

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

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

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

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.

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

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

Penn, R.B. *et al.*: **Regulation of G protein-coupled receptor kinases.** *Trends Cardiovasc. Med.* 2000, **10**:81–89.

Sato, T.K. *et al.*: **Location, location, location: membrane targeting directed by PX domains.** *Science* 2001, **294**:1881–1885.

©2004 New Science Press Ltd

There are several ways of targeting proteins in cells

Targeting of a protein within the cell is achieved in three main ways: by sequences in the protein itself; by various types of post-translational modification; and by binding to a scaffold protein. Localization to the membrane-bounded compartments of the cell—the nucleus, the endoplasmic reticulum (ER), the Golgi complex, and so on—is usually achieved by specific "localization signals" encoded in the protein sequence. For example, the sequence KDEL targets proteins for retention in the endoplasmic reticulum (by binding to a specific receptor) (Figure 3-4a); lysine/arginine-rich clusters such as KRKR target proteins to the nucleus; and the hydrophobic signal sequence found at the beginning of some proteins directs their secretion from the cell. Depending on the organelle to be targeted, localization sequences may be located at either end of the polypeptide chain or even internally. Intrinsic plasma membrane proteins are also targeted to the membrane via the ER by this type of mechanism.

Localization by sequence motifs is an intrinsic property of the protein. In contrast, post-translational targeting mechanisms are regulatable. The commonest mechanism is covalent chemical modification of the protein. Perhaps the most widely used modification of this type is phosphorylation of tyrosine, serine or threonine residues by protein kinases; such modifications target proteins for binding by other proteins that have a recognition site for the modification (Figure 3-4b). This mechanism is particularly important in the formation of signaling complexes and in biochemical switches, and we discuss it in more detail later in this chapter.

Another important post-translational modification targets proteins to the plasma membrane by covalent attachment of their carboxyl or amino terminus to a lipid molecule, often called a **lipid anchor**, which is then inserted into the phospholipid bilayer (Figure 3-4c). For example, the signaling GTPase Ras, which is found mutated in many human tumors, becomes covalently attached to the isoprenoid farnesyl at its carboxyl terminus via a thioester linkage, and this modification, which targets Ras to the cell membrane and thus brings it close to the signaling complexes attached to receptor proteins, is essential for its function in signal transduction (Figure 3-4c).

Although lipid anchors could insert randomly into membranes solely by virtue of their hydrophobicity, evidence is mounting that the membranes in eukaryotic cells are not just random soups of lipids. Membranes appear to have many patches (sometimes called lipid rafts) where specific lipids congregate. These "islands" target particular lipid anchors or lipid-binding domains on proteins not just to the membrane but to very specific places on the membrane. Control of the location and size of these rafts by enzymatic modification and hydrolysis of phospholipids is yet another level of functional organization. We discuss lipid rafts in more detail later, in the section on lipid modification of proteins (see section 3-19), in which we also describe the various types of lipid anchor.

The third common means of protein targeting is by binding to scaffold proteins. These are proteins that can bind several other proteins simultaneously, thereby promoting their interaction (Figure 3-4d). To carry out this function, scaffold proteins generally contain a number of small domains that recognize specific targets, modifications or sequence motifs. One example is the SH3 domain, which binds to proline-rich regions; another is the SH2 domain, which recognizes sequences containing a phosphorylated tyrosine; yet another is the PH domain, which binds to phospholipids in membranes (see section 3-1).

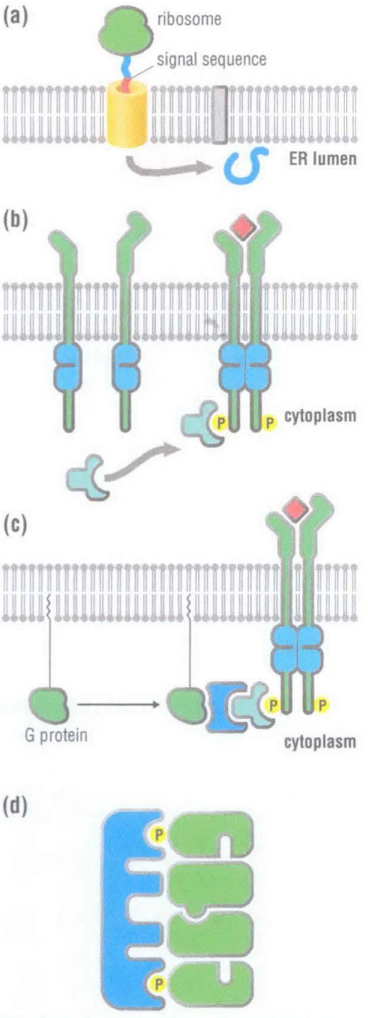


Figure 3-4 Mechanisms for targeting proteins **(a)** Proteins destined for export from the cell are first co-translationally targeted to the endoplasmic reticulum (ER) by signal sequences (red) at the amino terminus of the protein. The signal sequences are cleaved off when protein synthesis is complete, releasing the protein into the lumen of the ER, from which it is exported to the cell surface in transport vesicles. Transmembrane proteins are targeted in a similar way, but have additional internal hydrophobic signal sequences that retain the protein in the membrane. **(b)** Proteins may be targeted to specific signaling complexes by recognizing phosphorylated sites on their target protein. As shown here, a cell-surface receptor (blue and green) with intrinsic protein kinase activity dimerizes on binding an external ligand (red), and this triggers phosphorylation of the cytoplasmic domains. That generates a target site for a cytoplasmic signaling molecule (light blue) that recognizes the phosphorylated residue and binds to the receptor tail. **(c)** A lipid tail is attached to a small G protein (green), thereby anchoring it in the plasma membrane where it can interact with other membrane-bound signaling proteins. It is shown here interacting with an adaptor protein (dark blue) that is part of the signaling complex that builds up around phosphorylated receptor domains. **(d)** A scaffold protein (blue) binds to several different signaling molecules (green) and thereby targets their activities: the signaling molecules may, for example, be protein kinases that can now sequentially phosphorylate one another as part of a signaling pathway.

©2004 New Science Press Ltd

3-3 Control by pH and the Redox Environment

Protein function is modulated by the environment in which the protein operates

All proteins are adapted to fold and function optimally in the particular environment of the cellular compartment in which they operate. The cellular aqueous solution is highly viscous and contains many components besides proteins at high concentration, including ions, free polar organic molecules and, most important, the dissociated conjugate acid/conjugate base components of water: the proton and the hydroxide ion. If these two components are present at approximately equal concentration, as is the case in the cytosol of most cells, the solution is neutral with a pH of about 7. If protons are in excess, the solution is acidic, with a lower pH, and ionizable groups tend to be protonated. If hydroxide ions predominate, the solution is alkaline with a pH > 7 and ionizable groups will tend to be deprotonated. Distinct membrane-bounded compartments inside the cell often have a distinct internal microenvironment and the extracellular environment represents a different aqueous environment again from that of the interior. We describe here some examples of the adaptations of proteins for the environments in which they function.

Changes in redox environment can greatly affect protein structure and function

The interiors of cells are for the most part reducing environments: they furnish electrons, often in the form of hydrogen atoms. On the other hand, outside the cell, proteins and small molecules are typically exposed to an oxidizing environment in which electrons can be lost. The chief effect of this difference is that cysteine residues in proteins are usually fully reduced to –SH groups inside the cell, but are readily oxidized to disulfide S–S bridges when a protein is secreted. Cells can exploit this difference to trigger oligomerization by S–S bond formation in secreted proteins, or subunit dissociation and conformational changes when proteins are internalized. For instance, acetylcholinesterase is synthesized as a monomer, but when secreted from muscle and nerve cells self-associates to form dimers and tetramers. These oligomers, which are more stable than the monomer and thus better able to survive outside the cell, are composed of covalently linked subunits. Each monomer has a carboxy-terminal cysteine residue which forms an intersubunit S–S bond with the identical residue on a neighboring polypeptide chain.

Changes in pH can drastically alter protein structure and function

Most cytosolic fluid is maintained at near neutral pH, so that neither acids nor bases predominate. However, there are specialized compartments, such as endosomal vesicles, where the pH is quite acidic. As the surfaces of soluble proteins are chiefly composed of polar side chains, many of which are ionizable, both the net charge on a protein and the distribution of charge over the surface can vary considerably with pH.

If ligand binding depends on electrostatic interactions (see Figure 2-23), changes in the external pH (or ion concentration) can greatly influence binding strength by directly altering the ionization states of groups that interact with the ligand or of groups on the ligand itself. Modulation of the surface charge distribution of a protein by pH changes can also affect the biochemical function indirectly, by changing the extent of ionization of essential functional groups in an active site or binding site through long-range electrostatic interactions. For instance, endosomal proteases, which degrade internalized proteins, are only catalytically active

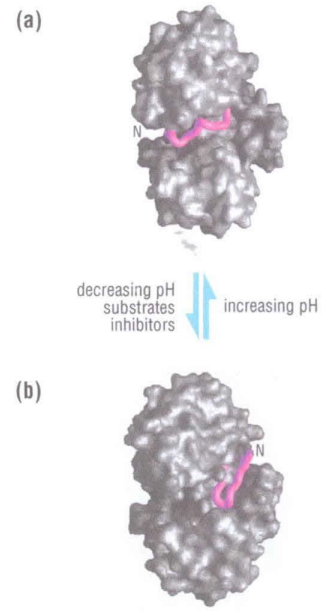


Figure 3-5 Cathepsin D conformational switching Cathepsin D undergoes a pH-dependent conformational change. At the neutral pH of the cytosol **(a)**, the amino-terminal peptide interacts with the active site and blocks substrate access. At the low pH of the endosome **(b)**, in which cathepsin D is active, the amino-terminal segment reorients, opening the active site. At the same time, the low pH sets the protonation states of the catalytic residues for activity. Graphic kindly provided by John Erickson. From Lee, A.Y. *et al.*: *Nat. Struct. Biol.* 1998, **5**:866–871, with permission.

References

Authier, F. *et al.*: **Endosomal proteolysis of internalized insulin at the C-terminal region of the B chain by cathepsin D.** *J. Biol. Chem.* 2002, **277**:9437–9446.

Authier, F. *et al.*: **Endosomal proteolysis of internalized proteins.** *FEBS Lett.* 1996, **389**:55–60.

Lee, A.Y. *et al.*: **Conformational switching in an aspartic proteinase.** *Nat. Struct. Biol.* 1998, **5**:866–871.

Weiss, M.S. *et al.*: **Structure of the isolated catalytic domain of diphtheria toxin.** *Biochemistry* 1995, **34**:773–781.

©2004 New Science Press Ltd

at acidic pH, when the charged groups in their active sites are in the proper ionization states. Endosomal proteases are implicated in the degradation of internalized regulatory peptides involved in the control of metabolic pathways and in the processing of intracellular antigens for cytolytic immune responses. Processing occurs in endocytic vesicles whose acidic internal environment is regulated by the presence of an ATP-dependent proton pump. The acidic environment (pH ~ 5) modulates protease activity, protein unfolding and receptor–ligand interactions. The endosomal compartment of liver cells contains an acidic endopeptidase, cathepsin D, that hydrolyzes internalized insulin and generates the primary end-product of degradation. The protease is only active at low pH because its active site contains two essential aspartic acid residues, one of which must be protonated for catalysis to occur.

However, there is an additional aspect to the pH-dependent regulation of cathepsin D activity. The conformation of the protein, particularly the amino terminus, differs at low and high pH. High pH results in a concerted set of conformational changes including the relocation of the amino terminus into the active site, where it blocks substrate binding; also, at this pH both catalytic aspartates are likely to be deprotonated and therefore inactive; their charge state also stabilizes this conformation of the amino terminus. Activation at low pH is due to the protonation of one of the two catalytic aspartates, weakening their electrostatic interactions with the amino-terminal peptide. The released amino-terminal peptide restores the accessibility of the active site to substrate (Figure 3-5).

In some cases, the effect of pH has been exploited to change the structure of a protein completely in an acidic compartment. For example, diphtheria toxin, a protein that is among the deadliest substances known, is synthesized in the neutral pH of the cytosol of the bacterial cell as a single polypeptide chain with three domains: from amino to carboxyl terminus these are designated as domains B, T and A. The B domain binds to a receptor in the target-cell membrane, leading to uptake of the toxin by receptor-mediated endocytosis. The A domain, which is joined to the B domain by a disulfide bond, is an enzyme that kills cells by catalyzing the ADP-ribosylation of elongation factor 2 on the ribosome, leading to a block in protein synthesis. The T domain is the most interesting, because it is responsible for delivering the catalytic domain into the cytoplasm of the target cell.

Endocytosis through binding of the B domain to the cell-surface receptor results in the sequestration of the toxin in the acidic interior of an endocytic vesicle. Exposure to the reducing environment inside the endocytic vesicle breaks the disulfide bond between the A and B domains, releasing the toxic A domain. At the same time, the acidic environment causes a massive conformational change in the T domain, which turns almost inside out, exposing the interior hydrophobic residues. Hydrophobic side chains, as we saw in Chapter 1 (section 1-9) are normally buried in soluble proteins, extensive exposed hydrophobic regions occurring only in the lipid interior of membranes (see section 1-11). Thus the exposure of hydrophobic residues of the diphtheria toxin T domain results in its insertion into the endosomal membrane, where it is thought to create a channel through which the toxic A domain is translocated into the cytoplasm (Figure 3-6).

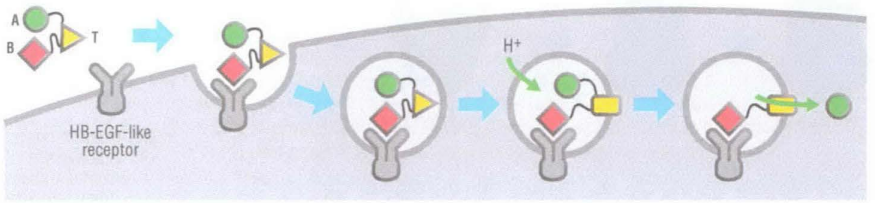


Figure 3-6 Schematic representation of the mechanism by which diphtheria toxin kills a cell The toxin has a receptor-binding domain (B), a membrane translocation domain (T) and a catalytic subunit (A) that kills cells by ADP-ribosylating a residue on the ribosomal protein EF-2. After receptor binding, the toxin enters the cell by endocytosis. In the low-pH environment of the endosome, a disulfide bond (not shown) in the toxin is reduced and a conformational change occurs that activates the membrane translocation domain, which is thought to release the active catalytic domain from the endosome. Weiss, M.S. *et al.*: *Biochemistry* 1995, **34**:773–781.

©2004 New Science Press Ltd

3-4 Effector Ligands: Competitive Binding and Cooperativity

Protein function can be controlled by effector ligands that bind competitively to ligand-binding or active sites

One of the most important ways in which the activity of proteins is controlled is by binding regulatory molecules, termed **effector ligands** or effectors, that alter the activity of the protein with which they interact. Effectors can be as small as a proton (pH-induced conformational changes can be enormous, as in the case of the diphtheria toxin T domain discussed in the preceding section), or as large as another protein. In some instances this regulation can simply be inhibition through competitive binding with the normal ligand. This is a common mechanism for inhibiting enzymes, in which the effector ligand binds to the active site instead of substrate: many metabolic enzymes are feedback-inhibited in this way by their own product or by the product of an enzyme downstream from them in the same metabolic pathway (Figure 3-7). Feedback inhibition ensures that the activity of an enzyme is diminished when there is an overabundance of its product in the cell. An example is the control by the tripeptide glutathione (GSH) of its own biosynthesis. GSH is ubiquitous in mammalian and other living cells. It has several important functions, including protection against oxidative stress. It is synthesized from its constituent amino acids by the consecutive actions of two enzymes. The activity of the first enzyme is modulated by feedback inhibition of the end product, GSH, ensuring that the level of glutathione does not exceed necessary values.

Cooperative binding by effector ligands amplifies their effects

In the previous example, binding of one molecule of a competitive feedback inhibitor inhibits one molecule of a target enzyme. But many physiological responses need to be rapid and total, which is hard to achieve with a linear system. Amplification would allow a single regulatory molecule to shut down many copies of a target protein or pathway. Amplification can be achieved in either of two ways. One is by covalent modification of the protein, which we discuss later in this chapter. The other is by **cooperativity**. This is a phenomenon of universal importance in biological systems and is as versatile as it is widespread. At the metabolic level, an enzyme in one pathway can cooperate with another pathway by providing a component that can serve as a substrate, enzyme or regulator of that pathway. We are concerned here, on the other hand, with **cooperative binding**, cooperativity between binding sites for the same

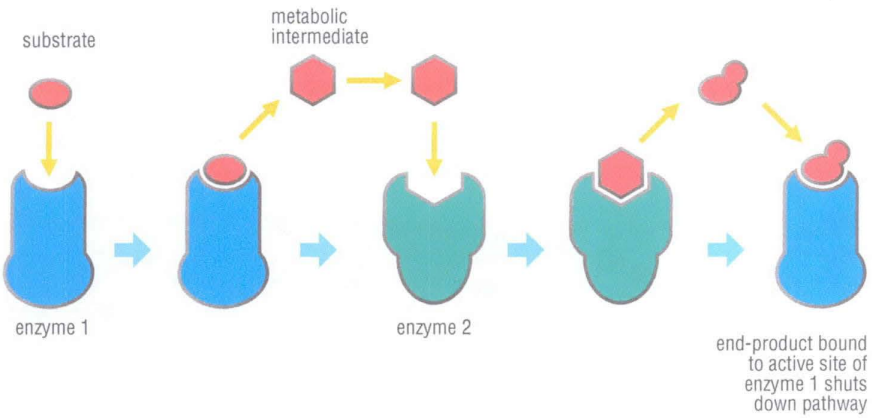


Figure 3-7 Competitive feedback inhibition In this simple two-enzyme pathway, the end-product of the pathway acts as a competitive inhibitor of the first enzyme.

Definitions

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.

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

effector ligand: a ligand that induces a change in the properties of a protein.

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.

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.

©2004 New Science Press Ltd

ligand on a protein, in which binding of the ligand at one site affects the ease or otherwise of binding of ligand at the other site(s). This type of cooperativity depends on the ligand–protein interaction resulting in a measurable conformational change in other regions, close by or distant, of that protein. Cooperativity is only present in oligomeric proteins, where there are two or more subunits each with a binding site for the ligand.

Cooperativity can be positive or negative: **positive cooperativity** means that binding of one molecule of a ligand to a protein makes it easier for a second molecule of that ligand to bind; **negative cooperativity** means that binding of the second molecule is more difficult. Thus cooperativity can amplify either the activation of a protein or its inhibition. If a protein has, say, four identical subunits, each of which has a binding site for an effector ligand, positive cooperativity occurs if binding of the first effector molecule to one subunit makes binding easier (that is, increases the binding constant) for the second molecule of the effector to the second subunit. In turn, this binding facilitates binding of a third ligand molecule to the third subunit even more, and so on. This is illustrated for a dimer schematically in Figure 3-8. Such positive cooperativity means that activation (or inhibition) of all subunits can be achieved at a lower concentration of the activating or inhibiting ligand than would be the case if each subunit bound the effector independently and with equal affinity. In the extreme case of absolute positive cooperativity, only a few molecules of ligand might be sufficient to activate or inhibit the protein completely. Cooperative binding is also seen in proteins that are not enzymes. The oxygen-transport protein hemoglobin, for example, binds molecular oxygen at its four binding sites with positive cooperativity.

Negative cooperativity in enzymes usually involves the binding of substrates or cofactors, and in this case the binding site in question is the active site. In the extreme case of absolute negative cooperativity, a dimeric enzyme, for example, may never be found with cofactor or substrate bound to both subunits at the same time. For example, the bacterial enzyme ATP-citrate lyase shows strong negative cooperativity with respect to citrate binding. Non-enzymes can display this property as well. For example, the heart drug bepridil binds to the muscle protein troponin C with negative cooperativity. The physiological significance of absolute negative cooperativity is uncertain.

The phenomenon of cooperativity reflects an important consequence of the flexibility of proteins: binding of even a small molecule to a protein surface can induce structural changes at a distance from the binding site. In the case of oligomeric proteins, as we have seen, this can result in communication of a ligand-induced conformational change from one subunit to another. In the next section, we shall see how this ability of proteins to undergo long-range structural adjustments enables regulatory effector ligands to act by binding at sites remote from the normal ligand-binding or active site.

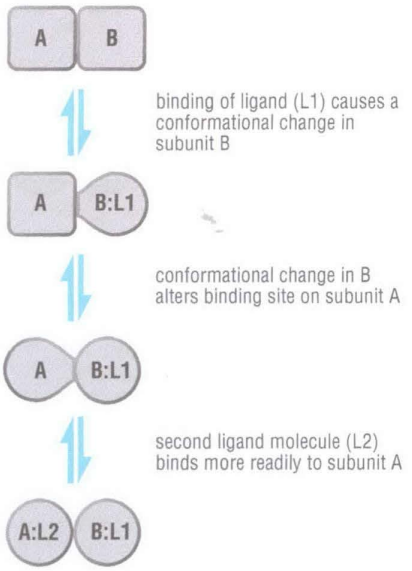


Figure 3-8 Cooperative ligand binding Shown here is a simple two-subunit protein displaying cooperative binding of a ligand L. The initial state of the protein (square) has a relatively low affinity for L. When the first molecule of L (L1) binds, it causes a conformational change in the subunit to which it is bound. But interactions between the subunits cause the unoccupied subunit to change its conformation as well, to one that is similar to the ligand-bound state (circle) and therefore has a higher affinity for L. Thus a lower concentration of L is required to bind to the second subunit, and the effect of L is amplified. If the protein is an enzyme, the effect of binding of the first ligand molecule may not only be to increase the affinity of the next subunit for ligand; its enzymatic activity may also be enhanced, increasing the amplifying effect.

References

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

©2004 New Science Press Ltd

3-5 Effector Ligands: Conformational Change and Allostery

Effector molecules can cause conformational changes at distant sites

Because of the close packing of atoms in globular protein structures, even small changes in side-chain and main-chain positions at one site can be propagated through the tertiary structure of the molecule and cause conformational changes at a distant location in the protein. Indeed, the most common type of regulatory effector ligand is one that is different from the normal substrate or functional ligand for the protein and which binds at a site distinct from the enzyme's catalytic site or from the site through which the protein's function is mediated (in the case of a non-enzyme). In hetero-oligomeric enzymes, the regulatory site is often located on a different subunit from the active site.

There are two extreme models for the action of such regulators and evidence exists for each in different cases. In one, the binding of successive effector molecules causes a sequential series of conformational changes from the initial state to the final state (Figure 3-9a). The other model postulates a preexisting conformational equilibrium between a more active and a less active form of the protein. The effector can only bind to one of these forms, and thus its binding shifts the equilibrium in a concerted manner in favor of the bound form (Figure 3-9b). Because structural changes are produced in both models, this form of regulation is called **allostery** from the Greek for "another structure", and a protein that is regulated in this way is called an allosteric protein. A molecule that stabilizes the more active form is an **allosteric activator**; one that stabilizes the less active form is an **allosteric inhibitor**. When the effector ligand is another protein, the protein that is regulated can be, and often is, monomeric, as in the case of the regulation of the monomeric cyclin-dependent kinase Cdk2 by the protein cyclin A, as described later in this chapter. When the effector ligand is a small molecule, however, although allosteric regulation does not formally require cooperative binding, in practice the allosteric protein is nearly always oligomeric, and binding of the effector ligand is usually cooperative (see section 3-4).

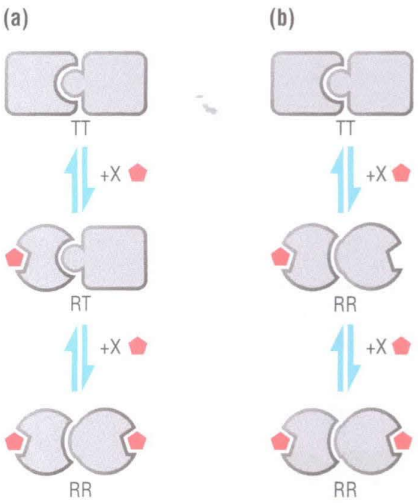


Figure 3-9 Two models of allosteric regulation (a) A sequential change in the conformation of subunits of a dimeric protein from a less active state (the "tense" state TT) to the more active state (the "relaxed" state RR) on binding of a regulatory ligand (X). **(b)** The case where the TT and RR states are in equilibrium and the regulatory ligand stabilizes one of them, shifting the equilibrium in favor of this form. To emphasize the conformational change, this diagram shows an asymmetrical protein; in reality, the subunits are usually identical.

ATCase is an allosteric enzyme with regulatory and active sites on different subunits

The allosteric enzyme aspartate transcarbamoylase (ATCase) is a hetero-oligomer made up of six catalytic and six regulatory subunits (Figure 3-10). ATCase catalyzes the formation of *N*-carbamoyl aspartate, an essential metabolite in the synthesis of pyrimidines, from carbamoyl phosphate and L-aspartate. The enzyme is allosterically inhibited by cytidine triphosphate (CTP), the end-product of pyrimidine nucleotide biosynthesis, and allosterically activated by adenosine triphosphate, the end-product of purine nucleotide biosynthesis. Thus, CTP is a feedback inhibitor and shuts down ATCase when pyrimidine levels are high; ATP activates the enzyme when purine levels are high and pyrimidines are needed to pair with them to make nucleic acids. Structural studies have shown that the enzyme exists in at least two states, with very different intersubunit contacts (Figure 3-10). The more compact state of the dodecamer (the tense or T state) is the less active form; it is stabilized by the binding of CTP to sites on the regulatory subunits. The more open arrangement (the relaxed or R state) is the more active enzyme; this structure is stabilized by binding of ATP, also to the regulatory subunits. Both these effectors bind cooperatively, as do the substrates, so there is great potential for amplification of the signals that regulate this enzyme. It is not surprising that the ATCase-catalyzed reaction is the major control step in pyrimidine biosynthesis.

Definitions

allosteric activator: a ligand that binds to a protein and induces a conformational change that increases the protein's activity.

allosteric inhibitor: a ligand that binds to a protein and induces a conformational change that decreases the protein's activity.

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.

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

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

References

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

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

White, A. *et al.*: **Structure of the metal-ion-activated diphtheria toxin repressor/tox operator complex.** *Nature* 1998, **394**:502–506.

©2004 New Science Press Ltd