12 – EXAMPLES OF STRUCTURE-FUNCTION RELATIONSHIPS IN ENZYMATIC SYSTEMS

As was shown in the preceding chapter, there is a great variety of enzymes catalysing very diverse chemical reactions. Some enzymes among the best known on the structural and functional level have been chosen in the present chapter to illustrate each of the reaction types. Enzymatic systems for which the three-dimensional structures are known both for the free enzyme and for the enzyme complexed to a substrate or to an analog are preferentially presented. Indeed, a precise knowledge of the position of substrate atoms with respect to enzyme catalytic groups is necessary for understanding reaction mechanisms; otherwise the amount of speculation remains too significant. However, even for particularly well described systems, the transition state and its interactions with the enzyme often remain hypothetical.

Group transfer reactions are treated by taking the examples of acyl transferases, serine proteases and more particularly chymotrypsin which is the paradigm, thiol proteases which include papain, acid proteases like pepsin, and a metallopeptidase, carboxypeptidase. Tyrosine tRNA synthetase and a kinase were chosen for phosphoryl group transfer reactions and lysozyme for glycosyl group transfer reactions. Two examples were retained for oxydoreduction reactions, alcohol dehydrogenase and flavocytochrome b2. Isomerisation reactions are illustrated by triose phosphate isomerase. Aspartate amino transferase, a pyridoxal phosphate enzyme, offers a well documented example of amino group transfer *via* the formation of a SCHIFF base. Aldolases, which induce the formation of a SCHIFF base over the course of the reaction that they catalyse, were retained to illustrate reactions of carbon-carbon bond breaking or formation.

Some of these enzymatic systems were already described under diverse aspects throughout this text. In this chapter, catalytic mechanisms are considered for each enzyme in close relation with its structure. At the risk of redundancy, it seemed useful to us to regroup all the significant relative data to a particular enzyme. Mechanisms of reactions catalysed by allosteric enzymes are presented in Part V in the chapter treating non-covalent regulations.

J. Yon-Kahn, G. Hervé, *Molecular and Cellular Enzymology*,
DOI 10.1007/978-3-642-01228-0_13, © Springer-Verlag Berlin Heidelberg 2010

12.1. PROTEASES

Proteases can be classified according to their activities and the nature of their functional groups. Most serine proteases are endopeptidases. With a few exceptions, thiol proteases are also endopeptidases; they possess a cysteine as the catalytic group which plays the same role as serine. These two types of proteases possess an optimal activity at neutral pH. Acid proteases or aspartyl proteases have carboxylic groups on two aspartyl residues at the catalytic site. Most of them are active at low pH; conversely, renin is active at neutral pH. Metalloproteases and Zn^{++} proteases are active at neutral pH. Among them, carboxypeptidases are exopeptidases and thermolysine is an endopeptidase.

12.1.1. SERINE PROTEASES

12.1.1.1. STRUCTURAL ASPECTS

There is in mammals a large number of serine proteases: trypsin, chymotrypsin, elastase, proteases of the blood coagulation cascade and among them thrombin. They are secreted as precursors or zymogens, their activation resulting from proteolytic cleavage. Many serine proteases of microorganismes are currently known, like subtilisine, α-lytic protease of *Myxobacter 495*, and that of *Streptococcus griseus*. Pancreatic proteases possess great similarities in sequence, from 40 to 45% identity and 50 to 55% homology. Structural analogs are even greater and extend to bacterial proteases of which the three-dimensional structure is known; however, these have a shorter polypeptide chain and only share 21% identity with the preceding ones.

The three-dimensional structure of several serine proteases and their zymogen is known at atomic resolution. Among the first resolved were α-chymotrypsin (BLOW, 1970; BIKTOFT & BLOW, 1972), γ-chymotrypsin (COHEN et al., 1969), chymotrypsinogen (KRAUT, 1971), pig elastase (SHOTTON & WATSON, 1970), trypsin (STROUD et al., 1974; FELDHAMMER & BODE, 1975), trypsinogen (BODE et al., 1976), protease B of *Streptomyces griseus* (DELBAERE et al., 1975), BPN' subtilisine (WRIGHT et al., 1970), Novo subtilisine (DRENTH et al., 1972) and α-lytic protease of *Myxobacter 495*. Since then, other protease structures have been determined and the list increases regularly. All these proteases possess the same catalytic groups. Those of the trypsin family are folded into two distinct domains stabilised by disulphide bridges; each of these domains is made of antiparallel β segments forming a β barrel (Fig. 12.1 opposite). They comprise a short helical segment at the C-terminal end. The principal structural differences are situated in the external loops. These structural similarities suggest that these diverse proteases derive from the same ancestor protein and that they diverged over the course of evolution by preserving their principal structural characteristics but developed different specificities. Subtilisines, although possessing the same catalytic groups, are not homologues of the preceding and present a different folding comprising an important α helix proportion. Carboxypeptidase II from wheat presents an even different structure; carboxypeptidase Y from yeast and two plant proteases belong to the same family.

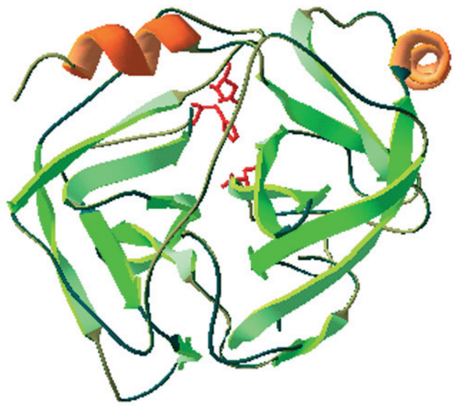


Fig. 12.1 Structure of α-chymotrypsin (from PDB: 3CHA, with SwissPDB viewer software, as well as all those referenced by their PDB identification) *The catalytic groups are indicated* in red

12.1.1.2. ACTIVATION OF ZYMOGENS

Most mammalian serine proteases are synthesised as prezymogens transformed into zymogens by elimination of the signal peptide when crossing the rough endothelial reticulum membrane and secreted under this form (see Chap. 8). Pancreatic proteases are found thus in the pancreas as inactive precursors; this situation constitutes a protection mechanism. Trypsinogen is activated into trypsin by enterokinase which cleaves the Lys6–Ile7 peptide bond, thus releasing the N-terminal hexapeptide. The activation of chymotrypsinogen is more complex. The zymogen is transformed into chymotrypsin by trypsin-catalysed breaking of the Arg15–Ile16 bond which gives rise to π-chymotrypsin. A second cleavage by active chymotrypsin thus formed is produced between Ile13 and Ser14 generating δ-chymotrypsin. Then in a third step, two other cleavages occur, those of Tyr146–Thr147 and Asn148–Ala149 bonds that give rise to α-chymotrypsin. Chymotrypsinogen itself can undergo proteolysis by chymotrypsin, giving rise to neochymotrypsinogens (Fig. 12.2 below). Proelastase is also activated by trypsin which cleaves the Arg15–Val16 bond (the numbering of amino acids is that of chymotrypsinogen).

The breaking of peptide bonds during zymogen activation drives the establishment of a salt bridge between the carboxylate of Asp194 adjacent to the reactive serine (Ser195) and the amino group of Ile16 (or Val16 in elastase); the formation of this salt bridge and its conformational and functional consequences were presented in Chap. 11. Although zymogens are inactive with respect to their physiological substrates that are peptide bonds, they present a weak activity for some specific esters such as p.nitrophenyl esters of different dipeptides as was shown by a study of the group of NEURATH [LONSDALE-ECCLES et al., 1978].

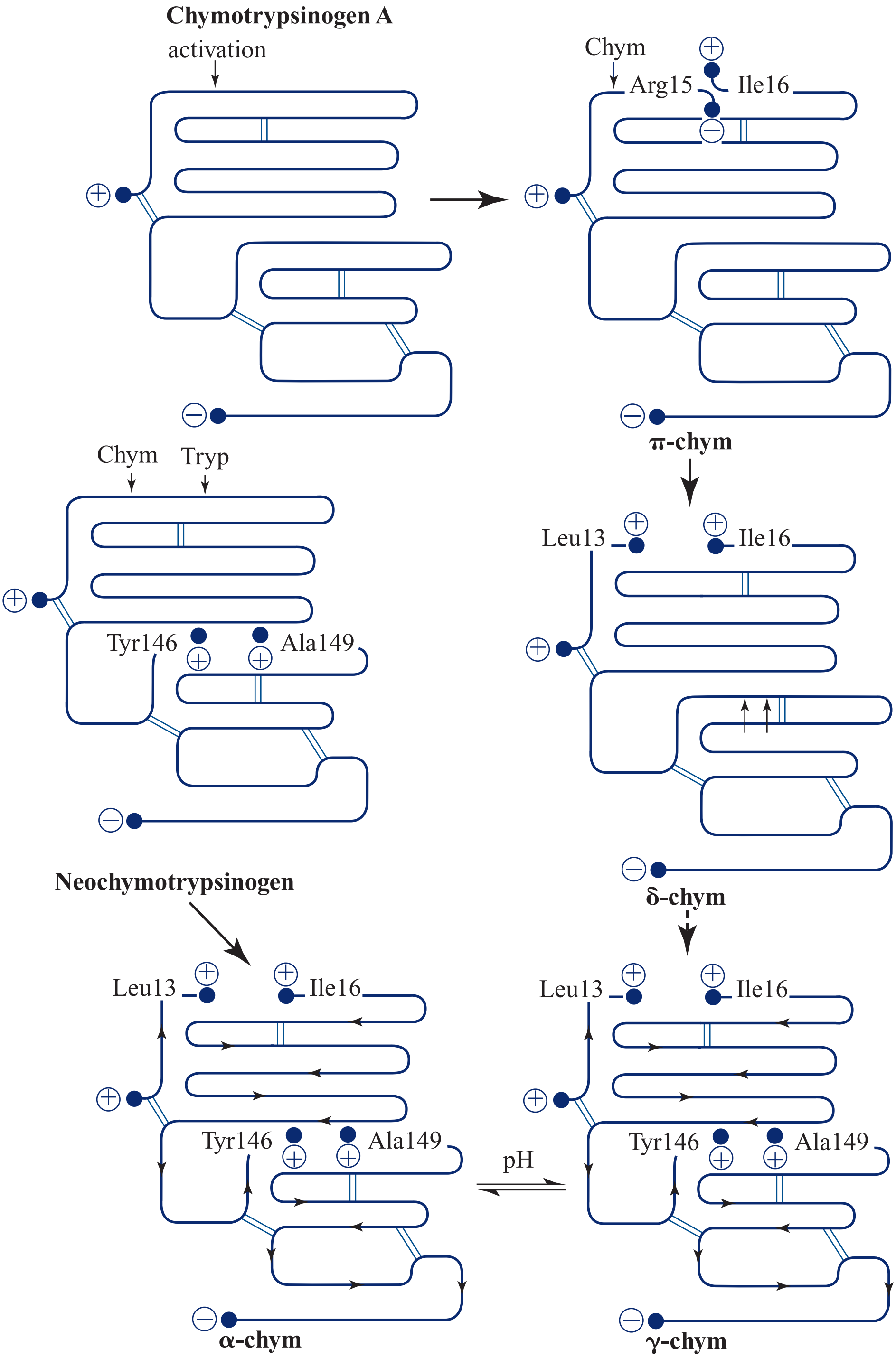


Fig. 12.2 Activation of chymotrypsinogen

◤ Table 12.1 indicates the values of the kinetic parameters obtained for the hydrolysis of diverse substrates by trypsin, chymotrypsin and their respective zymogens. The best substrate of chymotrypsinogen is tertiobutyl oxycarbonyl alanine p.nitrophenyl ester (Boc-Ala-ONP) and that of trypsinogen is benzyl oxycarbonyl hydroxyproline glycine p.nitrophenyl ester (Z-Gly-Hyp-Gly-ONP). Examination of the kinetic parameters obtained for diverse substrates clearly indicates that zymogens present a weak but significant catalytic activity. The comparison of values obtained for each en-zyme and the corresponding zymogen shows that the increase in the K_m value can reach two orders of magnitude whereas k_{cat} is decreased by a factor of 10 000 in the zymogen. On the basis of these results the authors attribute the weak activity of zymogens to a distortion in the primary substrate binding site. In addition, the same authors (LONSDALE-ECCLES et al., 1979) showed that strong ionic strength increases the activity of trypsinogen and chymotrypsinogen, indicating that in the absence of a salt bridge the functional structure of the protein can be stabilised by strong ionic strength. **These differences in efficiency between zymogen and enzyme were interpreted by GHÉLIS and YON (1979) on the basis of energy in terms of con-formational coupling between the two domains of the enzyme, the coupling being optimised upon salt bridge formation.**

Table 12.1 Hydrolysis of paranitrophenyl esters by zymogens and the corresponding enzymes (Reprinted with permission from *Biochemistry*, **17**, LONSDALE-ECCLES J.D. et al., 2805. © (1978) American Chemical Society)

	$k_{cat}(s^{-1})$	$K_m(M \infty 10^4)$	$k_{cat}/K_m(M^{-1} . s^{-1})$
Trypsinogen			
Boc-Gly-ONP	6.1×10^{-3}	26.4	2.3
Z-Ser-Gly-Gly-ONP	4.0×10^{-3}	6.3	6.4
Z-Gly-Hyp-Gly-ONP	7.6×10^{-3}	1.8	41.5
Trypsin			
Boc-Gly-ONP	0.22	5.2	0.43×10^{3}
Z-Ser-Gly-Gly-ONP	0.47	5.8	0.81×10^{3}
Z-Gly-Hyp-Gly-ONP	0.99	5.0	1.97×10^{3}
Chymotrypsinogen			
Boc-Gly-ONP	6.6×10^{-3}	6.3	10.5
Z-Ser-Gly-Gly-ONP	3.7×10^{-3}	7.1	5.2
Z-Gly-Hyp-Gly-ONP	9.3×10^{-3}	3.9	23.8
Boc-Ala-ONP	28.1×10^{-3}	4	72
Chymotrypsin			
Boc-Gly-ONP	0.52	0.22	26.6×10^{3}
Z-Ser-Gly-Gly-ONP	0.32	0.47	6.8×10^{3}
Z-Gly-Hyp-Gly-ONP	1.88	1.75	10.7×10^{3}
Boc-Ala-ONP	2.3	21	1.1×10^{3}

◢

The crystallographic data confirmed the existence of a salt bridge previously deduced by chemical, enzymatic and conformational studies (see Chap. 9). This led to the identification of Asp194 (Fig. 12.3). In spite of local rearrangements, the overall structure of chymotrypsinogen and the enzyme are very similar.

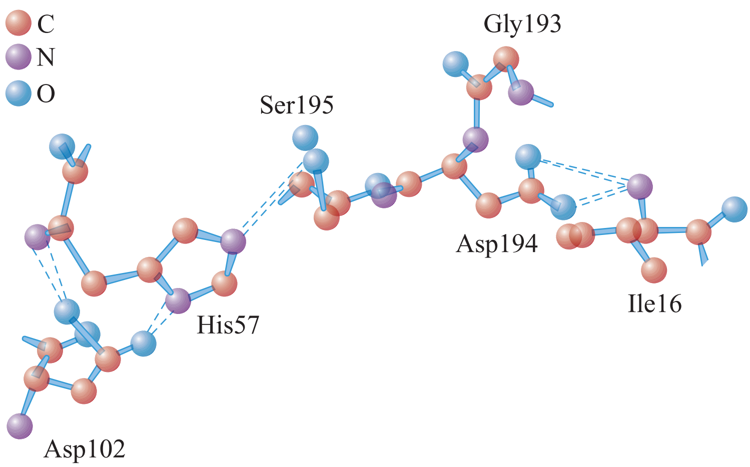


Fig. 12.3 The salt bridge in chymotrypsin according to the crystallographic data
(Reprinted from *The Enzymes*, 3rd ed., Vol. III, BLOW DM., The structure of chymotrypsin, 196. © (1971) Academic Press, with permission from Elsevier)

◤ Average displacements of α-carbon atoms from zymogen to enzyme are on the order of 1.8 Å. Rearrangements resulting from the activation of the zymogen are very localised. Some segments undergo displacements of a larger amplitude (Table 12.2). Segments I, III, and IV correspond to loops situated at the exterior of the molecule. The movement of segment II accompanies the formation of the salt bridge. Isoleucine 16 undergoes a displacement of 11.3 Å to be placed at the interior of the molecule by interacting with Asp194; this movement takes place by a rotation of the main chain of around 180°. Asp194 is localised in the interior of the molecule in the enzyme like in the zymogen.

Table 12.2 Residues for which the position of Cα differs by more than 3.6 Å between the structures of chymotrypsinogen and α-chymotrypsin
[Reprinted from *The Enzymes*, 3rd ed., Vol. III, KRAUT J., Chymotrypsinogen: X-ray Structure, 165. © (1971a) Academic press, with permission from Elsevier]

Segment	*Residue*	*Displacement in Å*
I	Gln7	4.8
	Pro8	10.0
II	Ile16	11.3
	Val17	6.6
III	Thr37	4.0
	Gly38	6.6

Segment	*Residue*	*Displacement in Å*
IV	Asp72	5.6
	Gln73	9.6
	Gly74	9.1
	Ser75	6.2
	Ser76	10.1
	Ser77	5.6
V	Thr144	5.9
	Arg145	8.7
	Tyr146	4.6
	Ala149	4.7
	Asn150	6.7
	Thr151	4.7
	Pro152	4.6
VI	Met192	8.4
	Gly193	6.6

In segment V, the side chain of Arg145 is displaced by 9 Å and becomes completely accessible to the solvent. In segment VI, Met192 and Gly193 undergo an important displacement that, with Ile16, leads to the setting up of a binding pocket for specific substrates. Met192, deeply buried in the zymogen, comes to the surface during activation permitting access of the substrate to the active centre. These crystallographic data are in agreement with results from chemical studies carried out by the group of NEURATH which showed that, in the enzyme, Met192 is very reactive towards iodoacetate whereas it practically does not react in the zymogen. The residue Gly193 is placed in proximity to His40 and, by the oxygen of the carbonyl and its peptide bond, establishes a hydrogen bond with $N^{\varepsilon 2}$ of His40 that, in the zymogen, forms a hydrogen bond with the oxygen of Asp194. The conformational changes implicating Ile16, Met192, Gly193 and Asp194 induce the formation of the specific substrate binding site during the activation of the zymogen. ◢

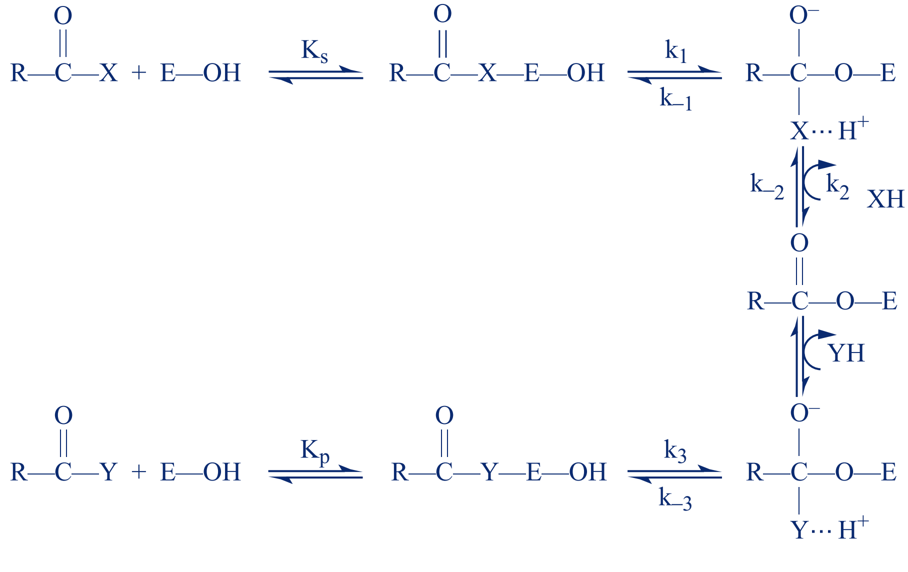
It is remarkable that the three catalytic groups Ser195, His57 and Asp102, thought to be linked by hydrogen bonds are found in exactly the same spatial arrangement in chymotrypsin and chymotrypsinogen. The catalytic triad is equally conserved in the same position upon trypsinogen activation into trypsin as was shown by FELHAMMER et al. (1975). This activation is provoked, in this case, by weak differences in the flexibility of regions well localised into the molecule. These regions become more rigid in the active enzyme. Trypsinogen adopts the conformation of trypsin in the presence of the Ile-Val dipeptide provided that it is covalently bound to the p.guanido benzoate group.

Numerous chemical labelling experiments associated with kinetic studies had shown the catalytic role of Ser195 and His57 before crystallographic studies were carried out (see Chap. 9). The structural analysis confirmed the existence of these two groups localised into the cleft and interacting by hydrogen bond, forming the active centre near the surface.

◤ The existence of a hydrogen bond between the $N^{\varepsilon 2}$ of His57 and the O^{γ} of Ser195 is based on the interatomic distance between these groups which is of 2.8 to 3 Å; it is 3 Å in trypsin (BODE et al., 1983) and 2.9 Å in elastase for which the binding of an inhibitor induces the alignment of serine and histidine at this distance. Moreover, crystallographic studies revealed that Asp102 is bound by hydrogen bond to a nitrogen proton of His57, which had driven crystallographers to propose the existence of a charge relay system between Ser195-His57-Asp102, and to propose a catalytic mechanism which was invalidated by more recent experiments; the role of Asp102 is discussed in Sect. 12.1.1.6. The intervention of this group in catalysis had not been revealed by chemical analysis. ◢

12.1.1.3. THE REACTION PATHWAY

Reactions catalysed by serine proteases proceed by a sequence of reaction steps common to enzymes of the chymotrypsin family and those of the subtilisine family that possess the same catalytic groups as mentioned above. Most experimental data were obtained on chymotrypsin. The reaction sequence, of which the kinetic evidence was presented in Chap. 5, implicates the formation of the non-covalent MICHAELIS complex and the formation of an unstable tetrahedral intermediate in which a covalent bond is established between the O^{γ} atom of the reactive serine and the substrate, the proton of serine being partially transferred to the leaving group. The tetrahedral intermediate gives rise to the acyl-enzyme and releases the leaving group after the breaking of the C—X bond (see scheme below). Deacylation is the inverse process in which the leaving group is replaced by water or by another nucleophilic compound on which the acyl transfer takes place. The complete reaction scheme is represented below:



R-CO is either a polypeptide, an amino acid or an acyl group, XH is a peptide, an amine or an alcohol; YH can be water or any other acceptor.

As was shown in Part II (Chap. 5), the existence of the acyl-enzyme was proved experimentally from kinetic studies in different conditions. The tetrahedral intermediate is too unstable to be revealed in this manner. Its existence was deduced from indirect arguments, either by analysis of substitution effects on the acylation rate during hydrolysis of synthetic substrates, or by the study of isotope effects. A more direct argument was provided by spectrophotometry data during the hydrolysis of chromophore substrates like N-furyl acryloyl-L-tryptophan amide and the corresponding methyl ester; an additional intermediate situated between the MICHAELIS complex and the acyl-enzyme was observed. FINK and MEEHAN (1979) detected by spectrophotometry the presence of a tetrahedral intermediate upon hydrolysis by elastase of specific di- and tripeptides in carrying out kinetic studies at low temperature (–39°C) as a function of pH. It was assumed that this tetrahedral intermediate is very similar to the transition state.

The crystallographic data revealed the geometry of most of these intermediates and the interactions that stabilise them and contributed to elucidating the catalytic mechanisms. However, the structural data does not permit unambiguously the establishment of a mechanism; it is important to provide complementary data by means of all methods in enzymology. Among serine proteases, chymotrypsin was the focus of the most complete studies; it is therefore this enzyme which is used as an example to present the mechanism of action of these proteases.

12.1.1.4. THE BINDING SITE OF SUBSTRATES AND THE MICHAELIS COMPLEX

Physiological substrates of proteases being polypeptides, they can establish numerous interactions with their specific enzyme. In order to determine their importance, it is useful to divide these types of substrates as a function of the amino acid residues they contain by the notations P_1, P_2, P_3 upstream of the peptide bond to break, that is in the acyl group region, and P'_1, P'_2, P'_3 downstream, that is in the region of the leaving group. The corresponding sub-sites of the enzyme are designated by S_1, S_2, S_3 and S'_1, S'_2, S'_3 (Fig. 12.4 below), the sub-site S_1 defining the protease specificity. Different crystallographic data were obtained in the presence of substrate analogs or specific inhibitors, in particular the complex formed by chymotrypsin and N-formyl-L-tryptophan and the complex of γ-chymotrypsin with polypeptides possessing at the C-terminal end an L-phenylalanine chloromethyl ketone. These studies showed that the binding site of the substrate acyl part is constituted by a region of the polypeptide chain in a β structure containing the amino acids Ser214-Trp215-Gly216. This region is represented in Fig. 12.5 below which corresponds to the complex of chymotrypsin with N-formyl-L-tryptophan. The amido NH group of formyl tryptophan points towards the carbonyl of Ser214. The carboxyl group has a carbon atom in contact with the O^{γ} of Ser195 and a C—O bond practically parallel to the hydrogen bond between the O^{γ} of Ser195 and the $N^{\varepsilon 2}$ of His57.

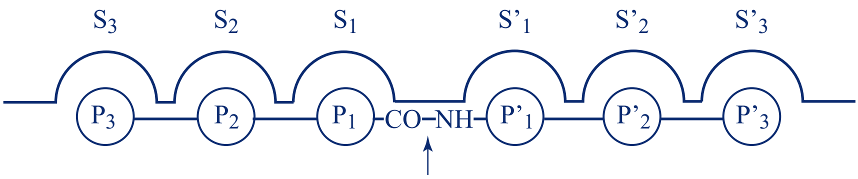


Fig. 12.4 Binding sub-sites of protease substrates

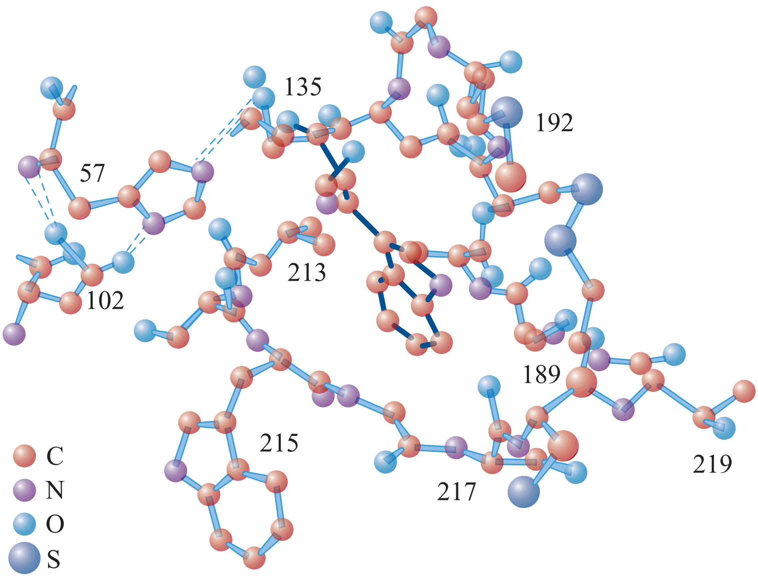


Fig. 12.5 Binding site of N-formyl tryptophan at the active centre of chymotrypsin

(Reprinted from *The Enzymes*, 3rd ed., Vol. III, BLOW DM., The structure of chymotrypsin, 199. © (1971) Academic Press, with permission from Elsevier)

◤ If one considers the different sub-sites, three hydrogen bonds can be formed between a substrate and the enzyme in the region of the acyl group, one with the carbonyl oxygen of S_1 and the amido nitrogen of P_1, another between the amido nitrogen of the peptide backbone of S_3 and the carbonyl oxygen of P_3 and a third between the carbonyl oxygen of S_3 and the amido nitrogen of P_3. Only this last bond can be formed with N-formyl tryptophan which only interacts with the S_1 site. ◢

The indolyl group of tryptophan is placed in the pocket of specificity, the planes of neighboring peptide bonds being practically parallel to the plane of the indol ring. This cavity comprises Ser189, Val213, Ser217; Met192 constitutes the hydrophobic and flexible opening of the cavity. The site of specificity of chymotrypsin accommodates aromatic residues whereas trypsin accepts substrates that possess an L-arginine or an L-lysine. In the crevasse of specificity of this last enzyme, Ser189 is replaced by Arg189 which contracts an electrostatic interaction with the positive charge of the side chain of substrates. The structure of elastase and of its complex with formyl alanine shows that the two glycines which are at the entrance of the substrate binding pocket in chymotrypsin are replaced by Val216 and Thr226; these residues hinder the entry of encumbering side chains, limiting substrate binding to the dimensions of the Ala residue (SHOTTON & WATSON, 1970).

The formation of the MICHAELIS complex is not accompanied by any significant conformational change. No displacement of the peptide backbone greater than 0.2 Å, which represents the limit of detection of the method, was observed.

The geometry of enzyme interactions with the leaving group is less well established. This results from the fact that, on one hand this interaction is much weaker, and on the other hand it is not easy to have inhibitors that bind strongly and in a productive manner on this part of the active centre. Kinetic studies of nucleophilic competition showed however that chymotrypsin possesses an S'_1 site capable of binding a P'_1 residue provided that the latter does not have a free carboxyl group. In trypsin, analogous studies using aliphatic alcohols of different chain sizes as nucleophile reagents revealed the hydrophobic character of the binding site of the leaving group (SEYDOUX et al., 1969). The hydrophobic character of the sub-site S'_1 was confirmed by crystallography studies of complexes of trypsin with two of its inhibitors, bovine pancreatic inhibitor (HUBER et al., 1974) and soybean inhibitor (BLOW et al., 1974). In elastase the sub-sites S'_1 and S'_2 are also hydrophobic (ATLAS, 1975).

12.1.1.5. THE TETRAHEDRAL COMPLEX AND THE BINDING SITE OF THE OXONIUM ION

The most important particularity of enzymatic catalysis resides in the complementarity of the enzyme for the transition state of the substrate (see Chap. 11). Analogs of the transition state of protease substrates that form tetrahedral compounds must therefore possess a high affinity for the enzyme. The tetrahedral addition compound yields an oxyanion formed by the carbonyl oxygen of the substrate. On the basis of crystallographic data obtained on an acyl-enzyme, indole acryloyl chymotrypsin, HENDERSON (1970) suggested that the oxyanion could contract a pair of hydrogen bonds, one with the NH of the main chain of Ser195, the other with that of Gly193; these hydrogen bonds could be stronger than those which would be formed by the carbonyl oxygen of the acyl-enzyme.

In complexes formed by trypsin with its protein inhibitors, the carbonyl group of the inhibitor is tetrahedral, its distance to the O^{γ} of the reactive serine being too long to establish a VAN DER WAALS interaction. In addition, this carbon has the same geometry in the complex that the inhibitor forms with the modified trypsin in which the serine has been transformed by chemical modification into dihydroalanine. Magnetic resonance studies of the proton indicate that the extra proton would be bound to His57 in the complex formed by pig β-trypsin and bovine pancreatic inhibitor.

The crystal structure of two aromatic derivatives of boronic acid bound to subtilisine was resolved by MATTHEWS et al. (1975). These authors concluded that these systems represent good models for determining the geometry of the transition state. An analogous study was carried out by TULINSKY and BLÉVINS (1987) on complexes formed by chymotrypsin and phenyl ethane boronic acid. This compound is associated to chymotrypsin with a very good affinity and forms a stable tetrahedral complex with the O^{γ} of Ser195 which occupies one of the tetrahedron positions. The phenyl ethyl boronate molecule occupies a part of the specificity site. The formation of the tetrahedral complex brings about a rotation of –14° in the angle χ_1 of the reactive serine.

12.1.1.6. THE CATALYTIC TRIAD AND THE MECHANISM OF ACYLATION

The catalytic triad is common to all serine proteases which are defined by the presence of a single reactive serine forming an acyl-enzyme with the substrates. Crystallographic data indicated the invariance in the respective position of Asp102, His57 and Ser195 in proteases of the chymotrypsin family. Although the global structure differs noticeably, the same geometry of the catalytic triad is found in subtilisines with the groups Asp32, His64 and Ser221. The distance between these groups is such that they can be connected by hydrogen bonds, which led crystallographers to propose a catalytic mechanism (Fig. 12.6).

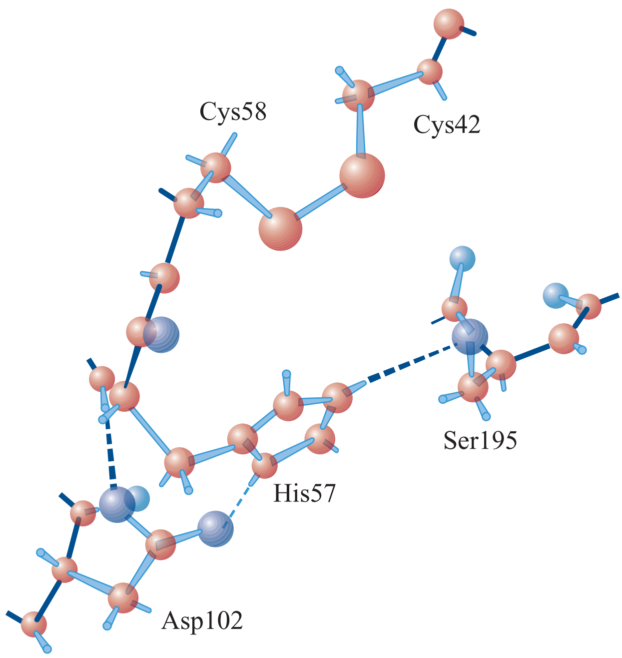


Fig. 12.6 Residue positions in the catalytic triad

(Reprinted from *The Enzymes*, 3rd ed., Vol. III, BLOW DM., The structure of chymotrypsin, 202. © (1971) Academic Press, with permission from Elsevier)

According to their hypothesis, the polarisation of the system would be carried out *via* a ***charge relay system*** in which the negative charge of the carboxylate would drive the increase in the nucleophilic ability of the serine and consequently would make it very reactive. At low pH where the protein crystallises, the system would remain inactive, but at pH 8 which is the pH of optimal activity of the enzyme, the charge relay would become functional. **However, the hypothesis of the charge relay system proposed by crystallographers was the subject of numerous controversies. Nuclear magnetic resonance studies suggested indeed that the catalytic base is histidine 57 and not aspartate 102, which was ultimately proved.**

One of the limitations of structural studies by X-ray diffraction resides in the fact that this method does not permit localising hydrogen atoms. In contrast, neutron diffraction constitutes an appropriate method to directly determine their exact position. An analysis by neutron diffraction at a resolution of 2.2 Å was carried out in 1981 by KOSSIAKOFF and SPENCER on trypsin covalently bound to an analog of the transition state, the monoisopropyl phosphoryl group. By this method, the authors determined the protonation states of Asp102 and His57. They concluded that the catalytic base is His57 and not Asp102; the serine proton is not transferred to Asp102 as was suggested by the hypothesis of charge relay system. Asp102 would intervene in the transition intermediate by modifying the dielectric environment of His57 of which the pK is usually 7, allowing attainment of a value higher than 9.5; the histidine would remain protonated in the transition state (IT) forming with Asp102 an ionic interaction (Asp^- His^+ IT^-). Figure 12.7 represents the mechanism proposed for the deacylation process which is similar to acylation, the water molecule replacing the catalytic serine. This result is in agreement with the theoretical studies of UMEYAMA et al. (1981) who, using quantum mechanics methods, attributed to Asp102 a purely electrostatic role which has the effect of lowering the energy barrier of proton transfer from serine to histidine, but in which the aspartate does not accept the histidine proton.

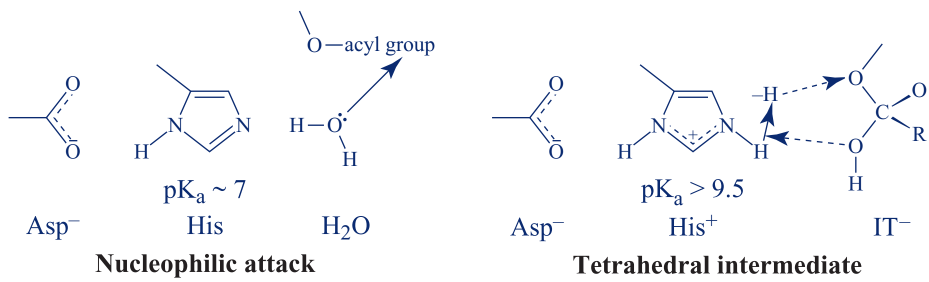


***Fig. 12.7** Mechanism of action of serine proteases in the deacylation step*

This clearly shows that crystallography, if it is necessary, is insufficient to permit by itself the establishment of a catalytic mechanism.

12.1.1.7. THE ACYL-ENZYME AND THE DEACYLATION STEP

Among the first crystallographic studies, several were carried out on proteases bound covalently and irreversibly by serine 195 to inhibitors in order to avoid autolysis. The structures of tosyl α-chymotrypsin, tosyl γ-chymotrypsin, tosyl elastase, and phenyl methane sulfonyl subtilisine BPN' were resolved. The structure of a more specific acyl-enzyme, indole acryloyl α-chymotrypsin was determined by HENDERSON (1970).

This acyl-enzyme had been previously characterised by kinetic and spectrophotometric studies by the group of BERNHARD et al. (1970). It has a lifespan 10 000 times longer than that of acyl-enzymes of specific substrates and is deacylated more slowly.

◤ Crystallographic studies clearly show that the indolyl part is placed in the site of hydrophobic specificity at a depth of 0.5 to 1 Å, as is the aromatic part of N-formyl tryptophan and of N-formyl-phenylalanine. The carbonyl part of the indole acryloyl group occupies a position close to that of the sulfonyl group in tosyl chymotrypsin. The O^{γ} of serine 195 moves by 2.5 Å upon acylation, which can correspond to a rotation of 120° around the Cα—Cβ bond. A movement of the imidazole of His57 of 0.3 Å towards the solvent region is observed. A water molecule is found localised in a position such that it can form a hydrogen bond with the imidazole $N^{\varepsilon 2}$ of His57 and the carbonyl oxygen of the acyl-enzyme (Fig. 12.8). This water molecule plays the role of the nucleophilic group in the process of deacylation; Met192 is displaced by 1 Å and the Cys191-Cys220 disulphide bridge by 0.35 Å. These movements are produced during acylation but do not bring about important displacements of the peptide backbone. Thus chymotrypsin as well as the ensemble of serine proteases from this family, can be considered as a relatively rigid enzyme, the most important movements accompanying the processes of acylation and deacylation.

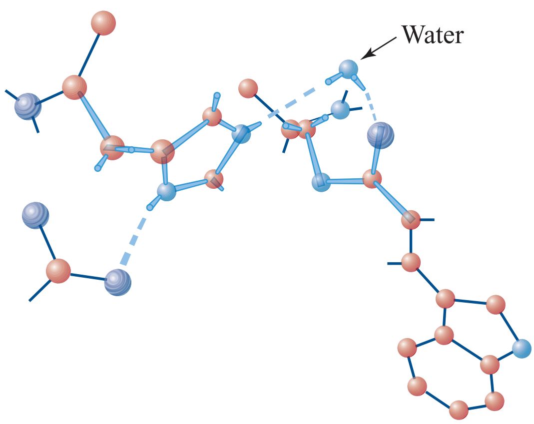


Fig. 12.8 Position of the water molecule with regard to the ester bond of serine 195 in indolacryloyl chymotrypsin ◢

On the basis of structural data obtained for indolacryloyl chymotrypsin, HENDERSON proposed a reactive acyl-enzyme model (Fig. 12.9 opposite). In this model, the N-H amide group very likely forms a hydrogen bond with Ser214. The orientation of the carbonyl oxygen is different from that observed in indolacryloyl chymotrypsin; it is correctly oriented with regard to the water molecule. Thus, deacylation can proceed rapidly by the extraction of a water proton in a general base catalysis, the formation of a tetrahedral intermediate and the appearance of the product.

A problem often discussed is that of the structural identity of a protein in a crystal and in solution. It is indeed very important to resolve this to know if one can transpose directly the conclusions from crystallographic studies to catalytic events that are produced in solution. BERNHARD and ROSSI succeeded in directly measuring in the crystal the rate of deacylation of indolacryloyl chymotrypsin. They found it equal to that determined in solution in the same conditions of pH and temperature.

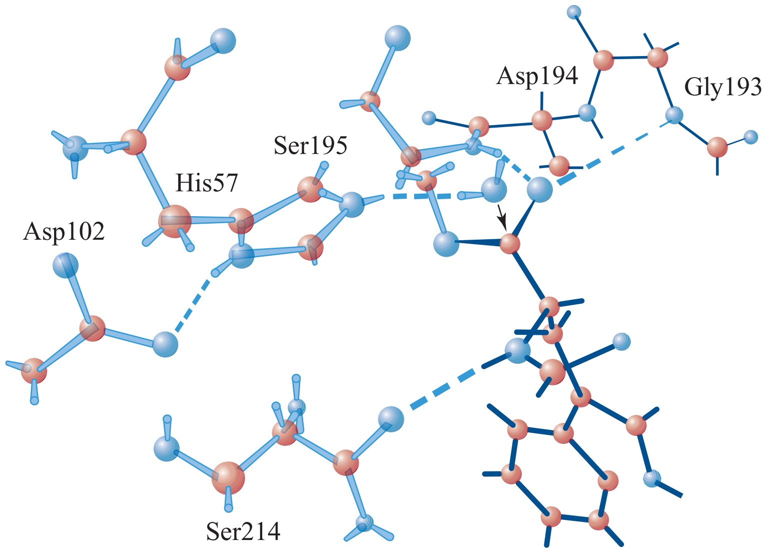


Fig. 12.9 Hypothetical model of the position of atoms in a reacting acyl-enzyme
(Reprinted from *J. Mol. Biol.*, **54**, HENDERSON R., Structure of crystalline α-chymotrypsin: IV. The structure of indoleacryloyl-α-chymotrypsin and its relevance to the hydrolytic mechanism of the enzyme, 351. © (1970) with permission from Elsevier)

The ensemble of structural data from the crystal and the functional and structural studies in solution, completed by theoretical approaches, has therefore allowed to specify the chemical mechanisms implicated in catalysis by serine proteases.

12.1.2. THIOL PROTEASES

Thiol proteases constitute a large family of enzymes which includes papain that was identified first, actinidine, chymopapain, ficine, bromelain, streptococcal protease, clostrinopeptidase B and cathepsins. These enzymes exist in different animal and plant species and in microorganisms. The best known are of plant origin; papain has been particularly well studied. Cathepsins are present in animal tissues. Table 12.3 below indicates the known thiol proteases and their origin.

Thiol proteases are generally monomeric proteins of which the molecular weight is situated between 25 000 and 28 000 with some exceptions. Some are glycoproteins like bromelaine, chymopapain and most cathepsins. All thiol proteases of plant origin are endopeptidases as well as most cathepsins with the exception of cathepsin B that is primarily a carboxypeptidase.

The biological role of most thiol proteases is not known. Enzymes of plant origin have a very broad specificity. It has been suggested that their role is to protect the fruit against attack from insects or mold. In mammals, cathepsins account for a large part of the proteolytic activity of lysosomes. Some even have specific known functions, for example cathepsin N which has a collagenolytic function and cathepsin P which converts proinsulin into insulin.

Table 12.3 Thiol proteases of different origins

(From The thiol proteases: structure and mechanisms in *Biological macromolecules and assemblies*, **3**, BACKER E.N. & DRENTH J., ed. by F.A. JURNAK & A.M. MCPHERSON, 313-368. © (1984 John Wiley and Sons). This material is reproduced with permission of John Wiley & Sons, Inc.)

Enzyme	*Origin*	*Molecular mass*
Plant enzymes		
Papain	*Carica papaya*	23 350
Chymopapain*	*Carica papaya*	35 000
Peptidase A	*Carica papaya*	26-28 000
Actinidine	*Actinidia chinensis*	24 000
Ficine	*Ficus*	23 500
Bromelain (stem)*	*Ananas comosus*	24-26 000
Bromelain (fruit)*	*Ananas comosus*	33 500
Asclepain A	*Asclepias syriaca*	31 000
Asclepain B	*Asclepias syriaca*	23 000
Calotropine DI	*Calotropis gigantea*	21 000
Calotropine FI*	*Calotropis gigantea*	23 400
Mammalian enzymes		
Cathepsin B*	Rat liver	23-26 000
	Human liver	24-27 500
Cathepsin B_2	Rabbit lung	52 000
Cathepsin H	Rat liver	28 000
	Human liver	28 000
Cathepsin L	Rat liver	23-24 000
	Bovine spleen	23-25 000
Cathepsin I	Rabbit lung	26 000
Cathepsin N	Bovine spleen	18-20 000
	Human placenta	34 600
Cathepsin T	Rat liver	33-35 000
Cathepsin P	Rat islets	31 500
Calpain	Rat liver	90 000
Bacterial enzymes		
Streptococcal proteinase	*Streptococci hemolyticus*	32 000
Clostripain	*Clostridium histolyticum*	55 000

* *glycoproteins*

12.1.2.1. STRUCTURAL ASPECTS

Among all the thiol proteases, papain is the best known from the structural and functional points of view. Its complete sequence was determined in 1978 by CARNE and

MOORE, and its three-dimensional structure is known and has been refined to 1.65 Å resolution (KAMPHUIS et al., 1985). It is constituted of a single polypeptide chain of 212 amino acids folded into two structural domains. The N-terminal domain is essentially made of antiparallel β segments. The C-terminal domain is composed of α helices: a large helix A (residues 24–42) which extends at the interface between the two domains and two shorter helices, helix B (50–57) and helix C (67–78) (Fig. 12.10). The ensemble of the structure is stabilised by three disulphide bridges, two in the N-terminal part and the third in the C-terminal part. The interface between the two domains presents a predominantly polar character comprising four charged residues (Lys17, Glu35, Glu50 and Lys174), and encloses a network of 9 water molecules. Numerous hydrogen bonds are established between the two domains. As in serine proteases, the active centre is localised between the two domains with one of the two catalytic residues, Cys25, in the N-terminal domain, and the other, His159, in the C-terminal domain.

Fig. 12.10 Schematic representation of the three-dimensional structure of papain (PDB: 6PDA)
Catalytic residues Cys25 on the main helix and His159 are represented in red

The structure of other thiol proteases has been determined. Actinidine possesses a sequence and a three-dimensional structure very similar to that of papain, with a small number of insertions and deletions (Fig. 12.11 below). Of the 212 amino acids in papain, 98 are identical in actinidine. The differences are unequally distributed. The sequences forming the helices A and B of the first domain and the β structures of the second domain present 69% homology; 81% of the residues situated at the interface between the two domains are conserved. In contrast, the residues present in the hydrophobic core of the protein present only 41% identity. In the median part of the chain made up of residues 82 to 116, one notes only 11% identity; the helix C presents 17% homology. Despite some differences between the sequences of papain and actinidine, these proteins possess a very high similarity in their three-dimensional structure.

The complete sequences of two mammalian proteases, cathepsins H and B, are known; the other thiol protease sequences have been partially determined. Cathepsin H presents a very high degree of homology with papain and actinidine. The residues situated around the catalytic groups are so well conserved in different thiol proteases that whether they come from plants or mammals, they probably share a common origin (Table 12.4). The largest difference is observed in the staphylococcal enzyme which does not possess disulphide bridges. Either this enzyme has considerably diverged, or its structure results from a convergent evolution selecting the essential catalytic groups. Thus, there is a great similarity in the three-dimensional structure of known thiol proteases, the most significant differences being localised in the regions that comprise insertions or deletions. Figure 12.11 below indicates the approximate positions of deletions and insertions in some of these enzymes.

Table 12.4 Amino acids around catalytic residues of thiol proteases

	Sequence around the reactive cysteine																					
			20					**25**					30					35				
Papain	N	Q	G	S	C	G	S	C	W	A	F	S	A	V	V	T	I	E	G	I	I	K
Actinidine	S	Q	G	E	C	G	G	**C**	W	A	F	S	A	I	A	T	V	E	G	I	N	K
Bromelain (stem)	N	Q	N	P	C	G	A	**C**	W	A	F	G	A	I	A	T	V	E	S	V	A	S
Bromelain (fruit)	N	Q	N	P	C	G	A	**C**														
Ficine	Q	Q	G	Q	C	G	S	**C**	W													
Chymopapain B	R	V	P	A	S	G	E	**C**	Y													
Chymopapain	N	Q	G	S	C	G	S	**C**	W	A	F	S	T	I	A	T	V	E	G	I	N	K
Cathepsin H	N	Q	G	A	C	G	S	**C**	W	T	F	S	T	T	G	A	L	E	S	A	V	A
Cathepsin L	Y	Q	G	A	C	G	S	**C**	W	A	F	S	A	V	V	L	A	Q				
Cathepsin B	D	Q	G	S	C	G	S	**C**	W	A	F	G	A	V	E	A	M	S	D	R	I	C
Streptococcal proteinase	G	Q	A	A	T	G	H	**C**	V	A	T	A	T	A	Q	I	M	K	Y	H	N	Y
	Sequence around the reactive histidine																					
	155				**159**					165												
Papain	N	K	V	D	**H**	A	V	A	A	V	G	Y	G									
Actinidine	T	A	V	D	**H**	A	I	V	I	V	G	Y	G									
Bromelain (stem)	D	K	L	N	**H**	A	V	T	A	I	G	Y	N									
Ficine	T	S	L	D	**H**	A	V	A	L													
Cathepsin H	D	K	V	N	**H**	A	V	L	A	V	G	Y	G									
Cathepsin B	V	M	G	G	**H**	A	I	R	L	G	W	G										
Streptococcal proteinase	K	V	G	G	**H**	A	F	V	I	D	D	G	A									

12.1.2.2. ACTIVATION OF THIOL PROTEASES

The mechanism of activation of thiol proteases differs from that of serine proteases for which exists a zymogen that must undergo a proteolytic cleavage. Thiol proteases require the presence of an activator to expose the reactive sulfhydril group and permit the expression of the activity. They can be activated by cyanide or reducing agents like cysteine and reduced glutathione as well as 2,3-dimercaptopropanol. Papain is very sensitive to oxidation. The thiol group of the active centre easily participates in exchange reactions with disulphide bridges.

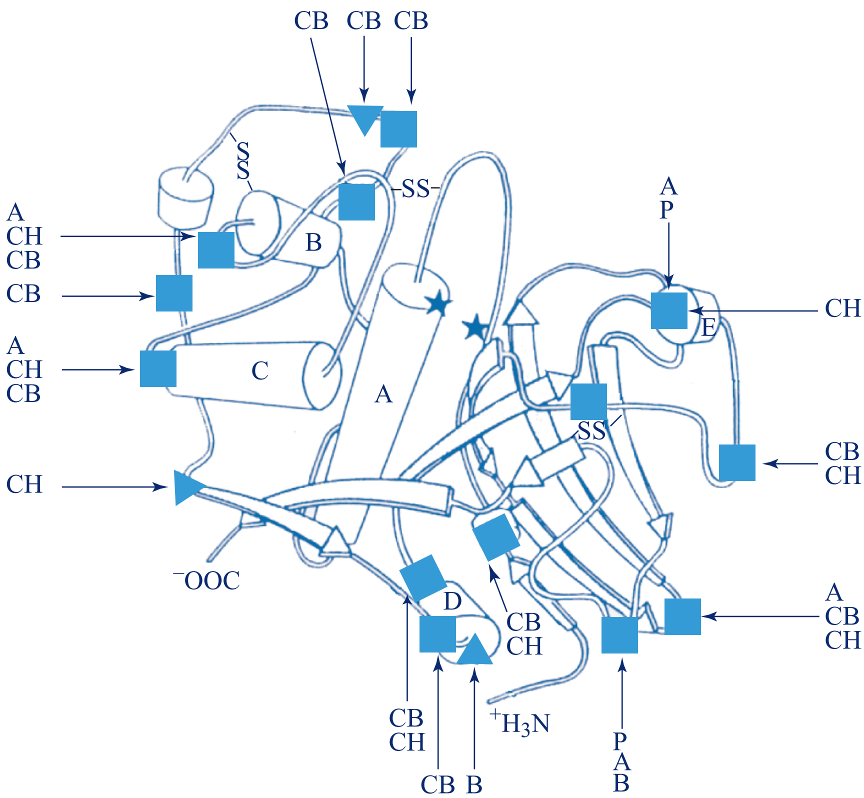


Fig. 12.11 Approximate positions of insertions and deletions in different thiol proteases: papain (P), actinidine (A), bromelain (B), cathepsin H (CH) and cathepsin B (CB), deduced from crystallography studies for papain and actinidine, and from sequence alignments for cathepsins H and B and bromelain

(From The thiol proteases: structure and mechanisms in *Biological macromolecules and assemblies*, **3**, BACKER E.N. & DRENTH J., ed. by F.A. JURNAK & A.M. MCPHERSON, 313-368. © (1984 John Wiley and Sons). This material is reproduced with permission of John Wiley & Sons, Inc.)

12.1.2.3. THE ACTIVE CENTRE

BERGER and SCHECHTER (1970) had shown that papain contains seven binding subsites of the substrate; four of them S_1, S_2, S_3 and S_4 are situated in the acyl part of the substrate scissile peptide bond, the three S'_1, S'_2 and S'_3 in the amino part of this. The specificity of papain and enzymes of this family is relatively broad. It requires the presence of a hydrophobic residue, preferentially phenylalanine in the sub-site S_2. Crystallographic studies carried out on complexes of papain with a series of inhibitors, chloromethyl ketones of di-, tri- and tetrapeptides permitted the study of the interactions, and to visualise the association of natural substrates to the binding site as shown in Fig. 12.12 below; hydrogen bonds between the substrate and the enzyme are represented. These studies, as well as NMR experiments, showed that the conformation of the substrate is not distorted in the complex.

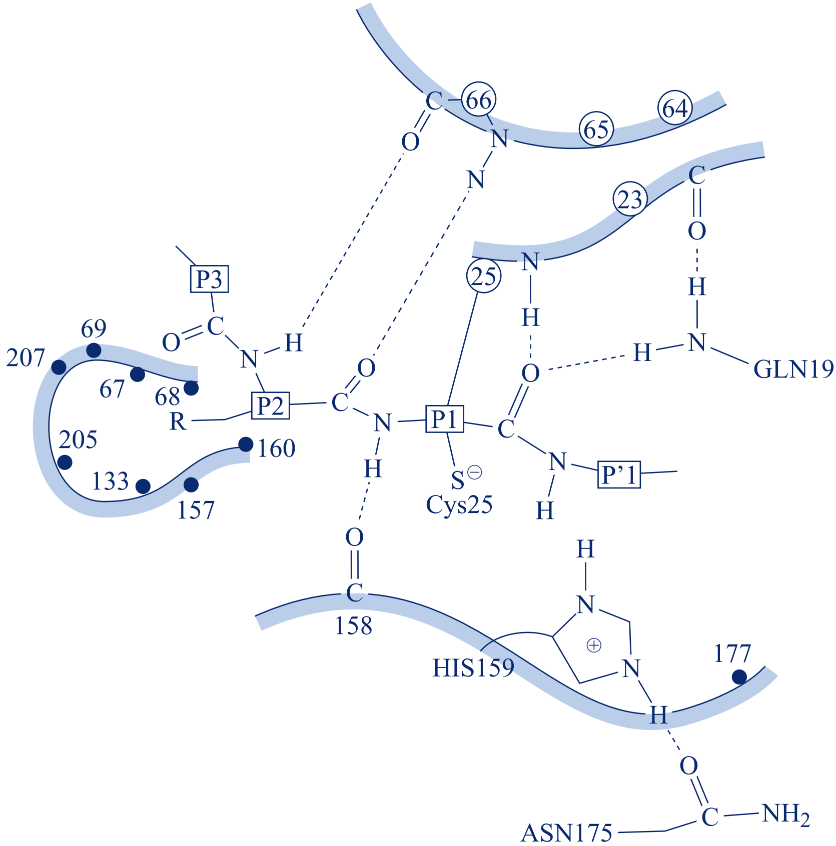


Fig. 12.12 Enzyme-substrate interactions in papain

(From The thiol proteases: structure and mechanisms in *Biological macromolecules and assemblies*, **3**, BACKER E.N. & DRENTH J., ed. by F.A. JURNAK & A.M. MCPHERSON, 313-368. © (1984 John Wiley and Sons). This material is reproduced with permission of John Wiley & Sons, Inc.)

Catalytic groups Cys25 and His159 which constitute the catalytic pair are found in a favorable position relative to the peptide bond to break. In addition, Asn175 interacts by hydrogen bond with the catalytic histidine. One finds again a catalytic triad similar to that of serine proteases (Fig. 12.13). The hydrogen bond His…Asn and the C^{β}—C^{γ} bond of His159 are approximately collinear. The electron density map revealed differences in the orientation of the cycle of the imidazole of His159 in the non-active oxidised papain in which a cysteinyl group is bound to the S^{γ} of Cys25. Following the respective orientation of His159 and Asn175, the rotation of histidine around the C^{β}—C^{γ} bond can be carried out without breaking of the hydrogen bond His…Asn.

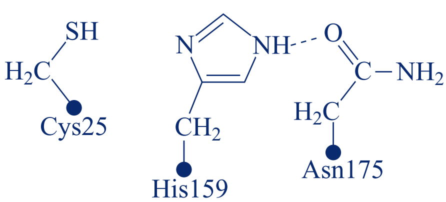


Fig. 12.13 The catalytic triad in papain
(from BACKER & DRENTH, 1984)

The amplitude of internal movements of the protein, evaluated from B factors by crystallographers, indicates that the active centre has a relatively rigid structure. In particular, the movements of the polypeptide chain carrying Cys25 and His159 are of a weak amplitude. They occur in regular structures; Cys25 is on the helix A and His159 on a β segment. In contrast, the corresponding side chains present a great mobility. The residues present at the active centre, principally Trp177 and residues 157–159, 22–24 and 64–67 deviate upon binding of an inhibitor and very likely that of a substrate, enlarging the binding site. No relative movement of domains was revealed; the active centre seems relatively rigid but can enlarge during substrate binding.

12.1.2.4. CATALYTIC MECHANISM

The catalytic mechanism implicates a nucleophilic attack by cysteine of the active centre to form a tetrahedral addition compound in which the oxyanion would be stabilised by a charge of the enzyme. The two groups NH_2 of Gln19 and NH of Cys25 would participate at the site of the oxyanion, interacting by hydrogen bond with oxygen (see Fig. 12.13). The breaking of the bond would occur by a proton transfer from a donor group to the nitrogen of the peptide bond, which would release the leaving group and would form the covalent intermediate, water intervening as a nucleophile in the final step of the reaction. The thiol group under the SH form only possesses a very weak nucleophilic power; in contrast, the S^- form is an excellent nucleophile. Numerous experimental arguments obtained by spectroscopy and NMR indicate that the active form of the enzyme corresponds to the existence of an ion pair between S^- and the protonated form of the imidazole of histidine:

$$S^- \cdots {}^+HIm$$

which is predominant between pH 4 and 8.5. At lower pH the enzyme is inactive, the catalytic groups being under the form $SH \cdots {}^+HIm$; at pH higher than 8.5, the predominant form $S^- \cdots Im$ is inactive. The existence of significant electrostatic fields created by the dipole of the N-terminal helix would have the effect of stabilising the ion pair forming the catalytic site. The role of Asn175 would consist therefore of properly orienting the imidazole nucleus of histidine.

12.1.3. ACID PROTEASES OR ASPARTYL PROTEASES

Acid proteases or aspartyl proteases constitute another class of proteases that includes gastric proteases like pepsin, gastricin, chymosin or rennet. Pepsin from pig has been studied for a long time; as early as 1930, NORTHROP succeeded with the crystallisation. It is secreted as a zymogen made of 370 amino acids. Other enzymes of this family like rennin intervene in some processes of biological control. One also finds acid proteases in some molds, penicillinopepsin of *Penicillium janthinellum* of which the molecule is made of 330 amino acids, pepsin of *Rhizopus chinensis* and of *Endothia parasitica*. No zymogen precursor was detected for these enzymes.

There are also aspartyl-proteases in retroviruses, in particular in the virus HIV 1 responsible for human immunodeficiency (AIDS). This protease plays an important role in the maturation of the virus; by limited proteolysis it produces proteins implicated in the replication of the virus.

Aspartyl-proteases catalyse the hydrolysis of peptide bonds; they are active between pH 1 and pH 7. Gastric enzymes are implicated in the degradation of alimentary proteins. Among the others, rennin has a very narrow specificity in the hydrolysis of the N-terminal peptide of angiotensinogen; this enzyme cleaves only the Leu10-Leu11 bond giving rise to angiotensin which is involved in the regulation of blood pressure.

Diverse approaches showed that the catalytic groups of these enzymes are two aspartic acids; however the catalytic mechanism was the subject of controversies, and different models have been proposed. They will be discussed in the light of structural data.

12.1.3.1. ACTIVATION OF ZYMOGENS

The activation of pepsinogen in pepsin occurs in acidic conditions (pH 3). In a first step, there is the breaking of the Leu16-Ile17 peptide bond which brings about pseudo-pepsin. The activation continues by breaking of the Leu44-Ile45 bond giving rise to an additional peptide which produces active pepsin. The protonation of the zymogen that precedes the proteolytic cleavage at pH 3 is accompanied by a conformational change. At low pH, the activation is essentially intramolecular. For strong zymogen concentrations and at pH 4, the process becomes autocatalytic; the activated molecules hydrolysing the zymogen molecules, the reaction is accelerated. Pepsinogen possesses 11 lysine residues that, with the exception of one (Lys364), are eliminated during activation. Pepsin once formed is susceptible to autolysis in acidic solution.

The group of SIELECKI (1991) solved the structure of pig pepsinogen at 1.8 Å. The results confirm conclusions from studies carried out in solution showing an important conformational change upon zymogen conversion into the active enzyme. The 44 N-terminal amino acids of the pro-segment form a long β strand followed by two approximately orthogonal helices. There are strong electrostatic and hydrophobic interactions between the pro-segment and the active centre of the enzyme. In particular, the interaction between Lys36 of the pro-segment and Asp215 of the enzyme, the hydrogen bonds between Tyr37 of the pro-segment and Asp215 of the enzyme and between Tyr9 and Asp32 of the enzyme block the access of the substrates to the active site in the zymogen. Upon activation Ile1 undergoes an ample movement of 44 Å and these interactions are broken in opening the entry of the active site.

12.1.3.2. STRUCTURAL ASPECTS

The first sequence determined entirely was that of pig pepsin; it comprises 327 amino acids and three disulphide bridges. Among mold acid proteases, only the primary structure of penicillinopepsin was entirely characterised.

The three-dimensional structures of several aspartyl proteases have been solved, including those of penicillopepsin to 1.8 Å, pig pepsin to 2.3 Å, then to 1.8 Å; those of rhizopuspepsin and endothiapepsin were obtained at medium resolution. The structure of aspartyl protease of HIV 1 was determined by the group of NAVIA (NAVIA et al., 1989). It possesses most characteristics of other aspartyl proteases with which it presents great structural analogies, the only difference being that it is a homodimer.

All acid proteases belong to the class of β proteins; besides some very short helical segments, they are essentially made of β segments (Fig. 12.14). They are folded into two structural domains; in pig pepsin, the N-terminal domain consists of 176 amino acids and the C-terminal domain 147. There is an approximate 2-fold symmetry axis between these domains; each of them is itself constituted of portions of similar structure connected also by an approximate 2-fold symmetry axis. This organisation suggested that evolution of acid proteases would result in the product of a quadruple gene formed from four copies of an ancestral unit of around 45 residues organised in a β sheet made of 6 antiparallel β segments. Diverse mutations would have generated the loops that are present today in these proteases. The sheet formed from 6 antiparallel β segments is a structure common to all acid proteases. The interdomain 2-fold symmetry axis is situated between the two central sheets as shown in Fig. 12.15 below that specifies the localisation of hydrogen bonds.

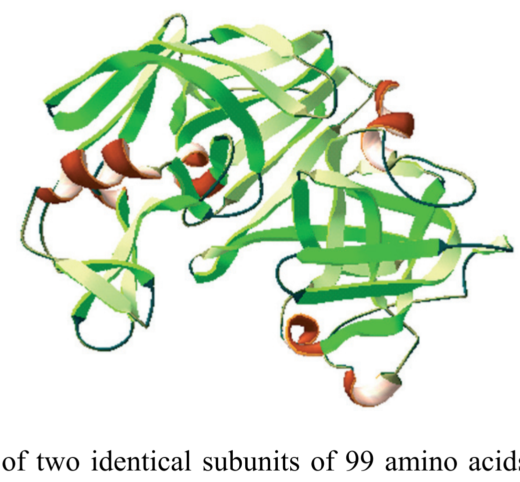


Fig. 12.14 Ribbon diagram of the structure of penicillopepsin (PDB: 3APP)

HIV 1 protease is a dimer consisting of two identical subunits of 99 amino acids which present between them a 2-fold symmetry axis. Each protomer contains the Asp, Thr, Gly triad present in all aspartyl proteases, with a 2-fold symmetry axis. Each protomer corresponds to a domain of other aspartyl proteases.

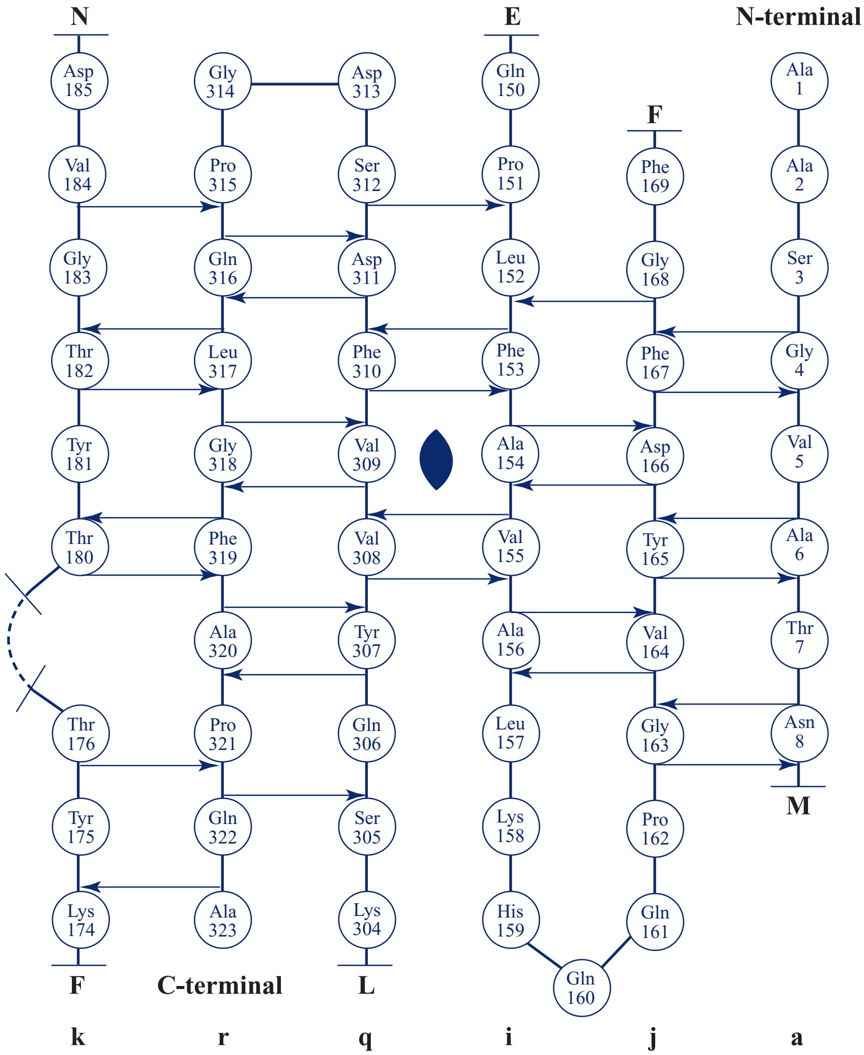


Fig. 12.15 β sheets of acid proteases with their symmetry axis

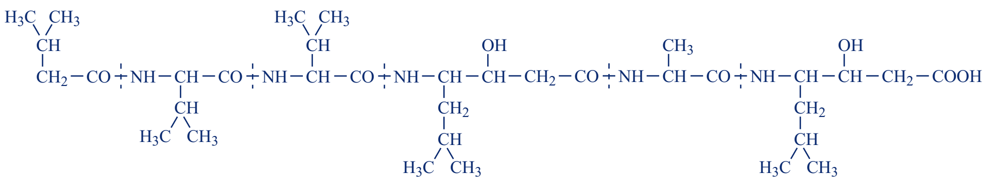
The position of hydrogen bonds is indicated by arrows

(From Aspartic proteinases and their catalytic pathway in *Biological macromolecules and assemblies*, **3**, JAMES M.N.G. & SIELECKI A.R., ed. by F.A. JURNAK & A.M. MCPHERSON, 431. © (1987 John Wiley and Sons). This material is reproduced with permission of John Wiley & Sons, Inc.)

12.1.3.3. ENZYME-SUBSTRATE ASSOCIATION

Pepsin hydrolyses peptide bonds with a specificity which implicates hydrophobic amino acids in position P_1 and P'_1. The analysis of catalytic constants and principally of the ratio k_{cat}/K_m for a series of synthetic polypeptide substrates suggests that the binding site of the substrate is stretched and can accommodate up to seven amino acids. The lengthening of the polypeptide chain, either near the N-terminal or the C-terminal extremity, increases the value of k_{cat} without modifying the value of K_m (FRUTON, 1970). These results were confirmed with other aspartyl proteases. From the data as a whole, it has been suggested that a conformational change of the enzyme associated with constraints in the substrate could bring an important contribution to the catalytic efficiency.

Crystallographic data were obtained on complexes formed by acid proteases and inhibitors, such as pepstatin and its analogs. Pepstatin is a powerful natural inhibitor of aspartyl proteases. Its composition is the following: isovaleryl(Iva)-Val-Val-Sta-Ala-StaOH, Sta being the statin residue (acid 4S, 3S-4-amido-3-hydroxyl-6-methyl heptanoic acid). Figure 12.16 gives the formula of this compound. The affinity of this inhibitor for aspartyl proteases is very high ($K_d = 4.6 \times 10^{-11}$ for pig pepsin and 1.5×10^{-10} for penicillinopepsin). This suggested that the central statin residue is an analog of the tetrahedral transition state of true substrates. Diverse analogs such as Iva-Val-Val-StaOEt as well as pentapeptide AcPro-Ala-Pro-Ala-PheOH were used for the crystallographic study of the complexes.



Iva Val Val Sta Ala Sta

Fig. 12.16 The pepstatin molecule

The structural data clearly indicate that the inhibitor approximately adopts a β type conformation, the side chains alternating left and right of the main chain. Thus, the binding site of the enzyme can accommodate a polypeptide of 7 to 8 amino acids. The association of pepstatin and its analogs to penicillinopepsin is accompanied by a large conformational change of the loop formed by residues Ile73-Ser80 (*flap*). No displacement was observed in this region during binding of the pentapeptide. This conformational change permits the establishment of hydrogen bonds between the enzyme and the inhibitor, and it drives a proper orientation of catalytic residues Asp33 and Asp213 towards the peptide bond to be broken.

These studies also allowed the localisation of the substrate binding sub-sites, in particular S_1 and S'_1 as well as S_4, S_3 and S_2 and to provide by molecular modelling, a representation of the association of a polypeptide substrate to the enzyme (Fig. 12.17 below). In the complex, the carbonyl oxygen of the peptide bond to be cleaved points towards Asp33 and Asp213 displacing a solvent molecule. In addition, the oxygen of a water molecule (O284) is found in an ideal position for a nucleophilic attack of the carbon atom of carbonyl.

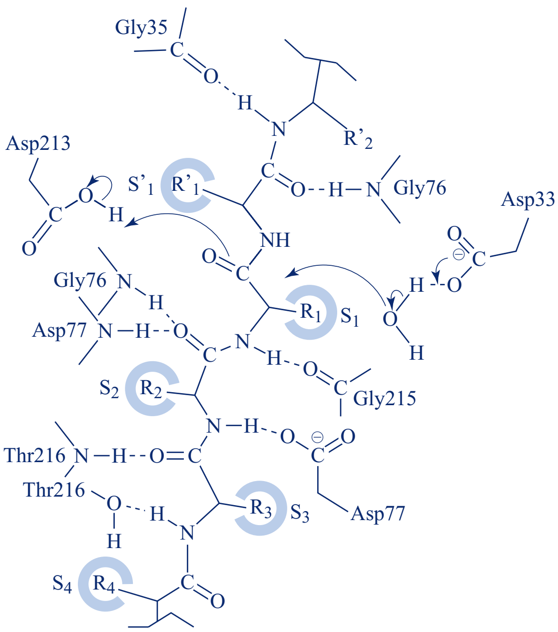


Fig. 12.17 Association of a peptide substrate to penicillopepsin
(From Aspartic proteinases and their catalytic pathway in *Biological macromolecules and assemblies*, **3**, JAMES M.N.G. & SIELECKI A.R., ed. by F.A. JURNAK & A.M. MCPHERSON, 469. © (1987 John Wiley and Sons). This material is reproduced with permission of John Wiley & Sons, Inc.)

12.1.3.4. THE CATALYTIC SITE

The two catalytic groups Asp33 and Asp213 in penicillopepsin, Asp32 and Asp215 in pig pepsin, are present in all aspartyl proteases. Several experimental results had permitted the identification of the catalytic role of two carboxyls before it was found by crystallographic studies. Kinetic studies as a function of pH indicated that the reaction rate depends on the ionisation of two groups of pK 1.2 and 4.7, respectively. Chemical modifications by affinity labelling with diazo and epoxyde compounds had allowed the identification of aspartates. The structural data confirmed these conclusions and showed that Asp33(32) and Asp213(215) are situated at the centre of the substrate binding site; the interdomain 2-fold symmetry axis goes between these two residues which are in a hydrophobic environment. Several solvent molecules were identified at the active site; some of them are displaced by specific inhibitor association with the enzyme.

The active site of the HIV 1 aspartyl protease is made up of Asp25 of one protomer and Asp25' of the other; it possesses an organisation very similar to the active site of pepsin. This active site is created by the symmetrical structure of the assembly of protomers; this represents "an elegant method for creating an active enzyme while encoding a minimal amount of genetic information" (NAVIA et al., 1989). The resolution of the structure of this aspartyl protease marked an important step in the development of therapies against AIDS, permitting the conception and synthesis of inhibitors directed against the active site.

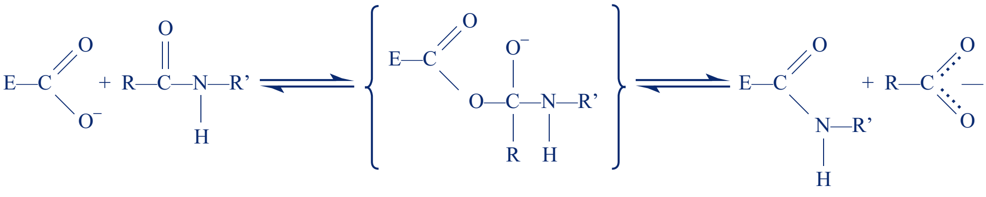
12.1.3.5. FORMATION OF THE TETRAHEDRAL INTERMEDIATE

The catalytic mechanism of acid proteases was the subject of controversies and has not yet been rigorously established. All mechanisms proposed implicate in a first step the formation of a tetrahedral addition intermediate. An electrophilic group polarises the carbonyl, to facilitate the nucleophilic attack of the carbon of the peptide bond and to stabilise the negative charge of the oxyanion. However, it is not yet established if the proton is yielded directly by one of the two groups Asp213 or Asp33 following an electrophilic mechanism or *via* a water molecule (O39) following a mechanism of general acid catalysis.

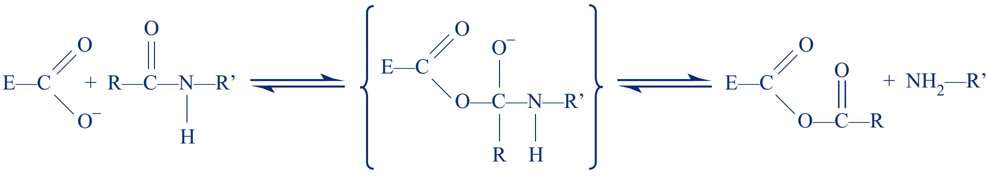
In the absence of proof, the first hypothesis illustrated in Fig. 12.17 is generally accepted (JAMES & SIELECKI, 1987). The catalysis would start by the protonation of the carbonyl oxygen by the residue Asp213. The electronic rearrangements resulting in this protonation would permit the nucleophilic attack of the carbon by the oxygen of a water molecule (O284) assisted by the COO^- group of Asp33 which would play the role of the base catalyst.

12.1.3.6. BREAKING OF THE TETRAHEDRAL INTERMEDIATE

The first hypotheses proposed the formation of a covalent intermediate between one of the two catalytic carboxyls and a part of the substrate, but the nature of this intermediate, amino-enzyme or acid anhydride, was the subject of discussions. According to the first hypothesis, an amino-enzyme is formed with the release of the acid part of the substrate according to the reaction:



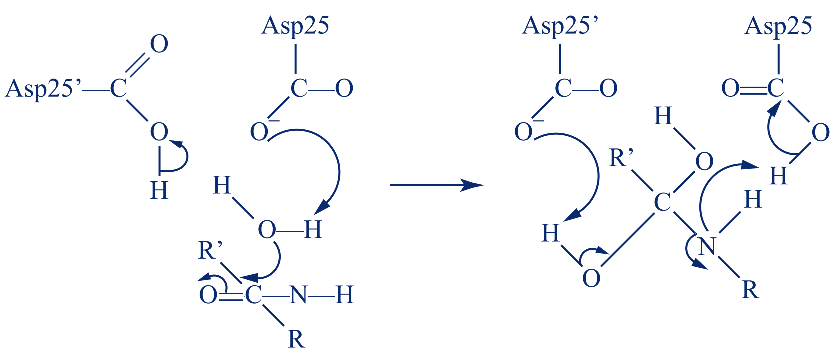
Objections based on arguments of steric order led to the rejection of this mechanism. According to another proposal, the tetrahedral intermediate would give rise to an acyl-enzyme anhydride whereas the amine part would be released according to the scheme:



The acid anhydride would be then decomposed *via* a water molecule. No attempt to detect the covalent intermediate succeeded, not even cryoenzymology experiments performed at –60°C (DUNN & FINK, 1984).

FRUTON in 1976 suggested for the first time that reactions catalysed by aspartyl proteases do not involve a covalent intermediate. Experiments of isotope exchange with ^{18}O brought arguments in favor of a mechanism implicating a direct breaking of the tetrahedral intermediate (ANTONOV et al., 1978). The hypotheses relative to the existence of a covalent intermediate had been formulated before the knowledge of the structure of the enzyme-substrate complex. Although the mechanism is not yet definitely established, the structural data permits to exclude the formation of a covalent intermediate. The breaking of the transition intermediate to give the products implicates first the protonation of the nitrogen of the NH_2 leaving group. This protonation could take place either from the solvent or by a transfer *via* Asp213; no data permits to distinguish between these two mechanisms.

Many theoretical approaches tried to elucidate the reaction mechanism of aspartyl proteases. One of them combining quantum mechanics and classical molecular dynamics was applied by the group of VAN GUNSTEREN (1996) to study the catalytic mechanism of the HIV aspartyl protease. The authors start from the principle that only one of the two Asp is protonated, which is in agreement with the difference in pK between these two groups, although it is not possible to distinguish them. The results of the simulations are in agreement with a mechanism of general acid-base catalysis in which Asp25' is protonated, according to the following scheme:



$\longrightarrow$ Asp25'COOH + Asp25COO$^-$ + RNH_2 + R'COOH

The proton transfers are predominant in all the proposed mechanisms. However, the X-ray data do not permit the localisation of the protons; it is not possible to define the limiting step of the reaction. The chemical experiments of the solvent isotope effect are difficult to interpret The method of neutron diffraction at high resolution should permit this determination.

12.1.4. *METALLOPROTEASES: CARBOXYPEPTIDASE A*

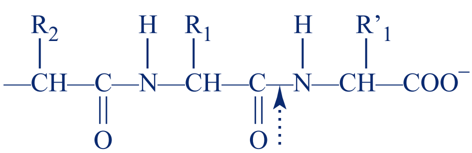
Zn^{++} metalloproteases form a very wide category of enzymes including pancreatic carboxypeptidases, the enzyme for the conversion of angiotensin very important in the regulation of blood pressure, and thermolysine that is a bacterial endopeptidase.

Pancreatic carboxypeptidases are exopeptidases which release the amino acid in the C-terminal position of a polypeptide chain. There are other carboxypeptidases from plants or microorganisms which are not metalloenzymes, in particular carboxypeptidase Y from yeast, dipeptidyl carboxypeptidase from *E. coli*, and D-D carboxypeptidases from *Streptomyces*.

Two carboxypeptidases, carboxypeptidases A and B, are secreted by the pancreas as zymogens or procarboxypeptidases. Carboxypeptidase A was the most studied and has been well-known for a long time; discovered in 1929 by WALDSCHMIDT-LEITZ, it was crystallised for the first time by ANSON in 1935. The presence of Zn^{++} in the molecule was evidenced in 1954; the enzyme as well as the zymogen contains one Zn^{++} ion per molecule. All the carboxypeptidase species resulting from the activation of the procarboxypeptidase contain Zn^{++}.

Carboxypeptidase A was the object of numerous studies, some carried out on the specificity of the enzyme, others on the role of the metal and on the nature of groups implicated in the chelation of this metal, others on groups of the active site, and on the catalytic mechanisms. The sum of works carried out on this enzyme was already impressive enough before the publication of results from crystallographic studies, which furnished solid bases for further studies.

The enzyme specificity was particularly well studied by the group of BERGMANN and FRUTON using synthetic substrates. Carboxypeptidase hydrolyses the last peptide bond at the C-terminal extremity of a polypeptide:



with the condition that the COO^- group is free and that the amino acid R'_1 is in the L configuration. Dipeptides having a free amino group are hydrolysed more slowly than those for which this group is blocked.

Although the specificity of pancreatic carboxypeptidases is relatively large, that of carboxypeptidase A implicates that the C-terminal amino acid is hydrophobic, with preference for an aromatic or branched aliphatic group, tryptophan, phenylalanine, tyrosine, isoleucine, alanine, and glycine; the presence of a proline diminishes or prevents hydrolysis. The residue R_1 can be a tryptophan, a glycine, an alanine, a methionine or a glutamate. Carboxypeptidase B is specific for charged residues (R'_1) lysine or arginine but does not recognise histidine. These enzymes are used to determine the N-terminal sequences of proteins as well as carboxypeptidase Y which releases most C-terminal amino acids including proline, but only slowly releases lysine, arginine and histidine.

12.1.4.1. ZYMOGEN ACTIVATION

According to the species, the precursor of the enzyme procarboxypeptidase A secreted by the pancreas appears either as a monomer or associated non-covalently to other zymogens. In mammals, procarboxypeptidase A forms a ternary complex with chymotrypsinogen C and proproteinase E (CHAPUS et al., 1987). The three-dimensional structure of this heterotrimer was resolved by the group of HUBER (GOMIS-RÜTH et al., 1997) to 2.35 Å by use of X-ray synchrotron. In the complex, procarboxypeptidase A occupies a central position; it is flanked by prochymotrypsinogen C and proproteinase E. Its site of activation is buried at the centre of the heterotrimer whereas the activation sites of chymotrypsinogen C and proproteinase E are localised at the surface and therefore are accessible. This organisation suggests a sequence of activation by trypsin in which chymotrypsinogen C and proproteinase E are activated first, the structure of the complex being very likely maintained and carboxypeptidase staying inactive. Next, the primary cleavage and then the secondary cleavage of procarboxypeptidase are produced bringing about the dissociation of the complex and the appearance of the activity of carboxypeptidase. It was suggested that this coordination in the appearance of proteolytic activities would have the effect of protecting carboxypeptidase from inactivation by the acidity of the gastric juice at the entrance of the duodenum. Procarboxypeptidase A is activated by trypsin cleavage at the N-terminal extremity of a peptide composed of 95 amino acids.

The three-dimensional structure of procarboxypeptidase A was determined by the group of HUBER (GUASH et al., 1992). It shows that the structure of the N-terminal peptide is made of three α helices and four antiparallel β segments forming a structural domain in the main part of the molecule (Fig. 12.18a opposite). This domain appears as an autonomous folding unit. It is suggested that this peptide is involved in the complexes that procarboxypeptidase A forms with chymotrypsinogen C or the proproteinase. In particular three aromatic residues exposed in the zymogen (Phe50, Phe58 and Phe94) would be implicated in this association. The activation of chymotrypsinogen is a slow process, the activation peptide staying associated to the enzyme until there occurs a new proteolytic cleavage at position 74. This peptide exerts an inhibitory effect on the enzyme. From structural considerations, the authors conclude that the segment connecting the activation domain to the enzyme contributes mainly to the inhibition of carboxypeptidase activity by positioning the activation domain on the substrate binding site. In fact, the activation gives rise to three carboxypeptidases of slightly different compositions, the trypsin proteolysis giving rise to α-carboxypeptidase which possesses 307 amino acids, the N-terminal extremity being an alanine. β-carboxypeptidase has 305 amino acids and a serine in the N-terminal position. γ-carboxypeptidase with 300 amino acids possesses an asparagine in the N-terminal position. The C-terminal residue is an asparagine for the three enzyme species.

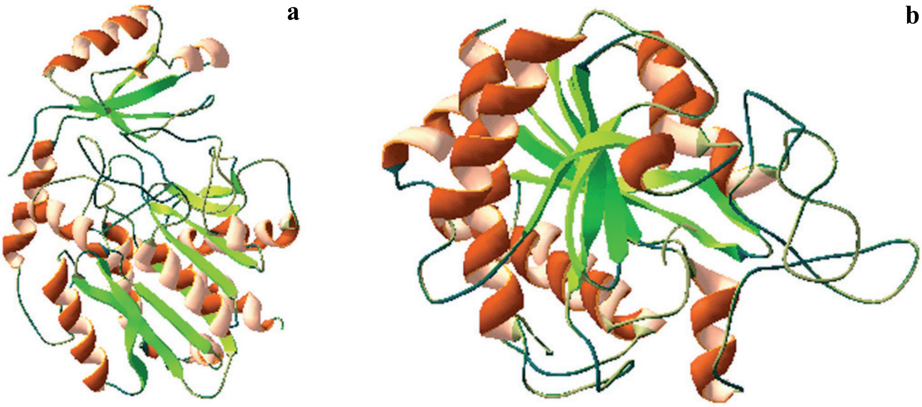


Fig. 12.18 (a) structure of procarboxypeptidase A (PDB: 1PCA)
(b) structure of carboxypeptidase A (PDB: 3CPA)

Carboxypeptidase A with molecular weight 34 000 consists of a single chain. The zymogen has a significant catalytic activity. It catalyses the hydrolysis of esters and also of peptides. It releases the C-terminal leucine of lysozyme with a rate only seven times lower than the enzyme. The binding site of substrates preexists in the zymogen. The difference in activity between the enzyme and the zymogen resides in part in the variations in pK of the carboxylate group of Glu270 which is two units higher in the enzyme (pK 7) than in the zymogen (pK 5). Such a difference suggested that the activation of the zymogen brings about a conformational change that modifies the hydrogen bonds contracted by the carboxylate of Glu270.

Carboxypeptidase B of molecular weight 34 600 is also synthesised as a zymogen which is activated by an enzymatic process. The sequences of the two carboxypeptidases A and B present 49% identity and their three-dimensional structures are very similar. The structure of procarboxypeptidase B was also resolved by the group of HUBER (COLL et al., 1991).

12.1.4.2. STRUCTURE OF CARBOXYPEPTIDASE A

The structure of carboxypeptidase A (form a) was resolved by the group of LIPSCOMB at Harvard at successive resolutions of 6, 2.8 and 1.5 Å (REES et al., 1983). This is the first metalloenzyme for which the three-dimensional structure was determined. Radiocrystallographic studies had been undertaken before the sequence was entirely determined by the group of NEURATH. The two structures were specified simultaneously by a confrontation of results obtained from each of the approaches. The free enzyme has a relatively open structure, the active centre forming a large crevasse on the surface of the molecule. Carboxypeptidase comprises large regions of β structures forming an internal sheet with parallel and antiparallel chains; 45 amino acids (about 15% of the molecule) are implicated in the β structures. The helical parts (25%) are located near the exterior of the molecule (see Fig. 12.18b). β turns of type I, II and III were characterised. Table 12.5 below indicates the parts of carboxypeptidase organised in regular structures. Water molecules interacting with the protein (195 water molecules per molecule of carboxypeptidase) were localised. Most are found at the exterior of the protein covering 62% of the surface and in tight contact (4 Å) with the atoms of the enzyme; 24 water molecules are buried in the interior, some of them associated by hydrogen bonds. The examination of temperature factors shows that the molecule is subjected to movements of a large amplitude, varying by 5 Å from the centre of the protein to 30 Å from the surface with an average of 8 Å. The localisation of Zn^{++} and its interactions with the side chains of carboxypeptidase were only known from crystallographic studies. It is the same for the amino acid residues which constitute the active centre of the enzyme (LIPSCOMB, 1980).

Table 12.5 Secondary structure of carboxypeptidase A

[Reprinted from *J. Mol. Biol.*, **168**, REES D.C. et al., Crystal structure of carboxypeptidase a at 1.54 Å resolution, 367. © (1983) with permission from Elsevier]

α *helices*	β *segments*	β *turns (type)*
14–28	32–39	3–6 (III)
73–90	47–53	29–32 (I)
93–101	60–66	41–44 (I)
112–122	104–108	69–72 (I)
173–187	191–196	72–75 (III)
215–231	200–204	89–92 (II)
254–261	239–241	99–102 (III)
285–306	267–271	123–126 (I)
		142–145 (II)
		150–153 (II)
		162–165 (I)
		169–172 (II)
		213–216 (I)
		243–246 (III)
		259–262 (I)
		277–280 (II)
		282–285 (I)

The structure of carboxypeptidase B was resolved a little later to 2.8 Å (SCHMIDT & HERRIOT, 1976). The two carboxypeptidases present great similarities; the only differences observed in the refolding of the main chain are situated at the level of external loops.

12.1.4.3. LOCALISATION AND ROLE OF Zn^{++}

Previous studies had shown that Zn^{++} is intimately associated to the activity of the enzyme. It can only be separated from the protein by powerful chelating reagents such as 1,10 phenanthroline, 8-hydroxyquinoleine and dithizone. Other compounds like cysteine, β-mercaptoethanol and thioglycollate also permit the elimination of Zn^{++}. The loss in activity is directly proportional to the removal of the metal. Apocarboxypeptidase, free of metal, is an inactive but stable protein. Its molecular characteristics are identical to those of the enzyme. The apoenzyme is still capable of forming a complex with a specific dipeptide. Zn^{++} therefore does not have a conformational role. It is possible to reconstitute the enzyme by addition of metal to the apoenzyme; the activity reappears proportionally to the concentration of the metal until one mole per mole of the enzyme is bound. The reconstituted enzyme can be distinguished from the native enzyme neither by its enzymatic properties, nor by its chemical properties, nor by its three-dimensional structure examined by X-rays.

◤ The activity of carboxypeptidase can be restored by other metals than Zn^{++}, such as Co^{++}, Ni^{++}, Mn^{++}, and Fe^{++}; all the metals of the first transition period restore both the peptidase and esterase activities of the enzyme. In the presence of Co^{++}, the activity is even greater than that of the enzyme with Zn^{++}. When Zn^{++} is replaced by a group II metal such as Cd^{++}, Hg^{++} or Pb^{++}, the metalloenzyme obtained still conserves an esterase activity but loses its peptidase activity. A model was proposed to account for differences between the enzyme with Zn^{++} and the enzymes in which Zn^{++} was replaced by Hg^{++} or Cd^{++}. Figure 12.19 shows the attack of the peptide bond by acid AH and base B: groups assisted by Zn^{++} in the peptidase activity of the enzyme. When this is replaced by Cd^{++} or Hg^{++}, the steric hindrance of the metal prevents the action of the basic group of the enzyme, and the amidase activity disappears. In contrast, the breaking of a more labile ester bond can use OH^- as a nucleophile group and therefore the reaction can proceed.

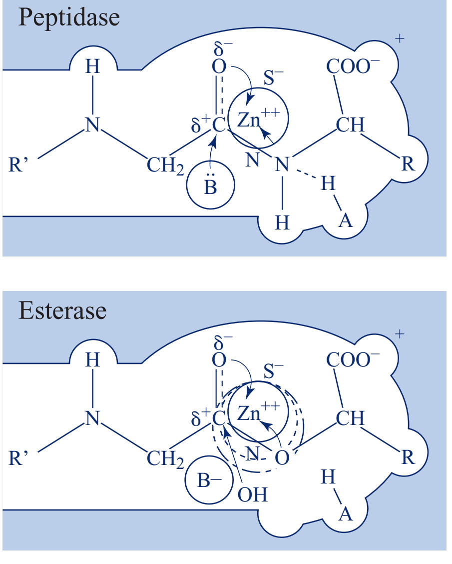


Fig. 12.19 Proposed model accounting for the esterase activity of carboxypeptidase when the Zn^{++} has been replaced by Hg^{++}

The constants of stability of carboxypeptidase complexes with diverse metals follow the following classification:

$$Hg^{++} > Cd^{++} > Zn^{++} > Cu^{++} > Ni^{++}\ Co^{++} > Mn^{++}$$

The dissociation constant has a value of 5×10^{-9} M for Zn^{++} and 1.5×10^{-6} M for Co^{++}. The analysis of stability constants as well as chemical labelling experiments had driven the group of VALLEE to suggest that the metal is chelated by a sulfur atom and a nitrogen atom. An SH group had been titrated by PCMB in the apoenzyme in the absence of metal but was masked in the presence of metal. The possible participation of imidazole had been turned down from experiments of photooxidation in the presence of methylene blue. Nitrogen had been attributed to the N-terminal α-amino group of the protein. ◢

Crystallographic studies showed that Zn^{++} is localised 25 Å from the N-terminal extremity of the enzyme. The groups that participate in chelation of the metal are the $N^{\varepsilon 1}$ nitrogens of residues His69 and His196 and the $O^{\delta 1}$ oxygen of residue Glu72, as well as a water molecule (W571). There are no cysteine residues in the molecule, but a disulphide bridge (Cys138-Cys161) far from the Zn^{++} region. It probably had been reduced in the chemical labelling experiments. His196 had first been identified with some ambiguity by crystallographic studies as a Gln or Lys residue; the authors had opted for lysine on the basis of experiments comparing the reaction of difluorobenzene with apoacetyl-carboxypeptidase and acetyl-carboxypeptidase. The resolution of the sequence established without ambiguity that residue 196 is a histidine (NEURATH et al., 1967). Furthermore, the sequence communicated by NEURATH in 1967 identified residue 69 as an isoleucine. In fact, the electron density map indicated that it was necessary to inverse the sequence 68–69 first identified as Ile-His, a result which was verified by a novel sequence study. The stability constant of the metal-protein complex obtained experimentally is perfectly compatible with the chelation by two nitrogens and an oxygen.

12.1.4.4. THE ACTIVE CENTRE

Before the crystallographic structure was established, a great number of studies using chemical labelling and enzymatic kinetics had been carried out with the goal of determining the catalytic groups of carboxypeptidase. The group of VALLEE in particular had deduced that one (or two) tyrosine residue(s) were implicated in the enzymatic activity. According to variations in the catalytic constant as a function of pH, these authors had suggested that a histidine was able to act as a general base catalyst. However, no chemical method had permitted to identify an essential histidine and it was difficult to choose which group, histidine or tyrosine, played the base catalyst role. Crystallographic studies were carried out by the group of LIPSCOMB not only with the free enzyme, but also with complexes of the carboxypeptidase with either a slightly reactive substrate or even inhibitors like p-phenyl propionate and L-lysyl-tyrosinamide. Data analysis drove to propose a catalytic role for residues Tyr248, Glu270 and Arg145. Nevertheless, later site-directed mutagenesis experiments producing the mutant Tyr248 $\longrightarrow$ Phe led to a revision of the role attributed to tyrosine (see Sect. 12.1.4.6).

12.1.4.5. ENZYME-SUBSTRATE ASSOCIATION

In the absence of an effector (substrate or analog), the active centre is presented as a deep pocket or by the metaphor of the mouth of a spider having long feelers. Upon substrate binding, one of these feelers moves. The conformational change consists of a coordinated movement of residues Arg145 and Tyr248 that draw near the substrate. Residue Tyr248 is displaced by 12 Å and is drawn to a distance of 2.7 Å from the nitrogen of the peptide bond to break and 3.5 Å from the amino group of lysyl-tyrosine. This movement implicates a rotation of about 120° around the C^{α}—C^{β} bond of this tyrosine. The residue Glu270 goes away from Zn^{++} about 2 Å by a rotation around the C^{α}—C^{β} and C^{β}—C^{γ} carbons. In the free enzyme, the residue Arg145 is interacting with carbonyl 155 of the peptide backbone and residue Tyr248 with residue Glu249. These interactions are suppressed in the complex and the catalytic residues interact with the substrate. On the other hand, the cavity filled with water in the free enzyme becomes a particularly hydrophobic region when the substrate is bound. The water molecule bound to Zn^{++} is chased by the substrate; the metal interacts with the carbonyl oxygen of the bond to break and polarises it. The movement of Tyr248 has the effect of closing the cavity. The ensemble of these movements, driving an adaptation of the enzyme to its substrate, represents an example particularly suggestive of induced fit. An interaction model of a tetrapeptide with carboxypeptidase was proposed by LIPSCOMB (Fig. 12.20).

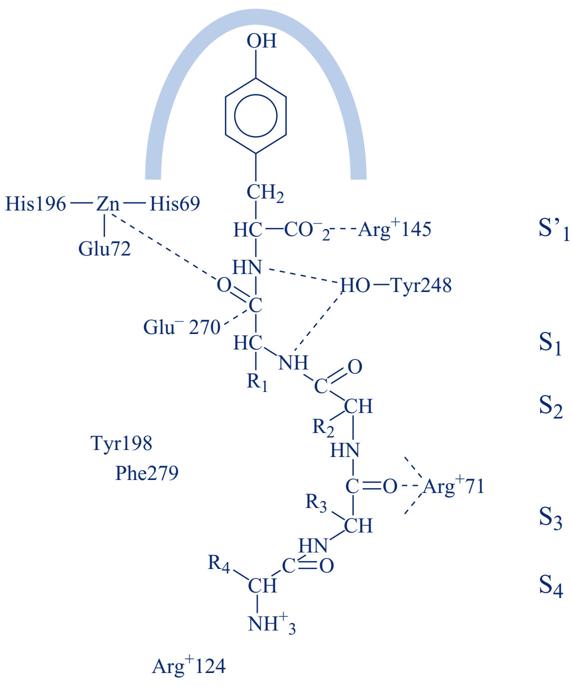


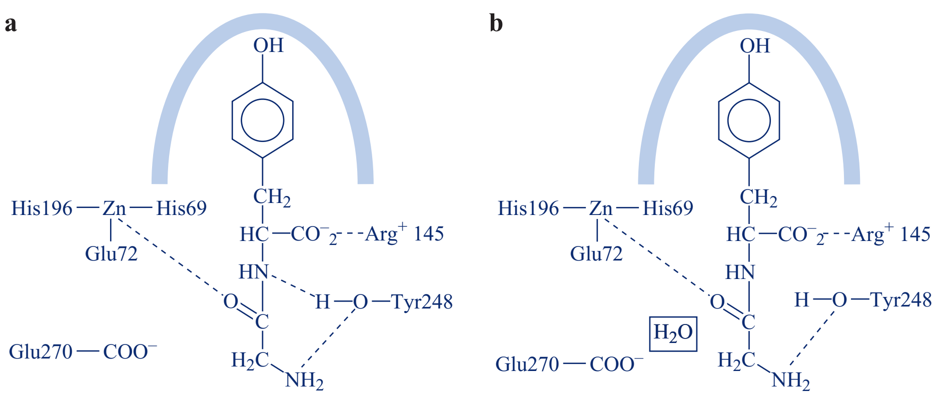
Fig. 12.20 Association of a peptide substrate to carboxypeptidase
(Reprinted from *The Enzymes*, 3rd ed., Vol. III, HARTSUCK J.A. & LIPSCOMB W.N., 1 Carboxypeptidase A, 45. © (1971) Academic press, with permission from Elsevier)

The mode of association of a ketonic substrate analog of ester, 2-benzyl-3-p-methoxypropionic acid, is similar to that of glycyl-tyrosine (LIPSCOMB, 1980). In the complex, the carboxy terminal of the substrate at site S'_1 is bound by two hydrogen bonds to residue Arg145. Residue Arg127 at site S_1 is 4 Å from residue Arg145. At site S_2 the residue Arg71 is found 4 Å from Phe279 and Tyr198 (S_2 to S_3).

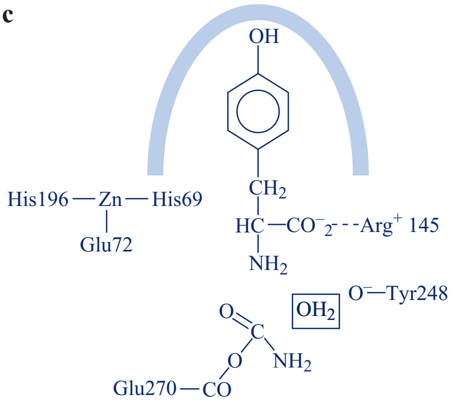
The active centre of carboxypeptidase B is very similar to that of carboxypeptidase A. The Zn^{++} ligands are identical (His69, His196 and Glu72). Residues Arg145 and Glu270 are present as well as residue Tyr248 of which the orientation around the C^{α}—C^{β} bond is undetermined. The presence of a residue Asp255 at the centre of the substrate binding pocket explains the difference in specificity between these two enzymes.

12.1.4.6. CATALYTIC MECHANISMS

The breaking of the peptide bond in the COOH terminal position of the substrate implicates a nucleophilic attack of the carbonyl and an electrophilic attack on the amide group of this bond. According to the position of the different atoms (Fig. 12.21a), two hypotheses had been formulated to explain the catalytic mechanism. According to one of them, residue Glu270 would function as a general base catalyst *via* a water molecule (Fig. 12.21b). According to the other hypothesis, a nucleophilic attack would be produced by the carboxylate of Glu270 to generate an anhydride intermediate. Without being possible to choose between these two hypotheses, some steric considerations, in particular the distances and orientations of the groups of reacting atoms, seemed rather in favor of the second. Tyr248 was assumed to act by yielding a proton to the amide group of the peptide bond. In this model, Arg145 would not have a catalytic role but would be indispensable to the stabilisation of the enzyme-substrate complex forming an electrostatic interaction with the carboxy terminal of the substrate.



c



d

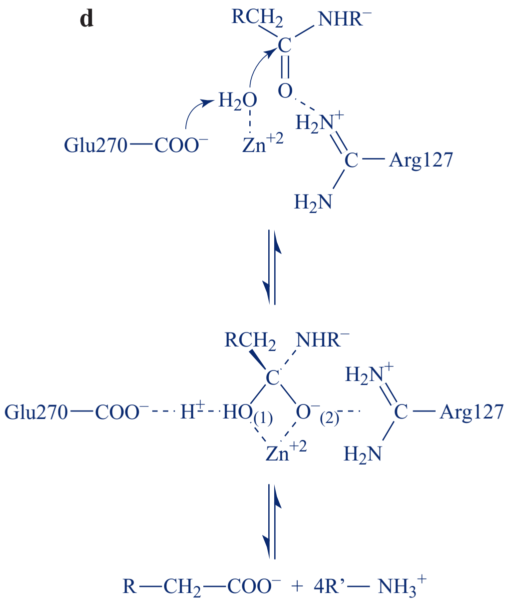


Fig. 12.21 Different catalytic mechanisms proposed for carboxypeptidase
*(**a**) position of enzyme groups with regard to the substrate – (**b**) proposed mechanism according to which Glu270 acts as a base catalyst – (**c**) hypothesis according to which Glu270 intervenes as a nucleophile to form an anhydride intermediate with the substrate (**d**) proposed mechanism in light of more recent data*
(12.21a, b, c – Reprinted from *The Enzymes*, 3rd ed., Vol. III, HARTSUCK J.A. & LIPSCOMB W.N., 1 Carboxypeptidase A, 49. © (1971) Academic press, with permission from Elsevier; 12.21d – Reprinted from *J. Mol. Biol.*, **163**, KUO L.C. et al., Catalytic conformation of carboxypeptidase A: the structure of a true reaction intermediate stabilized at subzero temperatures, 63. © (1983) with permission from Elsevier)

◤ For a long time, it was assumed that the mechanisms implicated in hydrolysis of esters and peptides by carboxypeptidase were different. On the one hand the Tyr248 residue did not seem to participate in the reaction; the acetylation of this group was shown to have no effect on the hydrolysis of esters. In fact, Tyr248 was localised to different positions in the molecule by radiocrystallographic studies. Its localisation in the free enzyme and in the enzyme-substrate complex differs by 12 Å; in the complex, this residue contracts a hydrogen bond with the NH group of the peptide bond of the Gly-Tyr dipeptide. The interaction of tyrosine 248 with Zn^{++} was observed in an arsanilazo derivative of the enzyme; finally it was found bound to the carboxylate of a ketonic substrate. On the other hand, during ester hydrolysis, a water molecule bound to Zn^{++} would seem responsible for the hydrolysis of the anhydride intermediate suggesting that, upon their association to the enzyme, these substrates would not chase completely the water molecule bound to Zn^{++}. However, the binding site S′1 seems common to the two types of substrates. The results of the group of CRAIK (GARDELL et al., 1985) invalidate the catalytic participation of Tyr248 in the hydrolysis of peptides. The authors obtained by site-directed mutagenesis the point mutant Tyr248 ⟶ Phe and evaluated the catalytic parameters for the hydrolysis of some peptide and ester substrates. Table 12.6 below gives the results compared to those obtained with the wild type enzyme. The modified enzyme conserves its catalytic activity; the most affected parameter is the MICHAELIS constant. The observed increase in the K_m values is compatible with the loss of two hydrogen bonds in the MICHAELIS complex. The authors concluded that the role of Tyr248 is not catalytic, but that this group is involved in the binding of substrates. Crystallographic data shows that residue Tyr248, by its oxygen, interacts tightly with the carboxylate group of the substrate; this situation suggested that its role would be to assist the release of the amino acid product of the reaction, from the active centre of the enzyme.

Table 12.6 Kinetic parameters of hydrolysis of several substrates by wild type carboxypeptidase and the mutant Tyr248 ⟶ Phe (Reprinted by permission from Macmillan Publishers Ltd: *Nature*, **317**, GARDELL S.J. et al., 551. © (1985))

Substrate	*Kinetic parameters*	*Wild type carboxypeptidase*	*Mutant Tyr248 ⟶ Phe*
Bz-Gly-Phe	$k_{cat}(s^{-1})$	17.7 ± 0.7	21.2 ± 0.2
	$K_m(\mu M)$	39.2 ± 5.1	199 ± 4
	$k_{cat}/K_m(M^{-1}.s^{-1})$	4.52×10^2	1.07×10^2
CbZ-Gly-Gly-Phe	$k_{cat}(s^{-1})$	51.6 ± 1.9	23.1 ± 1.6
	$K_m(\mu M)$	27.1 ± 3.4	194 ± 21
	$k_{cat}/K_m(M^{-1}.s^{-1})$	19×10^5	1.3×10^5
Bz-Gly-OPhe	$k_{cat}(s^{-1})$	1194 ± 62	1138 ± 41
	$Km(\mu M)$	95.6 ± 7.7	136 ± 13
	$k_{cat}/K_m(M^{-1}.s^{-1})$	125×10^5	83.8×10^5

Kinetic studies of the reaction in solution and at low temperatures were carried out with Zn^{++} carboxypeptidase and the enzyme in which the metal was substituted by Co^{++}, in order to specify the chemical nature of reaction intermediates and the catalytic conformation of the enzyme (KUO et al., 1983; GALDEN et al., 1986). According to KUO et al. (1983), the hydrolysis of esters and amides proceeds by the same mechanism, but the formation of a mixed anhydride is the limiting step in the hydrolysis of amides, and the deacylation is limiting in the hydrolysis of esters. In addition, the rotation around the C—O bond in the tetrahedral addition compound is required for breaking this bond. However, GALDEN et al. (1986) arrive at different conclusions and, according to their results, suggest that the intermediate formed is different during hydrolysis of esters and amides. ◢

The hypothesis formulated by PHILLIPS in 1990 is represented in Fig. 12.21d. The mechanism proposed relies on crystallographic data, site-directed mutagenesis, enzymatic studies and theoretical simulations. It involves a nucleophilic attack by the water molecule bound to Zn^{++}, assisted by the carboxylate of residue Glu270 that acts as a base catalyst. The oxyanion of the tetrahedral intermediate formed in the transition state would be stabilised by the positive charge of Arg127. The role attributed to Arg127 is supported by structural data of the LIPSCOMB's group who showed that this residue forms a hydrogen bond with the tetrahedral oxygen of the transition state analogs. In addition, the replacement of this group by site-directed mutagenesis by a methionine or an alanine brings about an important decrease of the catalytic constant. The ratio k_{cat}/K_m is lowered by a factor on the order of 10^4. The association energy of a transition state analog to the enzyme is decreased by 5.4 kcal . mol^{-1} whereas the interaction energy of an inhibitor in the fundamental state is only decreased by 1.7 kcal . mol^{-1}. In agreement with these experimental results, theoretical calculations carried out on the electrostatic contribution permitted the prediction that the positive charge of Arg127 stabilises the transition state from 6 to 8 kcal . mol^{-1}. In a further experiment (PHILLIPS et al., 1990), the authors obtained a partial restoration of the enzymatic activity by the addition of guanidine to mutant Arg127 $\longrightarrow$ Ala. The first step which drives the formation of the tetrahedral intermediate is the limiting step of the reaction. The hydrolysis of ester substrates like that of peptide substrates depends also on the contribution of Arg127 to catalysis. This mechanism which is currently the best supported by an ensemble of structural arguments and functional studies, seems the most plausible.

12.2. PHOSPHORYL TRANSFER ENZYMES

As was underlined in Sect. 11.3.1.2, the transfer of phosphoryl groups catalysed by enzymes can take place by a nucleophilic attack either on the α phosphorus of ATP or on the γ phosphorus, the attack on the β phosphorus being less frequent. The first type of mechanism is illustrated by reactions that catalyse amino-acyl tRNA synthetases, the second by reactions that catalyse kinases.

12.2.1. AMINO-ACYL TRNA SYNTHETASES

Amino-acyl tRNA synthetases catalyse two successive reaction steps; the first is the formation of an amino-acyl adenylate which consists of a nucleophilic attack of the α phosphorus by the amino acid carboxylate. This first reaction proceeds through the inversion of the α phosphate, without formation of a covalent intermediate. It produces an in-line attack driving a transition state in which phosphorus is pentacoordinated, followed by the release of the pyrophosphate. This mechanism was presented in a detailed manner in Sects. 11.3 and 11.4) in the case of tyrosyl tRNA synthetase, the enzyme for which the conjunction of structural studies, molecular modelling, and site-directed mutagenesis associated with enzymology studies, drove a very precise analysis of the reaction. The reader is deferred to this chapter.

The second reaction step, the charge of the amino-acyl adenylate on the tRNA is produced by a nucleophilic attack on the carbon of the ester phosphoric bond by the oxygen of the hydroxyl group in position 2′ with the release of AMP and formation of the amino-acyl tRNA. The migration from the position 2′ to 3′ is then carried out very rapidly.

12.2.2. *KINASES – PHOSPHOGLYCERATE KINASE*

Kinases reversibly catalyse the transfer of the γ phosphate group of ATP to a nucleophile acceptor. For some enzymes, other purine or pyrimidine nucleoside triphosphates can replace ATP. The presence of a metallic cation, generally Mg^{++}, is necessary for the reaction. In fact, in most cases the true substrate is Mg-ATP. The role of metal is to facilitate the catalysis; it acts as a super acid or an electrophilic catalyst, probably by stabilising the transition state. There are different types of kinases that one can distinguish according to the nature of the nucleophilic substrate. For hexokinase and phosphofructokinase the nucleophilic substrate is of the ROH type. For phosphoglycerate kinase it is of the R—COO^- type, for pyruvate kinase of the R—CO—COO^- type, for arginine kinase and creatine kinase of the RNH_2 type, and for protein kinases of the protein-serine OH type. In most of these cases, there occurs a direct transfer of γ phosphate on the nucleophilic substrate. For some protein kinases, the reaction would pass by a phosphoryl-enzyme intermediate, which is not the case for most other kinases.

Phosphoglycerate kinase is the first enzyme of the glycolytic pathway which catalyses the production of ATP by the transfer of a phosphoryl group from 1,3-diphosphoglycerate (1,3 diPG) to ADP with formation of 3-phosphoglycerate (3PG) according to the reaction:

$$\text{Mg-ADP} + \text{1,3-diPG} \rightleftharpoons \text{Mg-ATP} + \text{3PG}$$

The equilibrium constant of this reaction is in favor of the production of ATP:

$$K_{eq} = \frac{(\text{Mg-ATP})(\text{3PG})}{(\text{Mg-ADP})(\text{1,3-diPG})} = 3.4 \times 10^3 \text{ at } 25°\text{C}$$

In physiological conditions, this reaction is coupled to that which is catalysed by the preceding enzyme of the glycolytic pathway, glyceraldehyde-3-phosphate dehydrogenase:

$$\text{G3P} + \text{NAD}^+ + \text{P}_i \rightleftharpoons \text{3PG} + \text{NADH}$$

G3P being glyceraldehyde-3-phosphate. It follows that the ensemble of the two reactions equilibrates approximately at the physiological concentrations of substrates (SCOPES, 1973).

Phosphoglycerate kinase exists in a great variety of species and tissues, yeasts, microorganisms, muscles and red cells of mammals. It also exists in the plant kingdom where it is implicated in the taking up of carbon from CO_2. Indeed it makes part of the multienzymatic complex made up of ribose phosphate isomerase, phosphoribulokinase, ribulose bisphosphate carboxylate or Rubisco, phosphoglycerate kinase and glyceraldehyde-3-phosphate dehydrogenase.

12.2.2.1. STRUCTURAL PROPERTIES

The sequences of more than thirty phosphoglycerate kinases were determined, by either protein sequencing or gene sequencing. They show that this enzyme was particularly well conserved over the course of evolution. The three-dimensional structures of the enzyme of horse muscle (BANKS et al., 1979), yeast (WATSON et al., 1982), *Bacillus stearothermophilus* (DAVIES et al., 1993), pig (HARLOS et al., 1992) and *Trypanosoma brucei* (BERNSTEIN & HOL, 1998) were solved. Outside of some local differences, these structures present great similarities, particularly in the C-terminal part.

The molecule of molecular weight 44 500 is made of a single polypeptide chain folded into two domains of nearly equal size (Fig. 12.22). The elements of secondary structure are organised in repeating αβ motifs; the structure is of the α/β type. Each domain is made of six parallel β segments forming an internal sheet surrounded by helices. In total, 25% of the chain is implicated in the β structures and 42% in α helices. The twelve last residues of the polypeptide chain are strongly associated to the N-terminal domain. Residues 186–189 which join helix 7 and the segment βF (helix V and βG in the yeast enzyme) form the link between the two domains. The C-terminal domain comprises the ROSSMANN fold, characteristic of the NAD^+ binding domain of dehydrogenases. There is no sequence similarity between the two domains which differ also in the detail of their three-dimensional structure. The free enzyme presents a large groove between the two domains, an open structure in the absence of substrates, as was described for other kinases, in particular hexokinase.

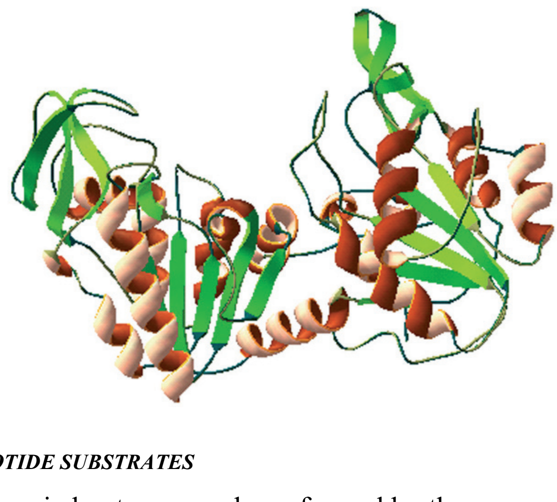


Fig. 12.22 Schematic representation of the phosphoglycerate kinase structure (PDB: 2PGK) showing the two structural domains

12.2.2.2. THE BINDING SITE OF NUCLEOTIDE SUBSTRATES

Radiocrystallographic studies were carried out on complexes formed by the enzyme with either Mg-ADP or MgATP. They reveal only a single binding site for nucleotide substrates; this site, located in the C-terminal domain, is identical for Mg-ATP and Mg-ADP. The enzymes of horse, yeast and *Bacillus stearothermophilus* show analogous characteristics. Figure 12.23 below illustrates the interactions of the horse enzyme with Mg-ATP. Bound to the enzyme, Mg-ATP like Mg-ADP, has a different configuration than in the ATP crystal. The base is in an *anti* configuration with respect to sugar, the ribose is C′2 *endo* and the C4′-C5′ exocyclic bond is left *trans*; the triphosphate group extends far from the adenine nucleus. This latter is completely buried in a very hydrophobic region of the site, its amino group likely contracting a hydrogen bond with the carbonyl oxygen of the polypeptide chain of Gly237. The adenine binding site is defined by the parts of the main chain consecutive to the segments βG (residues Gly212, Gly213 and Ala214), βH (Gly236, Gly237 and Gly238) and βJ (Val339, Gly340 and Val341). The side chains of Leu313 and Leu256 are in contact with the adenine nucleus. Ribose is localised in a depression close to Pro338, the hydroxyls in 2′ and 3′ forming hydrogen bonds with Gly343. In the triphosphate chain, only the α phosphate interacts with the ε-amino group of Lys219 forming an ion pair. In the native enzyme, this position is occupied by a sulfate or tartrate ion according to the composition of the crystallisation solution. α and β phosphates are situated 5 Å from the N-terminal part of helix 13, probably interacting with the positive dipole of the helix. The Mg-ADP substrate has the same configuration as Mg-ATP and the same interactions with the enzyme. The only difference resides in the position of the metal ion. In Mg-ADP, this metal is located between the α and β phosphates and the carboxylate of Asp374 of helix 13. In Mg-ATP, the situation is less clear; the metal interacts very likely with the β or the γ phosphate or both. Crystallographic studies only revealed a single binding site for the nucleotide substrates, and the existence of a second site was revealed by several other studies in solution.

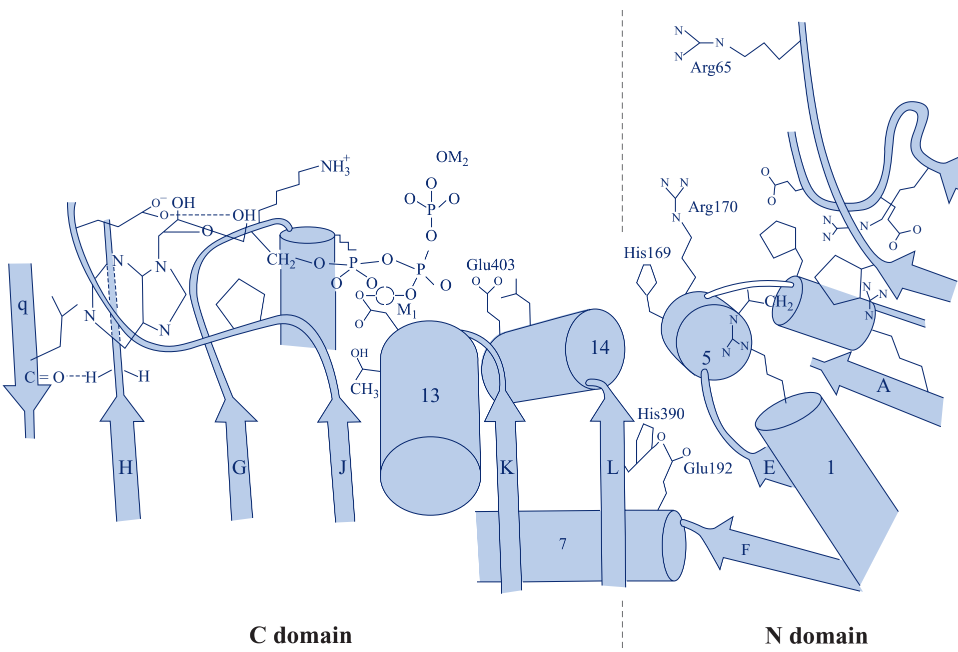


Fig. 12.23 Interactions of phosphoglycerate kinase with Mg-ATP
(Reprinted from *Phil. Trans. Roy. Soc. London*, **B293,** BLAKE C.C. & RICE D.W., Phosphoglycerate Kinase, 98. ©(1981) with permission from The Royal Society)

According to radiocrystallographic studies, nucleotide binding to the enzyme does not seem to be accompanied by significant conformational changes (BLAKE & RICE, 1981). However, studies of the fluorescence anisotropy decay of tryptophan residues of the enzyme reveal a displacement in helix 13 upon binding of Mg-ATP, permitting the free rotation of Trp335 which was impeded in the free enzyme (MOUAWAD et al., 1990).

12.2.2.3. THE BINDING SITE OF PHOSPHOGLYCERATE SUBSTRATES

The localisation of phosphoglycerate substrates was much less clear at the time of determination of the first structures. 1,3-diphosphoglycerate is too unstable to permit a crystallographic study of the complex. The only studies which attempted to localise the site of the second substrate were carried out with 3-phosphoglycerate. The localisation of this substrate on the enzyme remained however hypothetical, no precise structural data concerning the complex having been obtained. BLAKE and RICE (1981) suggested that 3PG is bound in the N-terminal domain between two arginine residues, Arg38 and Arg170 of the horse enzyme (Arg38 and Arg168 in the yeast enzyme); these two residues belong to a cluster of basic groups that include 5 arginines and 3 histidines. Data obtained by NMR showed also an interaction between 3PG and the cluster of basic groups. The substrate would be found thus at a distance of 12 to 15 Å from the γ phosphate. If such a localisation is correct, a relative movement of two domains is necessary for the chemical reaction to take place. From these considerations the group of BLAKE proposed a movement of domains around an axis localised between helix 7 and segment βF (*hinge bending motion*). A rotation of 20–25° would suffice to reduce the distance between the substrates from 12 to 5 Å; this rotation would be assured by a "scissor" movement of helices 7 and 14 (Fig. 12.24).

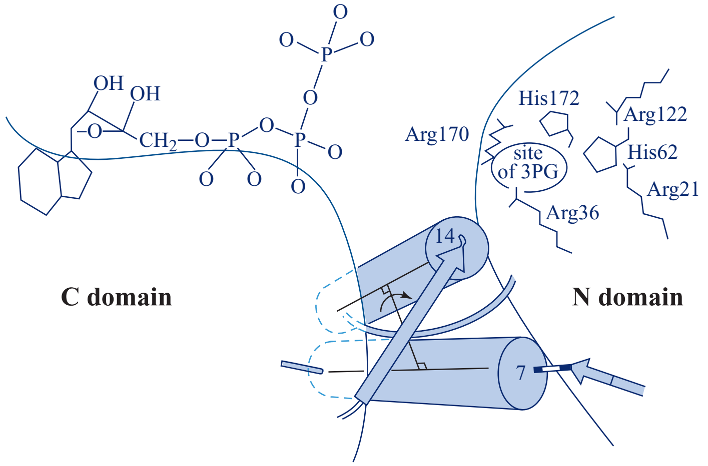


Fig. 12.24 Hinge bending movement which would bring together the two domains
(Reprinted from *Phil. Trans. Roy. Soc. London*, **B293,** BLAKE C.C. & RICE D.W., Phosphoglycerate Kinase, 98. © (1981) with permission from The Royal Society)

It would be induced by the binding of phosphoglycerate substrates as indicated in Fig. 12.25, the nucleotide substrates not bringing about conformational variations from the ensemble of the molecule.

In 1992, the group of BLAKE (HARLOS et al., 1992) succeeded in determining the structure of the complex of the pig muscle enzyme with 3-phosphoglycerate to a resolution of 2 Å. 3-phosphoglycerate is bound altogether by hydrogen bonds and electrostatic interactions with the cluster of basic residues of the N-terminal domain; the carboxylic group of phosphoglycerate points towards the C-terminal domain to which it is bound *via* a water molecule; this interaction involves nitrogen atoms of the main chain of helix 14 (Fig. 12.26 opposite). A rotational movement of 7°7 with respect to the free enzyme brings together the two domains. This value is weaker than that which had been predicted by the authors. Nevertheless, these results are in agreement with the preceding hypotheses of BLAKE.

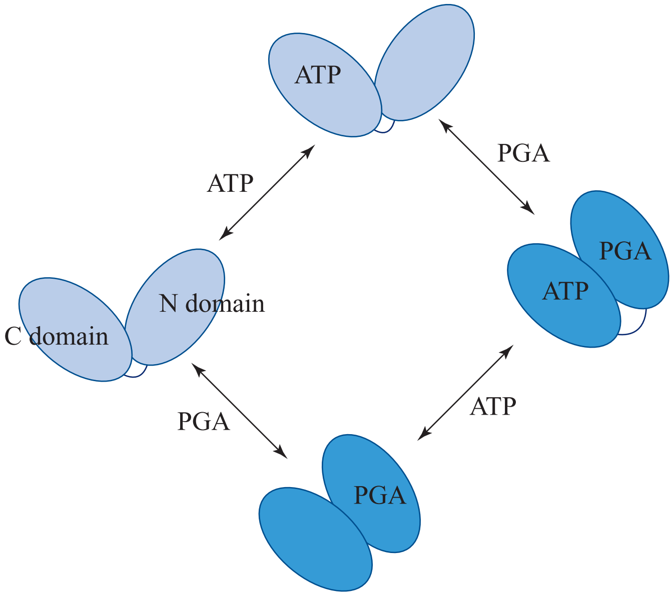


Fig. 12.25 Proposed hypothesis for the transition from the open to the closed structure of phosphoglycerate kinase

12.2.2.4. THE TERNARY COMPLEX AND THE MOVEMENT OF DOMAINS

For a long time, all the attempts to crystallise the ternary complex remained unsuccessful. In 1996, the group of BLAKE obtained the structure of the ternary complex of pig phosphoglycerate kinase with 3-phosphoglycerate and an analog of ATP, Mn-adenylimidodiphosphate. The same year, MCPHILLIPS et al. (1996) published the structure of a mutant of yeast phosphoglycerate kinase complexed with Mg-AMP-PNP and 3-phosphoglycerate. However, the movement of domains observed in the two cases was not more significant than in the binary complex with 3-phosphoglycerate and therefore was insufficient to permit the bringing together of substrates.

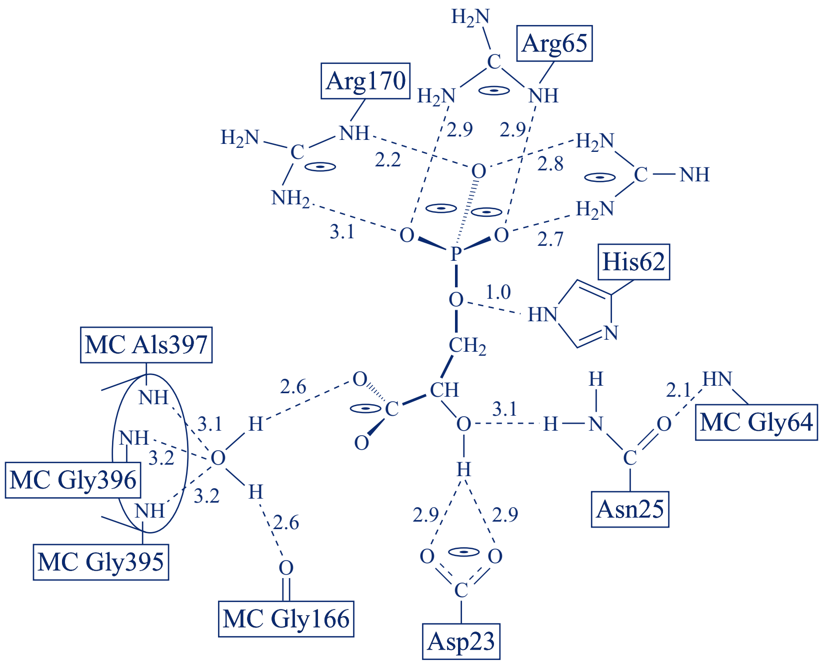


Fig. 12.26 Schematic representation of hydrogen bonds and electrostatic interactions of 3-phosphoglycerate with phosphoglycerate kinase of pig muscle

(From *Proteins: Struct. Funct. Genet.*, **12**, No. 2, HARLOS K. et al., 1992, 133–144. © (1992 Wiley-Liss, Inc.). Reprinted with permission of John Wiley & Sons, Inc.)

In return, several studies in solution had revealed major conformational movements upon binding of the two substrates. In addition, the group of PERAHIA showed by molecular dynamics simulations that the movement of the domains is an intrinsic property of the molecule that, in the absence of substrates, takes place in all directions. The binding of substrates by rigidifying the domains brings about a directionality of movement favouring the closure of the active site and bringing together the substrates (GUILBERT et al., 1995). In 1997, the group of VAN HOL (BERNSTEIN et al., 1997) obtained the structure of the ternary complex of the enzyme of *Trypanosoma brucei* in the presence of 3-phosphoglycerate and ADP; this shows a closure of the active site which brings together the substrates, thus allowing the enzymatic reaction.

12.2.2.5. CATALYTIC MECHANISM

Although phosphoglycerate kinase is a monomeric enzyme, the kinetics of the enzymatic activity do not follow the law of MICHAELIS in conditions of weak ionic strength. In return, at strong ionic strength, one observes a michaëlien behaviour. The deviation from the law of MICHAELIS recalls the behaviour of hysteretic or mnemonic enzymes (see Part V). In the case of phosphoglycerate kinase, this behaviour could result from the relative movement of domains, the return to the open form of the enzyme upon desorption of reaction products occurring according to slow kinetics, but no experimental proof has been produced. Whatever it may be, kinetic data obtained in conditions where the enzyme follows the law of MICHAELIS, indicate that the association of substrates to the enzyme occurs following a random mechanism to give the active ternary complex.

The catalytic mechanism does not proceed via the formation of a phosphoryl-enzyme intermediate. It consists of an attack in-line on the g phosphate of Mg-ATP with inversion of configuration and direct transfer of this group (WEBB & TRENTHAM, 1980). In the transition intermediate, the γ phosphorus is pentacoordinated. Figure 12.27 represents the most probable catalytic mechanism that involves one enzyme group as a general acid catalyst and another as a general base catalyst. The nature of the residues implicated in catalysis is not yet precisely known. The participation of His169 (His167 in the yeast enzyme), Lys215 (Lys213) and Glu403 (Glu401) was suggested but not really proven. Numerous site-directed mutagenesis experiments were undertaken with the yeast enzyme, in particular by the group of MAS. Most basic residues localised to the active centre were replaced. The results obtained up to now show that only residue Arg68 is critical for the activity. Residues His167, Arg168, Arg21 as well as His62 would be implicated in the binding of the anion activator.

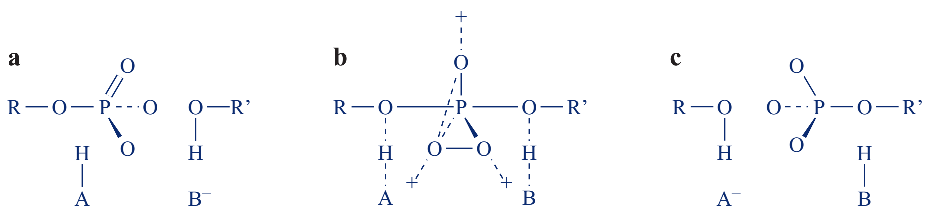


Fig. 12.27 Mechanism of the reaction catalysed by phosphoglycerate kinase
*(**a**) configuration of groups of atoms permitting the nucleophilic attack by phosphoglycerate – (**b**) transition intermediate in which the phosphorus is pentacoordinated – (**c**) products of the reaction*

12.3. GLYCOSYL TRANSFER ENZYMES – LYSOZYME

Glycosidases or glycosyl transferases catalyse the breaking of glycosidic bonds and the transfer of the glycosyl group to an acceptor which may be water. Among these enzymes, chicken egg lysozyme represents the best-known system on the structural and functional level.

The biological function of lysozyme is to hydrolyse bacterial cell walls. Bacteriological properties of egg white were recognised in 1909 by LASCHTCHENKO and lysozyme identified as responsible for these properties by FLEMING in 1922. There are numerous sources of lysozyme; besides egg white this enzyme is found in different tissues and secretions, in particular placenta, leukocytes, tears, saliva etc. There also are phage lysozymes. Since its discovery, chicken egg lysozyme was the object of many studies. This is the first enzyme for which the three-dimensional structure had been resolved (BLAKE et al., 1965, 1967; PHILLIPS, 1967). The substrate consists of alternating units of N-acetyl-glucosamine (NAG) and muramic acid (NAM) bound by glycosidic bonds 1 $\longrightarrow$ 4. Before crystallographic studies, neither the catalytic groups nor the mechanism of action of the enzyme were known. The examination of the structure of the enzyme and its complexes with inhibitors led to attribute a catalytic role to residues Asp52 and Glu35. These data stimulated a great number of following works.

12.3.1. STRUCTURAL PROPERTIES

The sequence of chicken egg lysozyme was established the same year (1963) by the group of JOLLÈS in France and the group of CANFIELD in England. The sequences of lysozyme from other species of vertebrates and bacteriophages were established later. The enzyme of chicken egg white is a small molecule of molecular weight 14 500 made of a polypeptide chain of 129 amino acids stabilised by four disulphide bridges (Cys6-Cys127; Cys64-Cys80; Cys76-Cys94; Cys30-Cys115). The three-dimensional structure (Fig. 12.28) shows that the enzyme is of the type ($\alpha + \beta$) according to the nomenclature of CHOTHIA. The molecule comprises an α structural domain with six helical regions, corresponding to 44% of the molecule, certain parts approaching a helix$_{3,10}$ conformation. Four antiparallel β segments are localised in the other domain. Residues 1-3 and 38-40 associated by two hydrogen bonds form a small antiparallel β sheet.

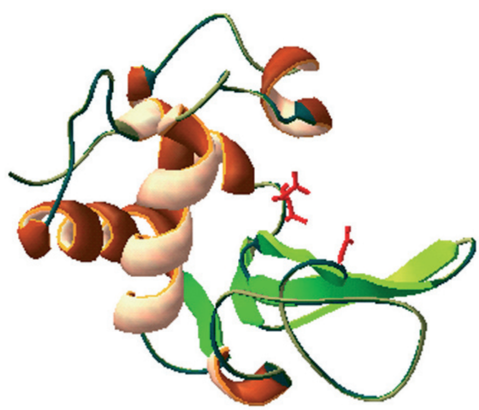


Fig. 12.28 Schematic representation of the chicken egg lysozyme structure (PDB: 6LYZ)

Figure 12.29 below represents the different hydrogen bonds that stabilise the segments of secondary structure. Three water molecules buried in the interior of the molecule were localised by radiocrystallography as well as a great number of water molecules situated at the surface.

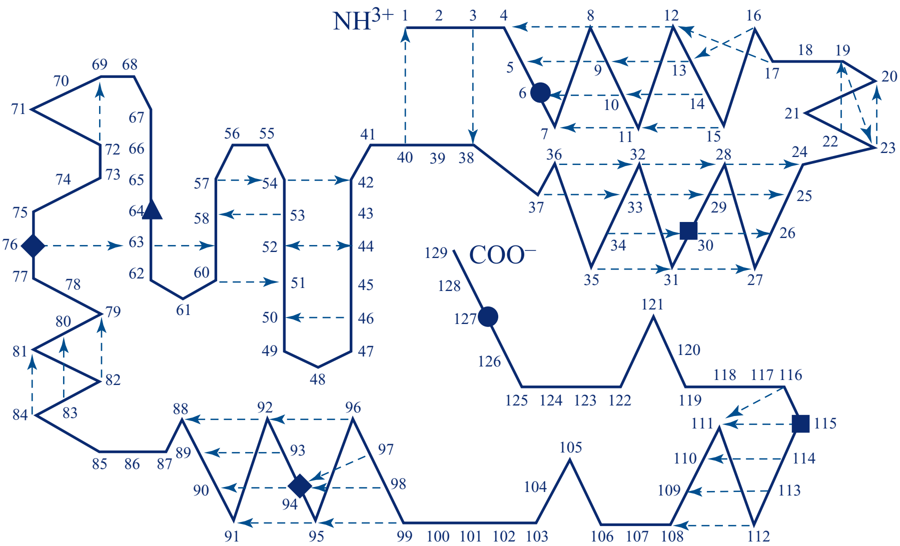


Fig. 12.29 Position of hydrogen bonds in lysozyme

(Reprinted from *The Enzymes*, 3rd ed., Vol. VII, IMOTO T. et al., Vertebrate lysozymes, 666-868. © (1972) Academic Press, with permission from Elsevier)

12.3.2. THE ACTIVE CENTRE

The active centre of the enzyme appears as a deep pocket crossing the molecule (see Fig. 12.28). Structural studies were carried out on complexes of lysozyme with diverse inhibitors, in particular (NAG)3; this latter forms a non-productive complex with the enzyme. Indeed, the minimum length of the substrate is a hexasaccharide, the cleaved bond being found between saccharide residues 4 and 5. Upon association of a trisaccharide to the enzyme, Trp62 undergoes a rotational movement of 0.75 Å towards the inhibitor, a movement which decreases the width of the binding site in the region of sub-site B. These conformational variations had previously been observed by difference spectrophotometry of the enzyme-substrate complex with respect to the free enzyme in solution. In addition, thermal relaxation studies had indicated the existence of an isomerisation of the enzyme-inhibitor complexes.

From the data obtained on enzyme-inhibitor complexes, a model of the hexasaccharide substrate was constructed and the complex was reconstituted by molecular modelling. The binding site of lysozyme substrates is formed from sub-sites A, B, C, D, E and F. Residues NAM are bound only to sites B, D and F, residues NAG to the other sites. It was proposed that non-covalent interactions between the enzyme and the saccharide ring in position D bring about the distortion of this latter which would take a half-chair configuration. Kinetic studies in solution gave an estimation of the interaction energies of different sub-sites with the substrate. Results obtained are given in Table 12.7 opposite; they seem to favor a constraint at the level of site D.

However, the theoretical calculations carried out by LEVITT (1974) have questioned the hypothesis of a substrate distortion in the fundamental state of the complex. It was suggested that the weak association of the substrate to site D results in fact in the displacement of two water molecules bound to the carboxylate of Asp52 in the free enzyme which are displaced upon complex formation. This was later confirmed by crystallographic studies (KELLY et al., 1979).

Table 12.7 Binding energies of substrates to sub-sites of lysozyme

Sub-site	*Bound residue*	*ΔG_{ass} (kcal . mol^{-1})*
A	NAG	–2
B	NAG	–3
	NAM	–4
C	NAG	–5
D	NAM	+3
	NAG	0
E	NAG	–4
F	NAG	–2

The saccharide bond is cleaved between sites D and E. It is localised between the carboxyl groups of Glu35 and Asp52 which are protonated and deprotonated, respectively (PARSON & RAFTERY, 1972).

12.3.3. CATALYTIC MECHANISM

The identification of the two residues Glu35 and Asp52 suggested two possibilities for the catalytic mechanism. One of these mechanisms had been proposed by KOSHLAND in 1953, well before the knowledge of structural data. He assumed a nucleophilic attack by Asp52 on the C_1 carbon of the saccharide ring to give an intermediate ester, a glycosyl-enzyme by analogy with the acyl-enzyme of serine proteases. According to this hypothesis (Fig. 12.30 below, pathway A), the carboxylate of Asp52 acts as a nucleophile to produce the covalent intermediate glycosyl-enzyme in a SN_2 reaction, with inversion of configuration. In a second step, a water molecule would intervene to hydrolyse the glycosyl-enzyme and a second inversion of configuration would restore the initial configuration of the enzyme.

However, this mechanism has been discarded in favor of a mechanism proposed by PHILLIPS on the basis of crystallographic data and modelling. According to PHILLIPS, the reaction would proceed with a retention of configuration according to a SN1 mechanism with the formation of a carbocation stabilised by the negative charge of Asp52 (Fig. 12.30, pathway B). The release of the product would occur following a general acid catalysis by the protonated residue Glu35. Water would intervene in a second step to form the second product of the reaction and to protonate residue Glu35.

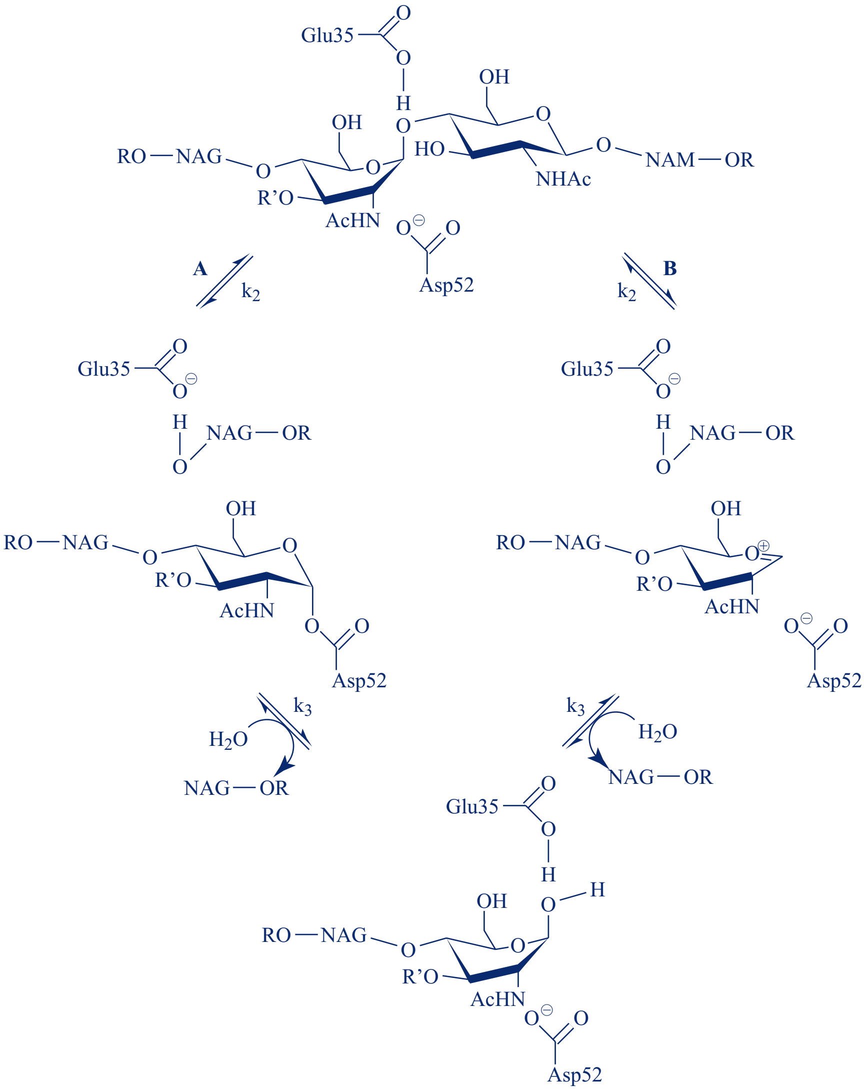


Fig. 12.30 Proposed mechanisms for lysozyme-catalysed reactions

Pathway A: formation of a covalent intermediate; pathway B: formation of a carbocation stabilised by the charge of Asp52

During more than forty years, diverse groups attempted without success to isolate and identify the intermediate. The difficulty resided in obtaining an intermediate which has a sufficiently long life time to permit the study. Recently, the proof of the validity of the first mechanism, that is the formation of a covalent intermediate, was carried out by VOCADIO et al. (2001). These authors succeeded to accumulate the covalent intermediate by combining two approaches, the first being the study of the reaction with a lysozyme mutant in which Glu35 was replaced by a glutamine (Glu35 $\longrightarrow$ Gln), the second being the use of a doubly fluorinated substrate. The covalent intermediate had a life span long enough to permit mass spectrometry studies, and to determine the crystal structure of this glycosyl-enzyme.

Such a mechanism seems valid for the ensemble of glycosidases.

This example shows once more that the knowledge of the enzyme structure alone is not sufficient to determine a catalytic mechanism.

12.4. OXYDOREDUCTION ENZYMES

In biological systems, oxidations are frequently coupled to the production of energy, and are carried out by electron transfers (see Chap. 11), which depend on the oxydoreduction potentials of the substrates (see Part I). Two oxydoreduction enzymes were chosen as examples; the first is a NAD^+ dehydrogenase, alcohol dehydrogenase, and the second a flavine enzyme, flavo-cytochrome b_2.

12.4.1. ALCOHOL DEHYDROGENASE

Alcohol dehydrogenase catalyses the oxidation of diverse primary and secondary alcohols in aldehydes and ketones in the presence of a coenzyme, NAD^+. The reaction proceeds through a ternary complex intermediate:

$$\text{enz-NAD}^+ + \text{alcohol} \rightleftharpoons \text{enz-NAD}^+\text{-alcohol} \rightleftharpoons \text{enz-NADH-aldehyde} \rightleftharpoons \text{enz-NADH} + \text{aldehyde}$$

The yeast enzyme and that of mammals differ in their specificity and their catalytic activity. Alcohol dehydrogenase of yeast is more specific for ethanol and acetaldehyde, whereas the mammalian enzymes have a much broader specificity. Among these latter, alcohol dehydrogenase of horse liver was the object of detailed structural and functional studies. Its three-dimensional structure is known with atomic resolution as well as that of enzyme-substrate complexes due to the work of the group of BRÁNDÉN in Uppsala, Sweden. The catalytic mechanism has been well established.

12.4.1.1. STRUCTURAL PROPERTIES

The horse liver enzyme is a dimer of molecular weight 80 000 constituted of two identical subunits. Each subunit possesses a binding site for NAD^+ and two strongly bound Zn^{++} atoms, one being essential to catalysis. The sequence of 374 amino acids of the polypeptide chain was determined by JÖRNVALL in 1970. Since then, other sequences of alcohol dyhydrogenases have been resolved. The human enzyme is a complex system of isoenzymes encoded by at least five genes. The sequence of the corn enzyme was deduced from the sequence of two genes ADH1 and ADH2. The known sequences of alcohol dehydrogenases of mammals, yeast, plants and bacteria permit the identification of conserved regions and differences. There is 87 to 89% homology between the human enzyme and that of horse. The horse enzyme and that of corn contain approximately 50% conserved residues; only 25% are identical between the yeast enzyme and that of mammals. The important residues at the catalytic site are conserved in all species as well as the catalytic Zn^{++} ligands. The residues that bind the coenzyme are less well conserved. Conversely, the entire region distant from the active site, which is not implicated in the association of the subunits, is found in all species. It includes residues 35, 77, 80, 86 and 87. Such a permanence suggests that these residues have an essential role in the folding of the catalytic domain.

Figure 12.31 shows the general organisation of the dimer molecule. This has the form of an ellipse of dimensions $40 \times 55 \times 100$ Å. In the absence of the substrate and the coenzyme, it presents an open structure. Each subunit is itself folded in two structural domains separated by a deep crevasse. One of the domains consisting of 143 residues binds the coenzyme; the other, made of 231 residues, contains the Zn^{++} ligands and most of the residues which are involved in the binding of the substrate. The two subunits are associated by the coenzyme binding domain.

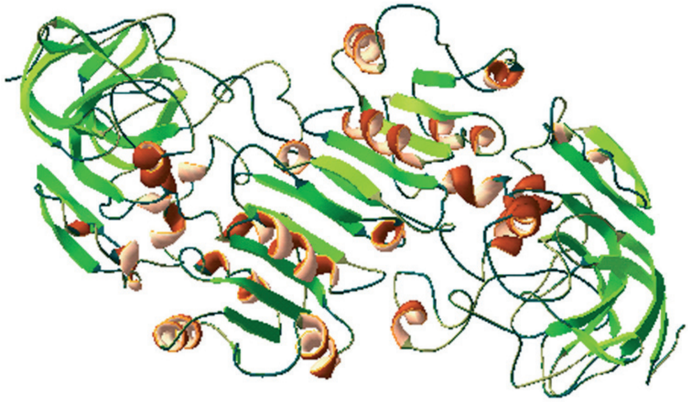


Fig. 12.31 Three-dimensional structure of horse liver alcohol dehydrogenase (PDB: 6ADH)

The NAD^+ binding domain has the typical α/β structure or ROSSMANN fold, with a regular sequence of six $\alpha\beta$ motifs, a central β sheet formed from six parallel β segments surrounded by the helices, the structure observed in most NAD^+ dehydrogenases. The comparison of the structures of malate dehydrogenase, glyceraldehyde-3-phosphate dehydrogenase, lactate dehydrogenase and alcohol dehydrogenase shows a remarkable similarity in the binding domain of NAD^+, whereas there is no sequence homology between these enzymes. In contrast, the catalytic domains totally differ. This observation had suggested to ROSSMANN et al. (1974) a relationship between these enzymes over the course of evolution. The dehydrogenases would have evolved by fusion of genes between ancestral polypeptide chains. The NAD^+ binding domain is common to these enzymes and would have fused with different catalytic domains. The discovery of introns and exons in eukaryotic genes suggested that the exons correspond to protein modules separated by introns. The intron-exon arrangement could thus constitute a mechanism that would permit an increase in the rate of evolution (EKLUND & BRÁNDÉN, 1984). On the basis of the position of introns in the gene of corn alcohol dehydrogenase, a diagram of the enzyme subunit is represented in Fig. 12.32.

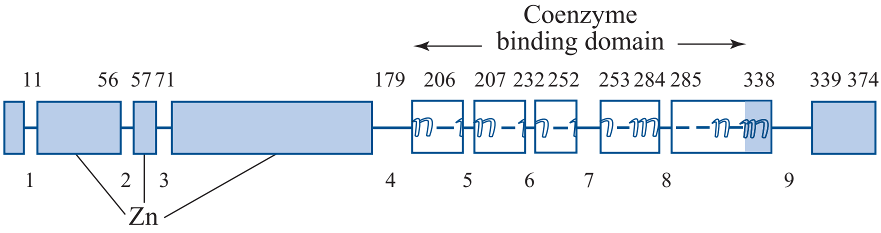


Fig. 12.32 Schematic representation of the polypeptide chain of alcohol dehydrogenase

The dark thin line *(non-framed) represents the position of the introns. The catalytic domain is indicated in* light blue

The catalytic domain is formed from two regions of the polypeptide chain localised between residues 1-174 and 319-374. The two Zn^{++} atoms are bound to ligands situated in this domain, the catalytic Zn^{++} by residues Cys46, His67 and Cys174. Each of these residues is encoded by a different exon. The second Zn^{++} atom is bound by four sulfur atoms belonging to residues Cys97, 100, 103 and 111; it is completely buried in the interior of the molecule. Figure 12.33 shows the arrangements of these two Zn^{++} atoms. The function of the second zinc atom is not known; a conformational role has often been attributed to it.

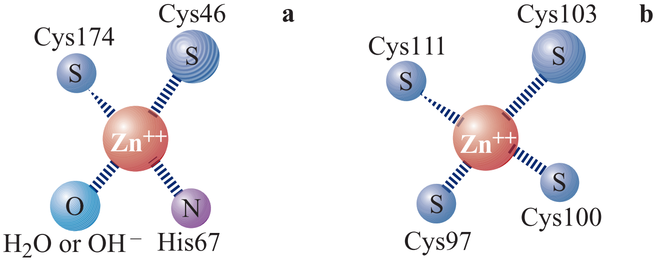


Fig. 12.33 Arrangement of the Zn^{++} atoms in alcohol dehydrogenase
*(**a**) catalytic Zn^{++} – (**b**) non-catalytic Zn^{++}*

12.4.1.2. CONFORMATIONAL CHANGE OF THE ENZYME INDUCED BY THE COENZYME BINDING

The binding of the coenzyme to alcohol dehydrogenase (LADH) brings about an important conformational change of the molecule that moves from an open to a closed form. The two binding domains of the coenzyme conserve the same orientation. The major change resides in a rotation of the catalytic domain with respect to the centre of the dimer (Fig. 12.34). However, the internal structure of the catalytic domains remains identical. This conformational change is therefore a rotational movement of one domain with respect to the other; this recalls the *hinge bending movement* described for phosphoglycerate kinase. Consequently, the interaction areas between the domains are modified.

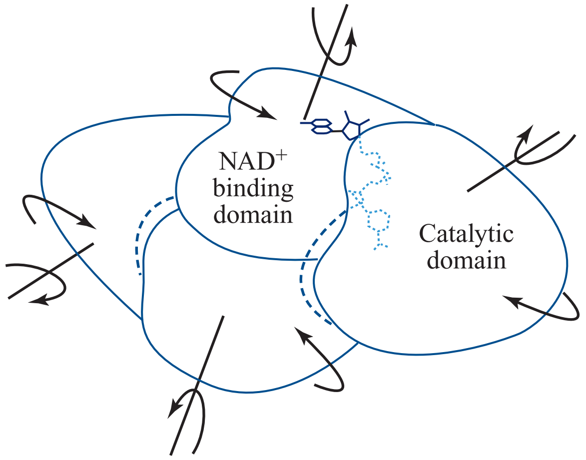


Fig. 12.34 Rotational movement of domains in alcohol dehydrogenase upon NAD$^+$ binding
(From Alcohol deshydrogenase in *Biological macromolecules and assemblies*, **3**, EKLUND H. & BRANDEN C.I., ed. by F.A. JURNAK & A.M. MCPHERSON, 73–142. © (1984 John Wiley and Sons). This material is reproduced with permission of John Wiley & Sons, Inc.)

The movements of the greatest amplitude are observed in this region. In particular, the loop comprising residues 292–298 of the binding domain of the coenzyme moves, as well as residues 51–58 of the catalytic domain (Fig. 12.35).

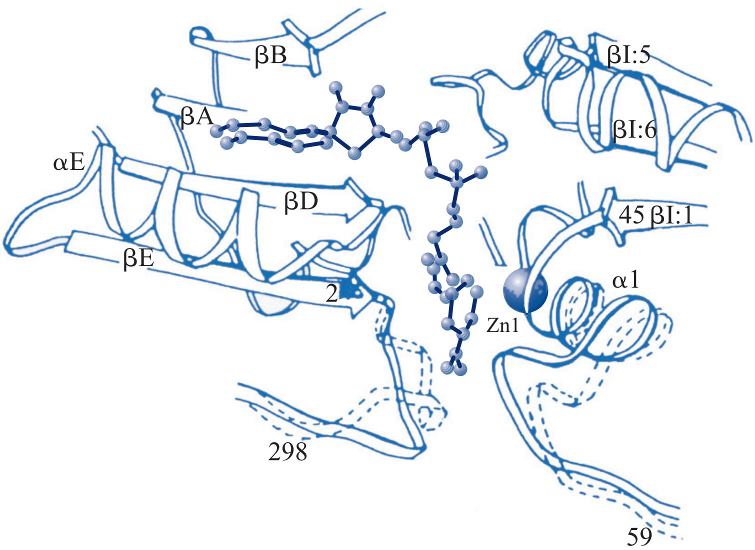


Fig. 12.35 Detail of alcohol dehydrogenase conformational variations upon NAD$^+$ binding (From Alcohol deshydrogenase in *Biological macromolecules and assemblies*, **3**, EKLUND H. & BRANDEN C.I., ed. by F.A. JURNAK & A.M. MCPHERSON, 73–142. © (1984 John Wiley and Sons). This material is reproduced with permission of John Wiley & Sons, Inc.)

This movement has the effect of burying the coenzyme more deeply into the enzyme molecule. The integrity of the coenzyme is necessary to induce the movement; the ADP ribose portion alone does not produce any change. The dissociation of the coenzyme, necessary for the return to the open structure of the molecule, constitutes the limiting step of the enzymatic reaction.

12.4.1.3. BINDING OF THE COENZYME

The structure of the enzyme was resolved in the presence of the coenzyme, the substrates and diverse inhibitors. The interactions of the enzyme with its different ligands are therefore well determined (Fig. 12.36).

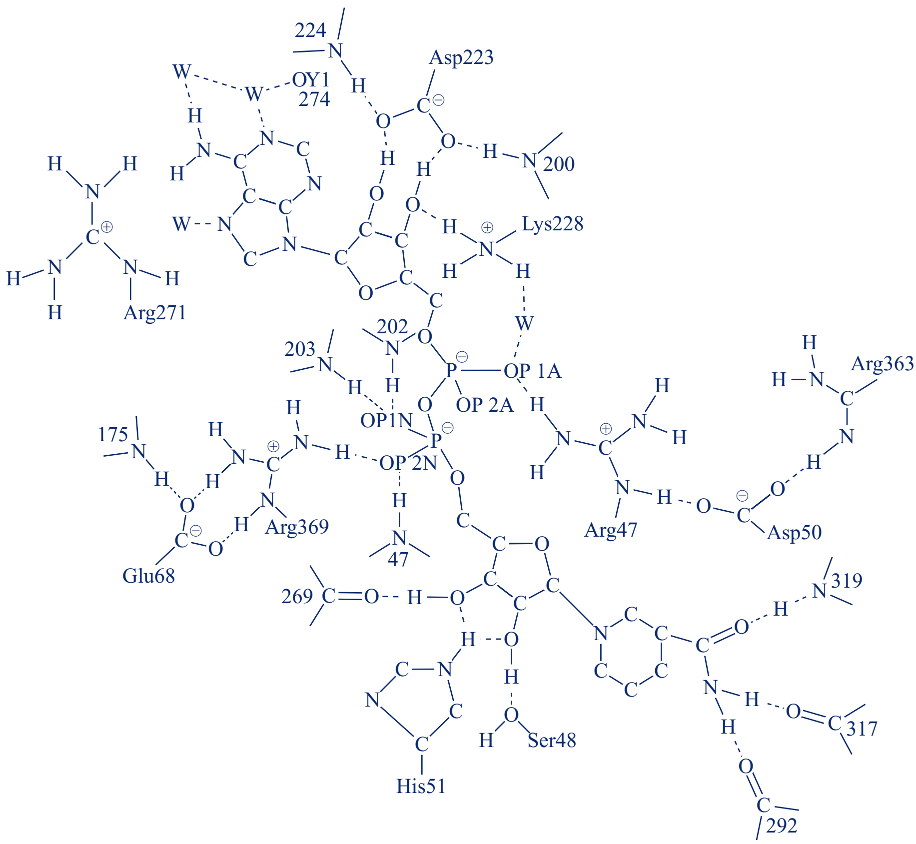


Fig. 12.36 Detail of interactions between NAD⁺ and the protein groups at the binding site

Hydrogen bonds are indicated in dotted lines. *They correspond to distances smaller than or equal to 3.5 Å between donor and acceptor groups*

The adenosine part is bound in the coenzyme binding pocket, and the adenine nucleus is in VAN DER WAALS contact with the side chains of Ile224 and Ile269. Nitrogen 6 is in contact with Asp271. Ribose is interacting with Asp223 and Lys228. The other parts of the coenzyme are bound in the region situated between the two domains of the same subunit. Pyrophosphate is localised in the central part of the crevasse between the two domains forming hydrogen bonds with nitrogens of the main chain and also with the side chains of Arg47 and Arg369. The dipole moments of the helices render this site very positive. The nicotinamide ring is in the crevasse at the interior of the protein 15 Å from the solvent. It interacts through one of its faces with Thr178, Val203 and Val294; the other face is directed towards the active site in proximity of the catalytic Zn^{++}. In ternary complexes, it interacts with substrates or inhibitors. The conformation of NAD^+ in the enzyme is represented in Fig. 12.35.

12.4.1.4. BINDING OF SUBSTRATES

The binding site of alcohol dehydrogenase substrates is larger and more hydrophobic than that of other dehydrogenases, in agreement with the larger specificity of LADH. The catalytic zinc is situated deeply in the interior of the subunit near the covalent bond which joins the two domains. It is found at the bottom of the substrate binding site. Zn^{++} ligands and Ser48 are the only polar groups of this region whose inner walls are made essentially of hydrophobic residues Leu57, Phe93, Phe110, Leu116, Phe140, Leu141, Leu294 and Ile318, as well as Met306 and Leu309 of the other subunit.

The mode of association of the substrate was studied by radiocrystallography in three different complexes, with trifluoroethanol, dimethylaminocinnamaldehyde, and a complex in equilibrium with bromobenzyl alcohol and bromobenzyl aldehyde. The substrate is bound to the catalytic Zn^{++} atom. It replaces a water molecule which, in the free enzyme, constitutes the fourth Zn^{++} ligand. The oxygen of alcohol interacts by hydrogen bond with Ser48 which itself is linked by hydrogen bond to His51. The system of hydrogen bonds extends to Ile269. Carbon 1 of the alcohol is near the nicotinamide ring. The other parts of the alcohol interact with the hydrophobic region of the site.

12.4.1.5. CATALYTIC MECHANISM

The enzymatic reaction proceeds following an ordered mechanism, the binding of the coenzyme preceding that of the alcohol [DALZIEL, 1975]. The transition of an open structure to a closed structure changes the properties of the active centre favoring the binding of the substrate. The mechanism of oxidation of alcohols is essentially an electrophilic catalysis by the Zn^{++}. One of the important points is the ionisation state of the alcohol in the complex. The proton cannot be localised by radiocrystallography. However, studies carried out in solution at various pH, using several substituted alcohols, led to establish that the substrate binds to the enzyme as an alcoholate ion. The system of hydrogen bonds including Ser48 and His51 would constitute a sort of shuttle for the migration of the proton. Zn^{++} would have the role of stabilising the alcoholate ion and orienting the substrate in a convenient position with regard to NAD^+. In the reverse reaction, Zn^{++} would act as an electron attractor increasing the electrophilic character of the aldehyde, which would facilitate the transfer of the hydride ion to the aldehyde (Fig. 12.37).

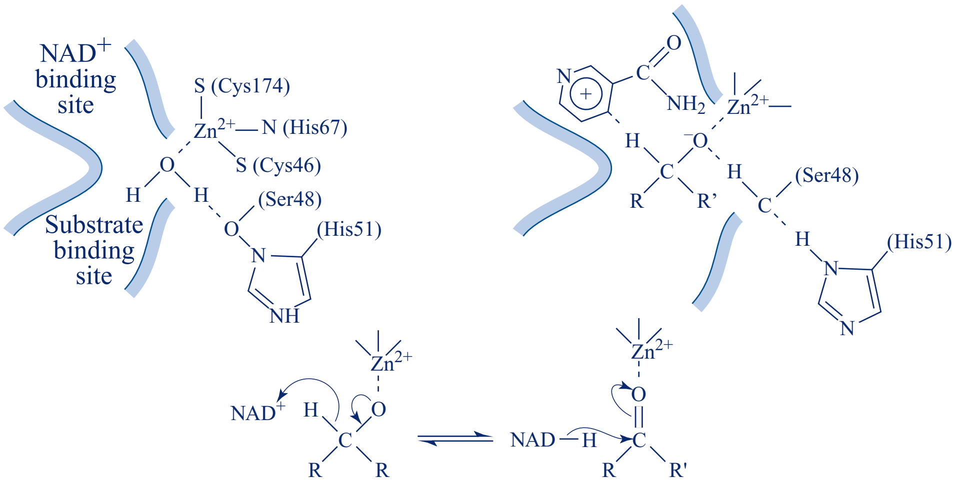


Fig. 12.37 Mechanism of action of alcohol dehydrogenase

Crystallographic data does not permit to establish a mechanism implicating a transfer of hydride. This was demonstrated by studies of isotope effects. Moreover, by molecular modelling it was shown that from very weak rearrangements in the position of bromobenzyl such as observed in the crystal, a correct geometry is assured for transferring the hydride between the substrate and the coenzyme. In addition, theoretical studies confirmed this mechanism.

12.4.2. FLAVOCYTOCHROME B₂

Flavocytochrome b_2 is a lactate ferricytochrome c oxydoreductase. This enzyme, localised in the intermembrane space of yeast mitochondria, catalyses the oxidation of lactate followed by the transfer of electrons to cytochrome c. It contains a heme, a flavine and a flavine mononucleotide or FMN, associated to the enzyme in a non-covalent manner. The amino acid sequence of the enzyme was established in France by the group of LEDERER in 1985; the three-dimensional structure was obtained to 3, then to 2.4 Å resolution by the group of MATTHEWS in the United States in 1987.

12.4.2.1. STRUCTURAL PROPERTIES

The enzyme is a tetramer of molecular weight 230 000 constituted of four identical subunits. Each chain contains 511 amino acids. The first hundred form a domain resistant to trypsin hydrolysis; this is the binding domain of heme or "cytochrome domain" whose sequence is homologous to that of mitochondrial cytochrome b_5. The four subunits are assembled according to a 4-fold symmetry axis (Fig. 12.38 below). Each subunit is itself organised into two domains, the heme binding domain (residues 1–100), the flavine binding domain (residues 101–486), and then a C-terminal part (487–511) containing a short portion of a helix that extends to the centre of the tetramer in contact with the three other subunits. Only two of the four heme binding domains are visible on the electron density map; the two others seem to present a disordered structure (Fig. 12.39).

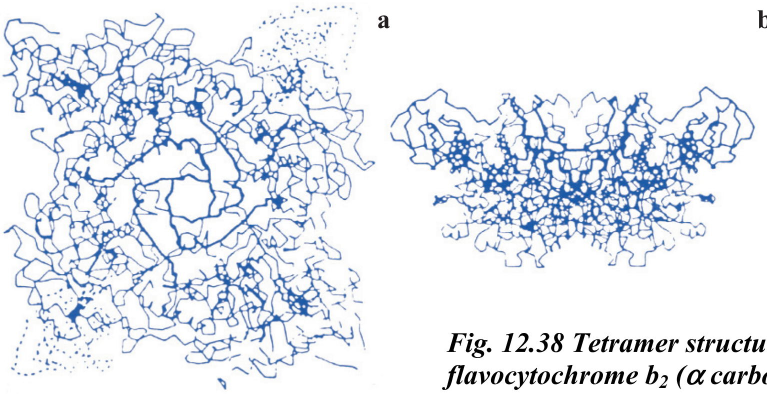


Fig. 12.38 Tetramer structure of flavocytochrome b_2 (α carbon tracing)

***(a)** view according to the 4-fold symmetry axis of the molecule – **(b)** view according to the vertical 4-fold symmetry axis.* (Reprinted from *Flavins and flavoproteins*, MATTHEWS F.S. & ZIA Z.X., 123. © (1987) Walter DE GRUYTER and C°, Berlin)

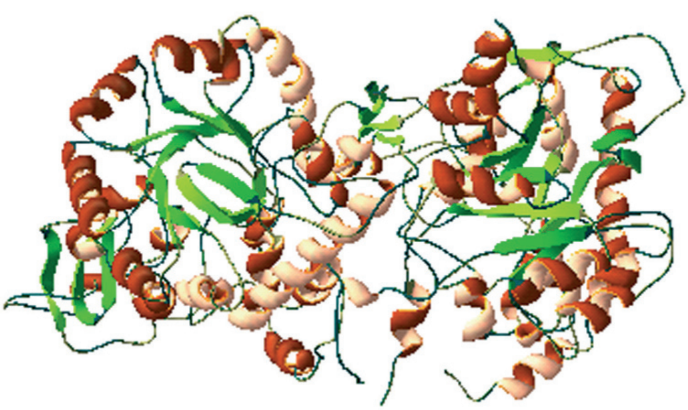


Fig. 12.39 Structure of flavocytochrome b_2 (PDB: 1FBC) showing the secondary structures

The cytochrome domains point towards the exterior. Each of them is folded similarly to the corresponding domain in cytochrome b_5; it is oriented such that the heme crevasse faces the flavine binding domain (Fig. 12.40 opposite). In the cytochrome domain, the heme crevasse has at the bottom a β sheet constituted of six β segments and the inner walls of two pairs of antiparallel helices, each pair being localised in each side of the crevasse. The flavine binding domain is constituted by the portion of the chain going from residue 192 to residue 465; it is folded into a structure of type α/β with eight αβ motifs. It comprised moreover a small segment of the N-terminal part (101–191) with three short helices. This organisation is similar to that observed in the flavine binding domain of other enzymes, in particular glycolate oxydase and trimethylamine dehydrogenase (TMADH), as is illustrated in Fig. 12.41 below. The interface between the two domains seems to be inaccessible to the solvent.

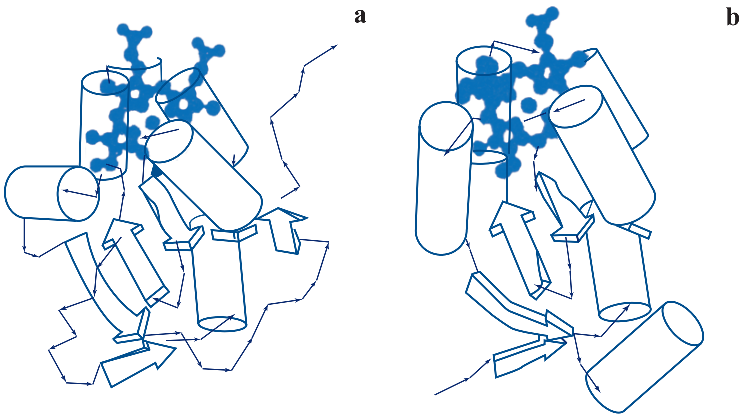


Fig. 12.40 Cytochrome domain
*(**a**) in flavocytochrome b_2 – (**b**) in cytochrome b_5*
(Reprinted from *Flavins and flavoproteins*, MATTHEWS F.S. & ZIA Z.X., 123. © (1987) Walter DE GRUYTER and C°, Berlin)

12.4.2.2. BINDING OF HEME AND FLAVINE

The heme is situated at the interface between the two domains which fits into the hydrophobic part of the N-terminal domain. The iron ligands of heme are the N^{ε} of histidines 43 and 66. The propionate groups extend towards the centre of the molecule and the flavine. Each contracts a hydrogen bond with a tyrosine, one with Tyr97 of the cytochrome domain, the other with Tyr143 of the flavine binding domain.

Flavine lies at the C-terminal extremity of the central sheet. The isoalloxazine nucleus is inclined and its plane is approximately parallel to segment β1. The pyrimidine nucleus is near the axis of the β barrel and the benzene ring near the surface. The ribityl chain is situated between segments β7 and β8. The phosphate group is near the N-terminal extremity of a short helix situated just after segment β8. FMN establishes contacts with the main chain and the side chains of residues situated in six of the eight β segments. The phosphate group contracts salt bridges with Arg413 and Arg433 (in β8) and hydrogen bonds with peptide nitrogens of residues 432 and 433 (in β8). The O^3 oxygen of the ribityl group interacts with Asp409 (in β7), the O^2 oxygen with Ser195 (in β1), Lys349 (in β5) and the carbonyl oxygen of residue 196 (in β1). The N^1 nitrogen and the O^2 oxygen of flavine interact with Lys349 (in β5) and Thr280 (in β4), respectively. The N^3 nitrogen interacts with Glu252 (in β3) and the O^4 oxygen with Ser228 (in β2). The N^5 nitrogen interacts with the peptide nitrogen of residue 198 (in β1). The plane of flavine forms with that of the heme an angle around 17°. The distance of the heme iron to the centre of the isoalloxazine nucleus is around 16 Å.

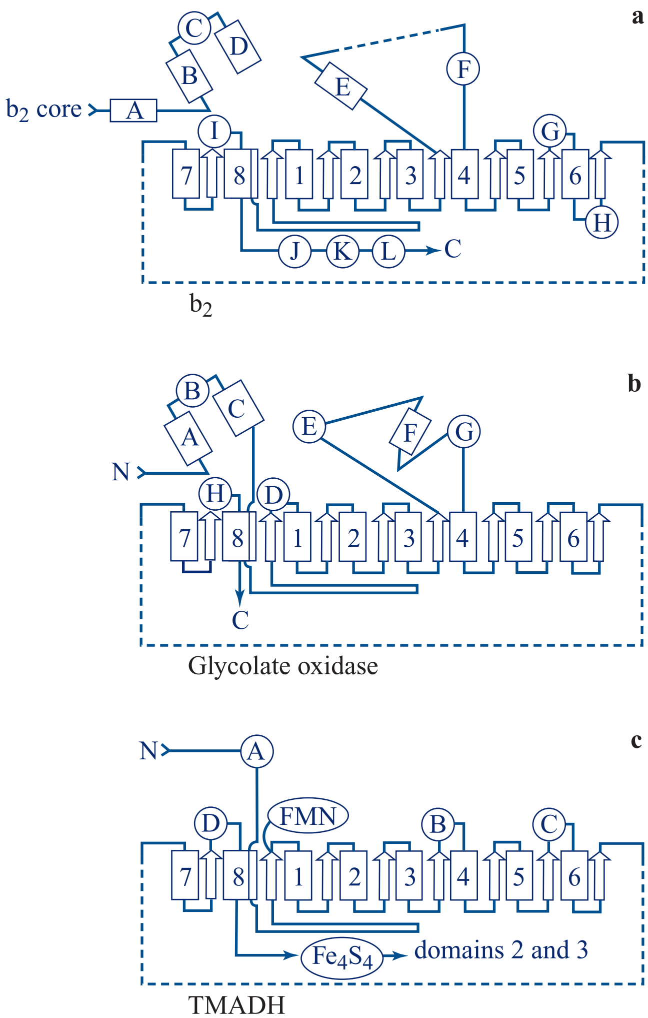


Fig. 12.41 Topology of the flavine binding domain in three FMN enzymes
*(**a**) flavocytochrome b_2 – (**b**) glycolate oxydase – (**c**) trimethylamine dehydrogenase*
(Reprinted from *Flavins and flavoproteins*, MATTHEWS F.S. & ZIA Z.X., 123. © (1987)
Walter DE GRUYTER and C°, Berlin)

12.4.2.3. BINDING OF THE SUBSTRATE

An isolated zone of electron density, localised in one of the subunits where the cytochrome domain is disordered, was attributed to pyruvate. The orientation of pyruvate was deduced from this. The carboxylate group of the substrate would interact with residues Arg376 and Tyr143 whereas the ketonic oxygen would be near residues Tyr254 and His373. The substrate is found thus in a position such that its carboxylate is near the benzene ring of the cofactor, its plane being practically parallel to that of flavine (Fig. 12.42 opposite).

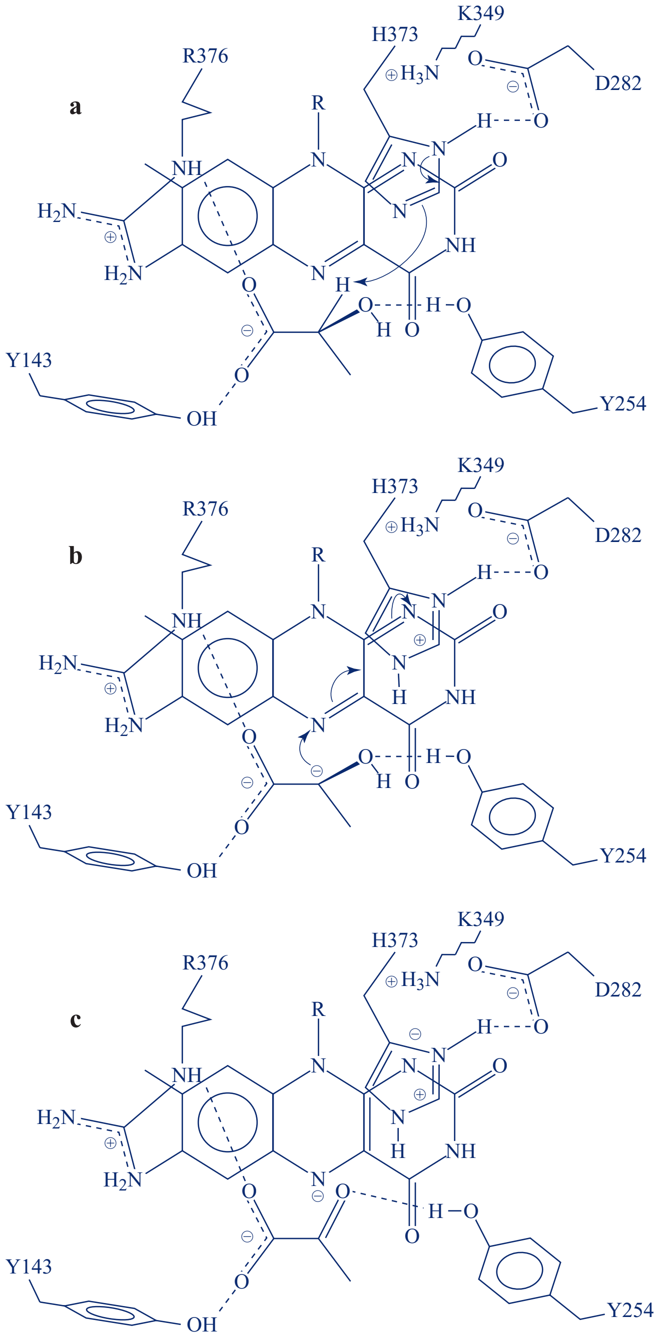


Fig. 12.42 Interactions of the substrate at the active centre of flavocytochrome b_2 and reaction mechanism *(courtesy of Florence LEDERER)*

12.4.2.4. REACTION MECHANISM

Flavocytochrome b_2 or L-lactate ferricytochrome c oxydoreductase catalyses the electron transfer from lactate to cytochrome c. The first step is the dehydrogenation of lactate to give pyruvate with reduction of flavine; it occurs according to a $2e^-$ transfer step process. Reoxidation of flavine takes place by a twice $1e^-$ transfer step process, the transfer being carried out from flavine to the heme, then to cytochrome c. Detailed kinetic studies permitted the establishment of diverse steps of the reaction and the determination of their rate constants (CAPEILLÈRE et al., 1975). Figure 12.43 illustrates the reaction pathway.

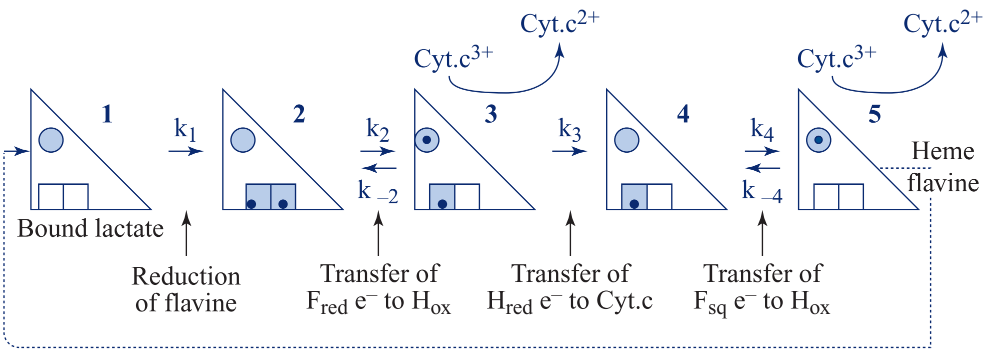


Fig. 12.43 Reaction pathway of the electron transfers from lactate to cytochrome c
(Reprinted from *Eur. J. Biochem.*, **54,** CAPEILLIAIRE-BLANDIN C. et al., Flavocytochrome b_2: Kinetic Studies by Absorbance and Electron-Paramagnetic-Resonance Spectroscopy of Electron Distribution among Prosthetic Groups, 549. © (1975) with permission from Blackwell Publishing)

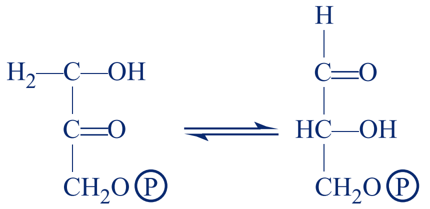
The first step of the reaction is the abstraction of a hydrogen in α. From an ensemble of experiments including the use of halogenated substrates such as bromolactate the group of LEDERER proposed a mechanism which implicates the extraction of a proton and the formation of a carbanion intermediate (see Chap. 11). The structural data confirmed this reaction mechanism and permitted specifying the details as well as the role of different enzyme groups in the reaction (see Fig. 12.42). This would implicate the following events: the binding of the carboxylate of the substrate by an electrostatic interaction with Arg376 and a hydrogen bond with Tyr143, a general acid-base catalysis on the C_2 by His373. Tyr254 stabilises the transition state. The entry of electrons in the cofactor is facilitated by Lys349 of which the positive charge stabilises the anionic character of the reduced flavine.

The transfer of electrons between flavine and heme is carried out certainly at the interior of a same subunit. Indeed, the distance between the edges of heme and flavine being 10 Å and their respective orientation being near coplanar, such a situation favors a rapid transfer of electrons as is revealed by kinetic studies (75–500 s^{-1}). The distance between the cofactors belonging to two different subunits is at least 45 Å and therefore an inter-subunit transfer would be much too slow.

The intermolecular transfer of electrons between the heme of flavocytochrome b_2 and cytochrome c is the last step of the reaction. These two proteins are capable of associating to form a complex with a stoichiometry of a heme c by heme b_2 facilitating thus the electron transfers.

12.5. TRIOSE PHOSPHATE ISOMERASE

Triose phosphate isomerase is an enzyme of the glycolytic pathway which catalyses the interconversion of dihydroxyacetone phosphate into D-glyceraldehyde-3-phosphate according to the reaction:



In the metabolic pathway, the reaction is favored in the forward direction, glyceraldehyde-3-phosphate being rapidly transformed. This enzyme is an aldose-ketose isomerase. The structures of several triose phosphate isomerases have been determined and many functional studies were reported.

12.5.1. STRUCTURE OF THE ENZYME

The amino acid sequences of triose phosphate isomerases (TIM) were determined in a great variety of prokaryotic and eukaryotic organisms. The first three-dimensional structures known for TIM were those of the chicken muscle enzyme (BANNER et al., 1975) and yeast (ALBER et al., 1981) solved to 2.5 and 3 Å, respectively. Since then, the structures of other triose phosphate isomerases were resolved: those of yeast (1.90 Å), of *Leishmania mexicana*, and of *Vibrio marinus*. In these different species the structures are very similar. The enzyme is a symmetrical dimer of molecular weight 53 000. The polypeptide chain of the chicken enzyme contains 247 amino acids; that of the yeast enzyme 248.

The three-dimensional structure (Fig. 12.44 below) is very typical with alternating β segments (22%) and α helices (55%) organised into a cyclic structure of type $(\alpha\beta)_8$, referred to as the *TIM barrel*. The practically parallel β segments are internal to the molecule, the α helices being situated at the periphery. These regular structures are practically superposable in the yeast and chicken enzymes, the differences residing in the position of irregular external loops.

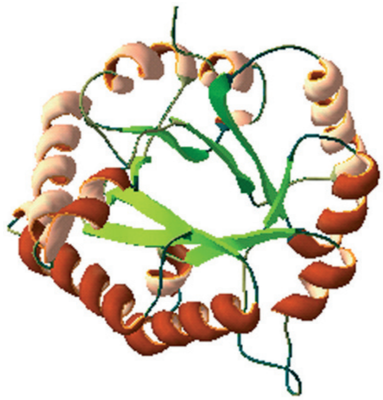


Fig. 12.44 Three-dimensional structure of a triose phosphate isomerase subunit (PDB: 8TIM) showing the $(\alpha\beta)_8$ or TIM barrel structure

12.5.2. STRUCTURE OF THE ENZYME-SUBSTRATE COMPLEX

The structure of the complex formed by triose phosphate isomerase and dihydroxyacetone phosphate was solved to 3.5 Å for the yeast enzyme and to 6 Å for the chicken enzyme. Although in the last case the resolution is not excellent, the structure of the complex formed by the yeast enzyme with a competitive inhibitor, phosphoglycohydroxamate was determined to 1.9 Å resolution. Previously, chemical labelling studies with affinity labels had permitted to attribute a catalytic role to Glu165. The structural data revealed that the substrate is effectively juxtaposed to residue Glu165. Residue His95 situated at the N-terminal extremity of the helix formed by residues 95–103 is in a position allowing its interaction with the oxygens in C_1 and C_2. It can also form a hydrogen bond by its nitrogen atom with the NH in the main chain of residue Glu97. Residue Lys12 of the yeast enzyme (Lys13 in the chicken enzyme) can also interact with the oxygen in C_2 or with that of the phosphoric ester. Residues 168 to 177 which belong to a mobile external loop, have their position modified and their mobility decreased in the complex.

12.5.3. REACTION MECHANISM

The dimer is the functional unit, however the enzymatic reaction follows the MICHAELIS law and does not show any cooperativity. **This enzyme approaches "catalytic perfection", the value of k_{cat}/K_m being on the order of magnitude of the diffusion rate of reagents ($k_{cat}/K_m = 2 \times 10^8\ mol^{-1} . s^{-1}$).**

The reaction proceeds through the formation of a cis-enediol intermediate [FERSHT, 1985]. Its existence was evidenced by isotope exchange experiments showing that the reaction proceeds by proton exchange with the milieu, excluding a direct transfer. Isotope effect experiments with tritium demonstrated the intramolecular transfer of the proton between C_1 and C_2. It had been deduced that this transfer is catalysed by a basic group of the enzyme; the proton is transferred on the same face of the substrate driving the formation of a cis-enediol. The conjunction of results from kinetic and spectroscopic studies had suggested the involvement of a residue susceptible to act as an acid catalyst according to the scheme given in Fig. 12.45.

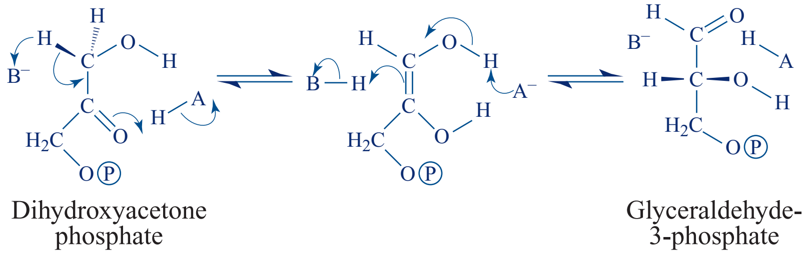


Fig. 12.45 Proposed mechanism for the reaction catalysed by triose phosphate isomerase mediated by an acid catalyst

Isotope exchange data (NICKBARG & KNOWLES, 1988) clearly showed that a rapid proton exchange with the solvent takes place at the level of the cis-enediol intermediate associated to the enzyme (Fig. 12.46).

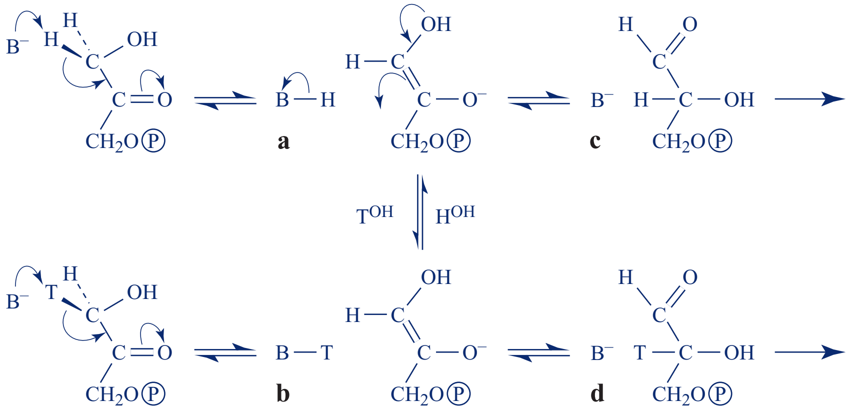


Fig. 12.46 Proton exchange with the solvent in the cis-enediol intermediate
(Reprinted with permission from *Biochemistry*, **27,** NICKGARD E.B. & KNOWLES J.R., 5939. © (1988) American Chemical Society)

A reaction mechanism was proposed on the basis of the crystallographic data; it involves a basic group of the enzyme, Glu165 which catalyses the extraction of the proton from C_1 and an acid group attributable to protonated His95 or maybe Lys12. NMR experiments showed indeed that His95 is not titrated between pH 5.4 and 9; however no formal attribution of non-titratable histidine residues has been carried out. The results from crystallographic studies are compatible with the establishment of a hydrogen bond between residue His95 and the hydrogen of the nitrogen amide of Glu97. On these bases, a mechanism was proposed in which the non-protonated histidine would catalyse the enolisation of the substrate by electrostatic stabilisation. In this case, there would not be a proton transfer from His95 to the oxygen. This alternative is presented in Fig. 12.47 below (a and b).

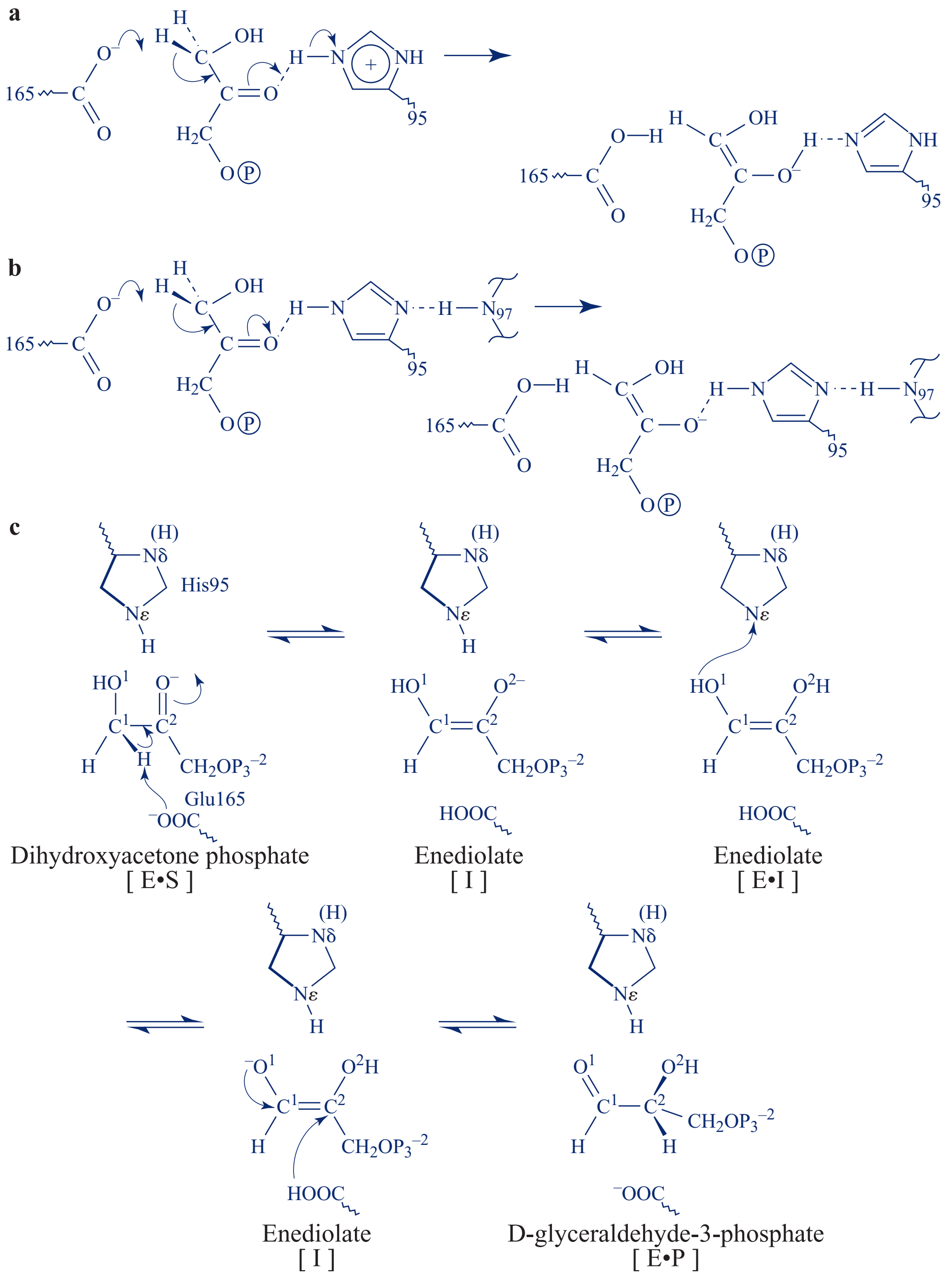


Fig. 12.47 Reaction schemes proposed from both crystallographic studies and NMR data *(**a**) histidine 95 would play the role of an acid catalyst – (**b**) histidine 95 would play the role of an electrostatic stabiliser – (**c**) histidine 95 under its neutral form would play the role of an acid catalyst* (Reprinted with permission from *Biochemistry*, **27**, NICKBARG E.B. et al., 5948. © (1988) American Chemical Society)

In order to specify the role of His95, an approach using site-directed mutagenesis (NICKBARG et al., 1988) consisted of replacing this histidine by a glutamine (mutant H95Q) in the yeast enzyme. This mutation brings about a decrease in the specific activity by a factor of 400. The inhibition constants determined for substrate analogs, 3-phosphoglycohydroxyacetate and 2-phosphoglycolate, indicate that these analogs have a respective affinity of 8 and 35 times weaker for the mutant than for the wild type enzyme. The analysis of isotope effects shows that there is practically no exchange between the intermediate and the solvent neither for the forward reaction, nor for the reverse reaction. These results suggested that the role of His95 would be to facilitate, by electrophilic assistance, the extraction by Glu65 of the proton in C_1 from dihydroxyacetone phosphate bound to the wild type enzyme. In the mutant, Glu165 would achieve this extraction without the catalytic assistance of His95 and therefore less efficiently, and the enediolate intermediate would be less stable.

The catalytic mechanism remained uncertain however. In 1991 KOMIVES et al., carried the proof of electrophilic catalysis by histidine 95. The experimental approach was driven by FOURIER transform infrared spectroscopy and the crystallographic analysis of mutants His95 $\longrightarrow$ Asn and His95 $\longrightarrow$ Gln. The decrease in the stretching frequency of dihydroxyacetone phosphate bound to the wild type enzyme revealed an electrophilic attack by the enzyme, which polarises the carbonyl group of the substrate towards the transition state for enolisation. No perturbation having been observed in the case of the two mutants, the electrophile catalyst can only be His95. Moreover, the crystallographic data showed that Glu165 is the only residue significantly displaced in the mutant His95 $\longrightarrow$ Gln, its new position permitting the proton transfer from C_1 to C_2. This role would be assured by this histidine in the wild type enzyme. Consequently, the authors conclude that residue His95 would play the double role of the electrophilic group and the general acid-base catalyst in the isomerisation of triose phosphate.

In 1991, DAVENPORT et al. determined at 1.9 Å resolution the structure of triose phosphate isomerase complexed to an analog of the reactive intermediate, phosphoglycolohydroxamate. The analysis of the refined structure permitted specification of the geometry of the active site residues and the interactions with the inhibitor and, by analogy, with the substrates. The study was completed by a theoretical simulation of the reaction pathway by the group of KARPLUS (1991). The data were in agreement with a mechanism of general acid-base catalysis in which the carboxylate of Glu165 extracts a proton from the carbon of the substrate during which His95, under its neutral form rather than its imidazolium form, gives a proton to oxygen to generate the enediol intermediate. The implication of this histidine under its neutral form was finally verified by NMR. Glu165 undergoes, in the complex, a movement of 2 to 3 Å with regard to its position in the free enzyme which brings it to a non-aqueous environment, increasing thus its basic character.

This mechanism is presented in Fig. 12.47c opposite, which also shows the enediolate intermediates. A theoretical analysis indicates that this mechanism seems the most energetically favorable (CUI & KARPLUS, 2001). Lys12 would orient or polarise the substrate oxygens by hydrogen bond or electrostatic interaction, stabilising the charged transition state. Asn10 would play a similar role.

The energy profile of the reaction was determined by an ensemble of studies. It corresponds to the following scheme:

$$E + D \rightleftharpoons ED \rightleftharpoons EX \rightleftharpoons EG \rightleftharpoons E + G$$

D, X, and G being respectively dihydroxyacetone phosphate, enediol intermediate and glyceraldehyde-3-phosphate. Figure 12.48 reproduces the free energy profile of the reaction catalysed by the yeast enzyme. This profile shows clearly that the steps of proton transfer are not limiting. The highest transition state corresponds to the binding step of the least stable substrate, glyceraldehyde-3-phosphate, which is the limiting step of the reaction; the rate 2×10^8 mol^{-1} . s^{-1}, corresponds to a diffusion process. The most stable state is that of the enzyme-dihydroxyacetone phosphate complex. The free energy profiles of the chicken enzyme and of the yeast enzyme are very similar despite the distance of the species in the phylogenetic tree and despite the sequence difference (50%). Triose phosphate isomerase seems therefore to have developed its catalytic potential very early over the course of evolution.

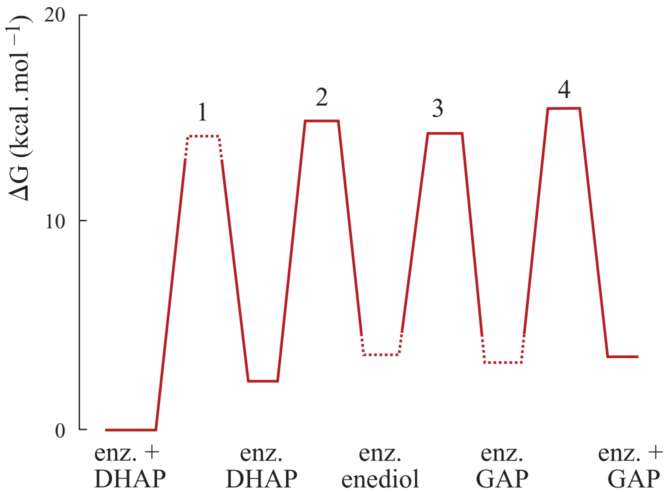


Fig. 12.48 Energy profile of the reaction catalysed by triose phosphate isomerase
(Reprinted with permission from *Biochemistry*, **30,** BASH P.A. et al., 5827. © (1991) American Chemical Society)

12.6. ASPARTATE AMINOTRANSFERASE

Aminotransferases are enzymes involved in the metabolism of amino acids. The metabolic scheme (Fig. 12.49 opposite) shows the role of aspartate aminotransferase with regard to the KREBS cycle and to the cycle of urea (from COOPER & MEISTER, 1989).

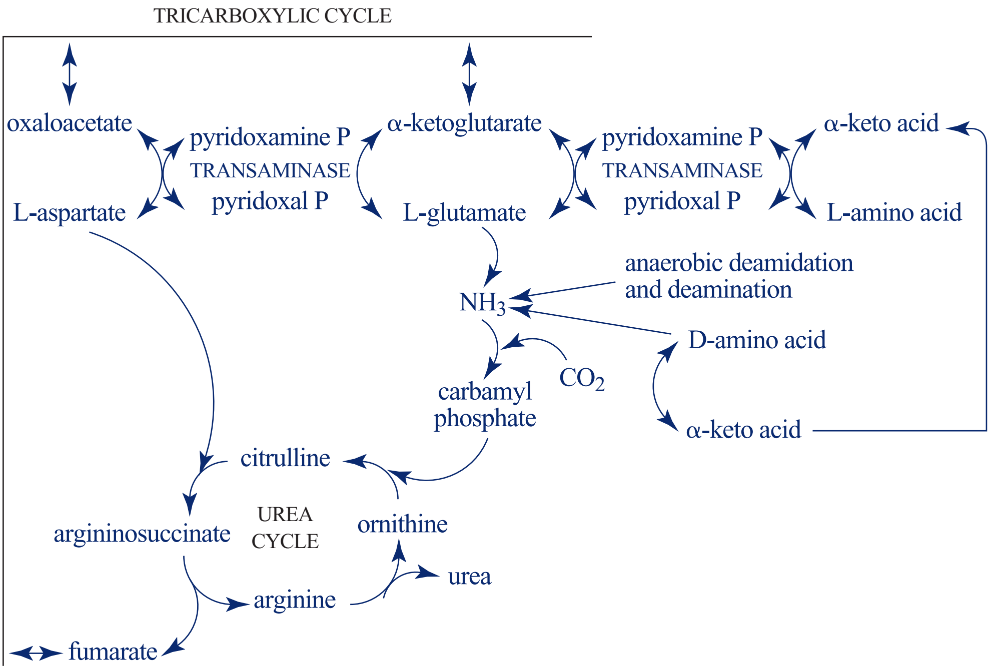


Fig. 12.49 Position of aspartate aminotransferase in cellular metabolism

These enzymes reversibly catalyse the exchange of the amino group between two amino acids through the intermediate of the corresponding keto acids by labilising the bond between the α carbon and the nitrogen of the amino group, according to the reaction:

$$\mathrm{R{-}\underset{NH_2}{\overset{H}{C}}{-}COOH} + \mathrm{R'{-}\underset{\|\atop O}{C}{-}COOH} \rightleftharpoons \mathrm{R{-}\underset{\|\atop O}{C}{-}COOH} + \mathrm{R'{-}\underset{NH_2}{\overset{H}{C}}{-}COOH}$$

This is carried out by the intermediate of a coenzyme, pyridoxal phosphate or vitamin B_6, and is decomposed in two half reactions:

$$\mathrm{R{-}\underset{NH_2}{\overset{H}{C}}{-}COOH} + \mathrm{E{-}C{=}O} \rightleftharpoons \mathrm{R{-}\underset{\|\atop O}{C}{-}COOH} + \mathrm{E{-}CH_2NH_2}$$

$$\mathrm{R'{-}\underset{\|\atop O}{C}{-}COOH} + \mathrm{E{-}CH_2NH_2} \rightleftharpoons \mathrm{R'{-}\underset{NH_2}{\overset{H}{C}}{-}COOH} + \mathrm{E{-}C{=}O}$$

in which the coenzyme is transformed from pyridoxal (E—C═O) to pyridoxamine (E—CH_2NH_2) (Fig. 12.50 below).

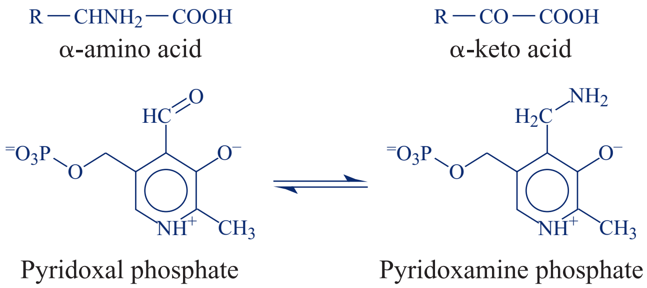


Fig. 12.50 Mechanism of transamination by pyridoxal phosphate

Transamination reactions were characterised by BRAUNSTEIN and collaborators in 1937, then by SNELL in 1945. These authors as well as the groups of METZLER and JENKINS carried a contribution decisive to the study of enzymatic transamination mechanisms. The coenzyme alone is capable of catalysing deamination reactions although with a much weaker efficiency; this property permitted its use as a model in the study of chemical mechanisms implicated in the reaction.

Each transaminase is specific for a couple of amino acids. Among these enzymes, aspartate aminotransferase catalyses the transfer of the amino group from L-aspartate to α-ketoglutarate (αKG) to give L-glutamate and oxaloacetate (OAA):

$$\text{L-Asp} + \alpha\text{-KG} \rightleftharpoons \text{L-Glu} + \text{OAA}$$

The kinetics of the reaction proceeds according to a Bi Bi ping-pong mechanism (see Chap. 5).

Aspartate aminotransferase of pig heart was the first pyridoxal phosphate enzyme which was purified (JENKINS et al., 1959). From this time, numerous enzymatic and physico-chemical studies of this enzyme were reported. Among these, one must cite the contributions of the groups of BRAUNSTEIN, KARPEISKY, OUVCHNINIKOV, FASELLA, YON, CHRISTEN, METZLER, MARTINEZ-CARRION and MEISTER. Since then, aspartate aminotransferases of other species were isolated and studied. In higher animals, there are two isoenzymes coded by two nuclear genes, a cytosolic form (cAAT) and a mitochondrial form (mAAT). This latter is synthesised as a precursor and imported in mitochondria where the signal peptide is released (see Chap. 7). The three-dimensional structure of the enzyme from several origins was obtained at high resolution. From these data, it was possible to specify the catalytic mechanisms.

12.6.1. STRUCTURAL PROPERTIES

Aspartate aminotransferase is a dimer of molecular weight 90 000 made of two identical protomers. Each protomer contains a molecule of pyridoxal phosphate (PLP) bound to the apoenzyme as a SCHIFF base (aldimine) with a lysine residue, Lys258, in the pig heart enzyme.

The complete sequence of several aspartate aminotransferases was established, including those of cytosolic and mitochondrial enzymes of pig heart (c and mAAT, respectively), of chicken (c and mAAT), as well as the rat enzyme (mAAt) and that of *E. coli*. Cytosolic enzymes consist of 411 or 412 amino acids per polypeptide chain; the mitochondrial enzyme is made of 401 amino acids and the *E. coli* enzyme of 396. Isoenzymes corresponding to different species present 85% identity whereas isoenzymes of a same species have only 46 to 48% identity. The *E. coli* enzyme is equally distant from the two types of isoenzymes presenting 40% identity with mAAT and cAAT of other species.

Radiocrystallography studies of the enzyme from several species were carried out by different groups of researchers, the cytosolic enzymes of chicken by the group of BORISOV, of pig by the group of ARNONE and METZLER, of yeast by the group of PETSKO, and the mitochondrial enzyme of chicken by the group of JANSONIUS. The enzyme structures of several microorganisms including *E. coli*, *Thermus thermophilus* and *Paracoccus denitrificans* were determined by the group of KAGAMIYAMA. These molecules present great structural similarities. The three-dimensional structure of the apoenzyme, the enzyme in the presence of PLP or pyridoxamine, as well as that of the enzyme in the presence of different substrate analogs like methyl aspartate, or dicarboxylic acids, competitive inhibitors, was obtained with a good resolution, as well as that of several mutants. One currently has access to 92 structures of aspartate aminotransferase from different species including the complexes and several mutants in the Protein Data Bank.

The structure of the mitochondrial enzyme of chicken (mAAT) bound to pyridoxal phosphate was obtained at 1.92 Å resolution (Fig. 12.51 below). The dimer molecule presents a 2-fold symmetry axis. Each subunit of identical structure has an ellipsoid form. It is folded in three parts, a stretched N-terminal region (residues 3 to 14), a small discontinued domain (15–47 and 326–410) and a large domain comprising residues 48 to 325 (the numeration used is that of the pig heart cytosolic enzyme). This latter is the domain of the coenzyme which is bound to Lys258 as an aldimine. Chicken mAAT, which only possesses 401 amino acids, presents deletions with respect to cAAT of pig heart in the following positions: 1, 2, 65, 127, 128, 131, 132, 153, 407 and 412. The interface between the two subunits is established *via* the binding domains of the coenzyme. The N-terminal part of the chain interacts with the base of the binding domain of the coenzyme of each subunit, contributing thus to the stability of the dimer. Each subunit comprises 16 helices including a very long one of 50 Å, helix 13 (Fig. 12.51), and a large β sheet composed of six parallel β segments and an antiparallel β segment, as well as two β sheets each constituted of two β segments, parallel for one and antiparallel for the other. The global structure comprises 50% of α helix, 14% of β structure and 14% of loops. The large β sheet is found in the coenzyme binding domain. The interface between the two domains is flat. This permits, over the course of the enzymatic reaction, a sliding movement of the small domain on the surface of the coenzyme binding domain (to see later).

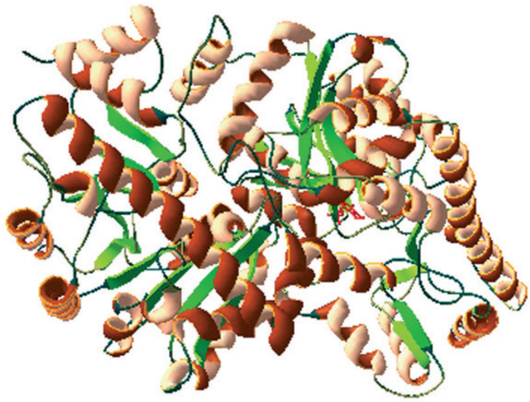


***Fig. 12.51 Three-dimensional structure of the aspartate aminotransferase dimer (PDB: 1TAR); PLP is indicated* in red**

The structure of the enzyme in the presence of pyridoxamine phosphate is slightly different. The only variations are localised to the immediate vicinity of the coenzyme. The inorganic phosphate occupies the same place as the phosphate group of the coenzyme. This structural analogy suggests that the binding of the coenzyme takes place without constraints.

12.6.2. BINDING OF THE COENZYME

The coenzyme PLP is bound to the enzyme under the aldimine form with lysine 258 (Fig. 12.52 opposite). Its face A is oriented towards the protein. The pyridine ring is surrounded by residues Ala224 and Tyr225 which contract a hydrogen bond with the hydroxyl group in 3′, as well as by residues Phe360, Asn194, Asp222 and Trp140. The $O^{3'}$ oxygen contracts a hydrogen bond with Asn194, the N^1 nitrogen forms a salt bridge with protonated Asp222. The phosphate group is associated to the enzyme by nine hydrogen bonds with, for OP^3 the OH groups of Ser107 and Ser255, the NH group of the main chain of Gly108, for OP^4 the NH group of the main chain and the OH of Thr109, the $N^{\eta 1}$ of Arg266, for OP^2 the $N^{\eta 2}$ of Arg266 and the phenolic group of Tyr70 of the other subunit. The role of three histidines 143, 189 and 193, situated under the coenzyme does not appear clearly yet; it could be related to the charge dissipation during the catalysis. According to the crystallographic data, the torsion angle of the phosphate group with respect to the pyridine ring is about–37°, i.e. smaller than what had been initially assumed.

When the coenzyme is under the pyridoxamine form, variations are observed in its orientation. The phosphate group remains localised and oriented as in PLP bound to the enzyme, but the aromatic ring is inclined 15° towards the front. The hydrogen bond with Asp222 remains. Rotations of weak amplitude between phosphorus and C^5 are sufficient for the change from one form to the other. This implies that the torsion angle C^4-C^5-$C^{5'}$-OP^1 passes through 0°, since in PMP, the OP^1 oxygen is behind face A of the pyridine ring with a torsion angle around 24°. Since PMP does not form a covalent bond with the enzyme, it can be inclined even more in front as that occurs in aldimine formed with methylaspartate.

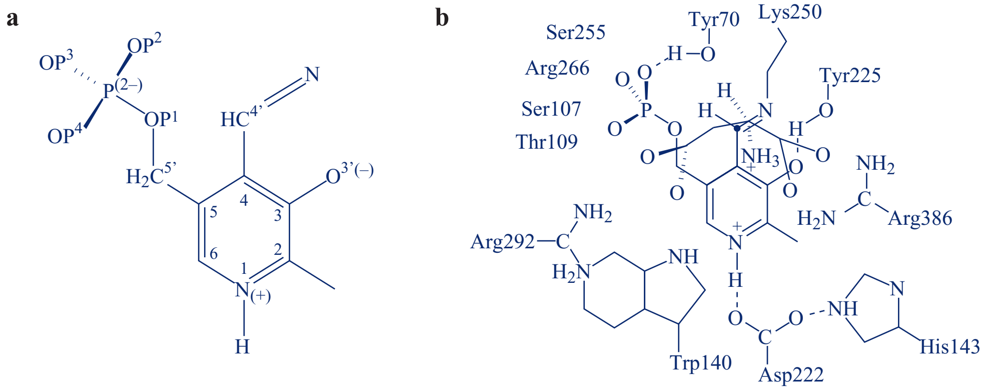


Fig. 12.52 (a) pyridoxal phosphate bound to the enzyme as a SCHIFF base (b) environment of the SCHIFF base and position of a dicarboxylic acid at the active centre of the enzyme

The inclination of the coenzyme is accompanied by movements of weak amplitude of the side chains of Tyr225 and Trp140, although the hydrogen bond between Tyr225 and $O^{3'}$ is maintained. The amino group of the coenzyme turns behind the plane of the pyridine ring with a torsion angle C^3-C^4-$C^{4'}$-$N^{4'}$ of –117°. The freed side chain of Lys258 extends towards the front, the NH_3^+ group establishing hydrogen bonds with the nitrogen of the amino group of the coenzyme and the carbonyl oxygen of Gly38.

12.6.3. BINDING OF THE SUBSTRATE: CONFORMATIONAL CHANGE OF THE ENZYME

Radiocrystallographic study of the enzyme in the presence of an inhibitor like maleate or the substrate analog, methylaspartate, showed that the binding of these ligands induces the closure of the enzyme structure. This movement corresponds to a rotation of 13° of the small domain towards the active site. It provokes the straightening of helix 13. During this conformational change, a sliding of the small domain occurs on the binding surface of the coenzyme, modifying the interface between the two domains. This involves residues 161–166, 192–199, 228–231 and 323–325 of the coenzyme binding domain, and residues 326–328, 348–364, 386–390 of the small domain. The residues of the coenzyme binding domain only undergo weak displacements, with the exception of the side chain of Phe228. The single hydrogen bond between the two domains (NH of Ile357 and O of Gly197) remains. In contrast, His352 undergoes a rotation of nearly 90° and Cys166 becomes accessible resulting in the "syncatalytic" increase in the reactivity of this cysteine, as was also observed in studies carried out in solution. The loop 37–39 undergoes a rearrangement. The largest movement affects helix 1 (residues 16–25) which orients towards the coenzyme, residues 15–18 blocking the entrance to the active site and enclosing the inhibitor or the substrate. The side chains undergo significant displacements. However, no change in the regular structures, helices or β segments, occurs.

12.6.4. THE ACTIVE SITE

The residues of the active site, His193, Phe360 and Arg386, are included in the interface between the two domains. There are important differences in the active site between the open and closed structures. In the closed structure whose formation is induced by maleate, the pyridine ring is reoriented following a rotation around N1-C4 and becomes coplanar with the double bond of aldimine; the torsion angle C5-C5′ is reduced to –16°. Following the movement of the coenzyme, the Trp140 nucleus is reoriented. Residues 37–39 undergo a movement which drives Val37 to establish a VAN DER WAALS interaction with Tyr70; the peptide bond 37–38 is oriented such that it contracts a hydrogen bond with the carbonyl of Gly38 and the εNH_3^+ group of Lys258 after transaldimination. The NH group of Gly38 forms then a hydrogen bond with the carboxyl which interacts with Arg386. The guanidinium group of Arg392 is reoriented towards the active site and contracts two hydrogen bonds with the carboxylate of maleate, whereas in the open structure it forms a salt bridge with residue Asp15.

12.6.5. CATALYTIC MECHANISM

The first step of the enzymatic reaction is the association of the aspartate substrate to the enzyme-PLP. In the coenzyme bound to Lys258 as an aldimine, the pyridine nitrogen N^1 is protonated and the 3′OH group deprotonated. When the aspartate is associated with the enzyme, it is oriented in the active site by electrostatic attraction of the group NH_3^+ towards $O^{3'-}$ of the coenzyme and by interaction of the α-carboxylate with Arg386 and of the β-carboxylate with Arg292. These charges are situated in such a way that the active centre of the enzyme can only accommodate L-amino acids. The β-carboxylate turns to a left orientation relative to the α-carboxylate, thus optimising the interactions. This is accompanied by the release of some water molecules and the transition towards the closed form of the molecule.

The non-polar environment created around the substrate increases the interactions of the charges, and the proton of the amino group of the substrate is yielded to aldimine which, once protonated, presents a characteristic absorption spectrum with a maximum at 430 nm (Fig. 12.53 opposite). The protonation of aldimine induces a rotation of the pyridine ring which turns –16° around the C^5—$C^{5'}$ bond and becomes practically coplanar with the double bond of aldimine. The α-carboxylate of the substrate remains implicated in a hydrogen bond with Arg386 and contracts another one with Asn194 ($N^{\delta 2}$). The remaining protons of the α-amino group are reoriented towards the carboxylates to form internal hydrogen bonds. It results that the pair of nitrogen electrons is oriented towards the $C^{4'}$ of the coenzyme permitting its nucleophilic attack. The first step of transaldimination is the transient formation of a diamine intermediate presenting an absorption spectrum centred at 340 nm. Then the non-protonated Lys258 is freed; the plane of the pyridine ring is inclined by 30° with respect to its position in the MICHAELIS complex and is clustered with Trp140 at a distance of 3.5 Å. It remains linked by hydrogen bond to Asp122. Tyr225, remaining bound by hydrogen bond to O′3, follows the movement of the coenzyme. The C^{α}—H bond has an optimal orientation, practically orthogonal to the plane of the pyridine ring. The NH_2 group of Lys258 hangs above the plane of the coenzyme; in this position, it can establish hydrogen bonds with OP^1, Tyr70, Tyr225 (O^{η}) and Gly38 (O). Hydrogen bonds with the last two oxygens place this group in proximity to C^{α}, permitting the transfer of the α proton to Lys258, which brings about the formation of a quinonoid intermediate. Indeed, the nitrogen and the C^{α} of the substrate are found then in the plane of the pyridine ring, forming a system of π electrons which extend to the α-carboxylate (intermediate III of Fig. 12.53).

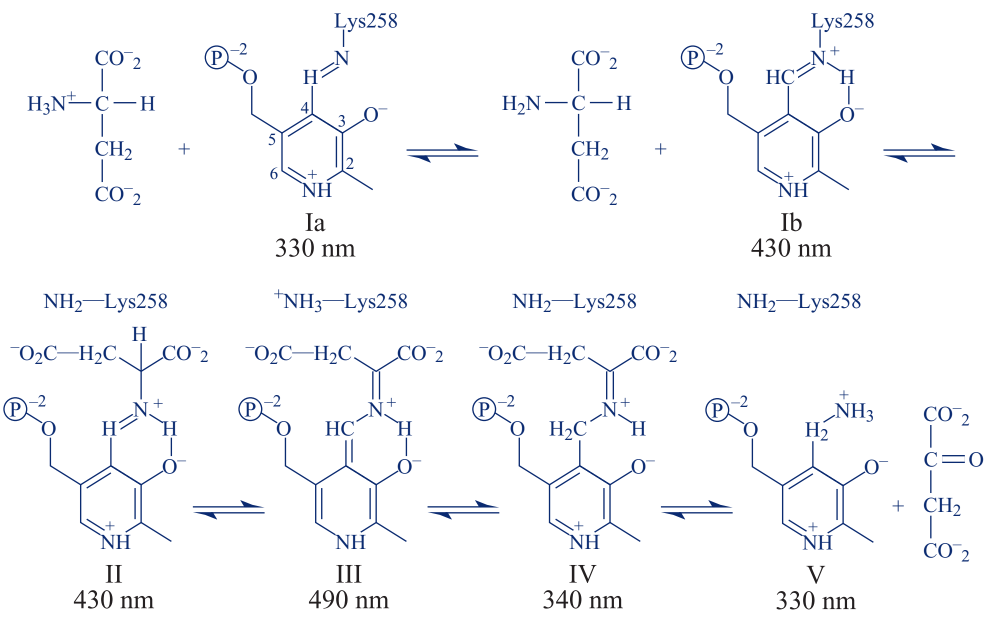


Fig. 12.53 Reaction intermediates during catalysis by aspartate aminotransferase

The transition of the external aldimine to this hyperconjugated form implicates a small rotation (< 10°) around the C^5—$C^{5'}$ bond. Trp140 follows this movement by a slight rotation which deletes its hydrogen bond with the β-carboxylate. The N^1 nitrogen in the sp^3 hybridation is uncharged; it can establish a strong hydrogen bond with the two oxygens of Asp222. This chemically unstable intermediate is thus stabilised by its interactions with the enzyme. It presents an absorption spectrum with a maximum at 490 nm (intermediate III).

The water molecule in position 7 can then enter the top of the active site, displacing Lys258 and protonating the $C^{4'}$ to give the ketimine intermediate (IV) which absorbs at 340 nm. This protonation splits the double bond between atoms N and C^{α}

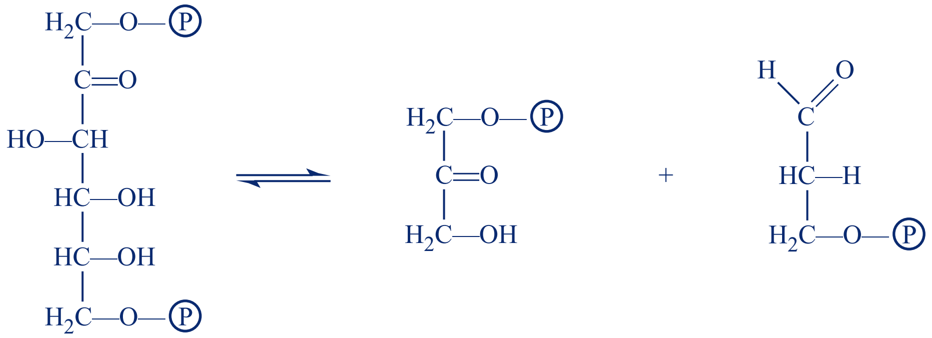
of the π electrons system in the pyridine ring. This latter is relaxed by a rotation towards the back which strengthens the hydrogen bond between Asp222 and the charged N^1 nitrogen. The part of the substrate which is coplanar with $C^{4'}$ optimises its position by establishing an internal hydrogen bond between the nitrogen of aldimine and the $O^{3'}$ oxygen, and another hydrogen bond between the β-carboxylate and Trp140 ($N^{\varepsilon 1}$). Lysine is displaced from its preceding position. It captures the proton of a water molecule situated in the vicinity. The OH group thus formed is oriented towards the C^{α} and produces a nucleophilic attack on this. The emergence of the sp^3 hybridation around the C^{α} repels the nitrogen that acquires a supplementary proton, probably originating from Lys258 before it is freed. The coenzyme turns towards the back. The nitrogen having taken back a second proton becomes oriented between the two carboxylates in order to optimise the hydrogen bonds. Lysine takes back a proton. The C^{α}—N bond thus broken produces oxaloacetate and the coenzyme as PMP. The transfer of the amino group to the α-ketoglutarate is carried out according to the reverse process.

On the basis of structural data, many questions concerning the catalytic mechanism of aspartate amino-transferase were resolved. The conception of mutants and their analysis permitted the specification of some aspects. CRONIN and KIRSCH in 1988 replacing Arg292 by an aspartate confirmed the key role of this arginine in the specificity of the enzyme for the substrate. The mutant Arg292 $\longrightarrow$ Asp has an activity (k_{cat}/K_m) decreased by five orders of magnitude for the transformation of its physiological substrates. Conversely, it presents an activity 9 to 16 times greater than that of the wild type enzyme for using cationic amino acids. The replacement of tyrosine in position 70 (Tyr70Phe) permitted the specification of the role of this residue. The OH group of Tyr70 stabilises the transition state by 2 kcal . mol^{-1}. In addition, the benzene ring is essential to the selectivity of the enzyme for L-glutamate and 2-oxoglutarate. His143 is not necessary for the catalysis but is involved in the formation of the enzyme-substrate complex. Studying mutants in which either Arg292 or Arg386 was replaced by a leucine, MIZAGUSHI et al. (2001) concluded that the constraints are more important than the electrostatic interactions in the control of the SCHIFF base pK.

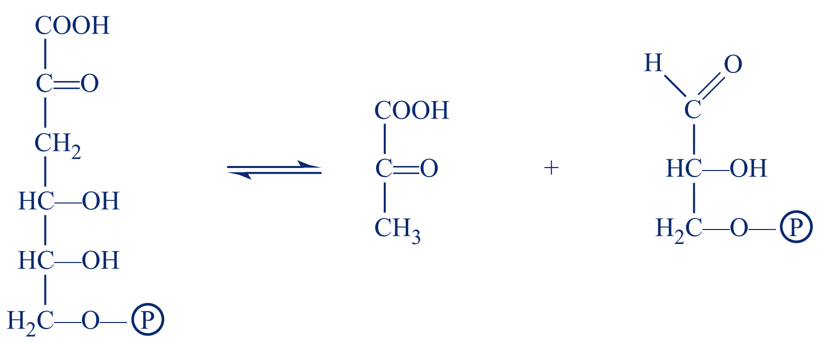
Thus, a wide ensemble of work involving structural and enzymatic studies and the application of genetic techniques permitted to establish in detail the catalytic mechanisms implicated in the activity of aspartate amino-transferase.

12.7. ALDOLASES

Aldolase of muscle was discovered in 1934 by MEYERHOF and LOHMAN who observed the reversible cleavage of hexose diphosphates catalysed by an enzyme which they name zymohexase. Later, MEYERHOF demonstrated that aldolase catalyses the transformation of fructose-1,6-diphosphate in dihydroxyacetone phosphate and 3-phosphoglycerate, according to the reaction:



Aldolase is a highly ubiquitous enzyme. It is found in animal and plant tissues and in most microorganisms. Aldolases can be divided into two distinct classes. Aldolases of class I are found in animals and higher plants; they do not require the presence of a metal ion and form a SCHIFF base intermediate with their substrate. They are inactivated upon reduction by sodium borohydride which blocks the intermediate. In vertebrate tissues, aldolase is tetrameric and there are different isoenzymes. Type A aldolase is present in skeletal muscle; that of type C in cardiac muscle. Liver aldolase is of type B. Aldolases of type A and C have a preferred catalytic specificity for fructose 1,6-diphosphate. Aldolase B plays a major role in the catabolism of fructose-1-phosphate. Aldolases of class II are met in bacteria and molds; their activity requires the presence of a metal ion. Their catalytic mechanism differs from that of aldolases of class I; they do not form a SCHIFF base with their substrate. However a microorganism, *Pseudomonas putida,* possesses an aldolase of class I specific for 2-keto-3-deoxy-6-phosphogluconate or KDPG aldolase. This enzyme catalyses the following reaction:



Only aldolases of class I are considered in this chapter. The primary structures of many of these aldolases are currently known, including the human and the rabbit muscle enzymes, KDPG aldolase of *Pseudomonas putida* as well as the aldolases of some parasites. The first three-dimensional structures resolved were that of KDPG aldolase to 2 Å (MAVRIDIS et al., 1982), that of rabbit muscle to 2.7 Å (SYGUSH et al., 1987), and that of human muscle to 2 Å (GAMBLIN et al., 1991). Today PDB contains more than a hundred aldolase structures of diverse organisms, some resolved to less than 2 Å. The sequences of human pathological mutants were determined. Some specific mutations provoke intolerance to fructose or hemolytic anemia. The mutation sites are generally localised to the vicinity of the active site or at the interfaces between subunits. The analysis of aldolase sequences of parasites responsible for malaria (*Plasmodium falciparum*) and sleeping sickness (*Trypanosoma brucei*) reveals some differences in the C-terminal part of the protein.

12.7.1. STRUCTURAL PROPERTIES

Rabbit muscle and human aldolases are tetramers made of four identical subunits of molecular mass 36 000. Each subunit comprises 365 amino acids. The two enzymes exhibit a high degree of sequence homology and also a great analogy in their three-dimensional structure. KDPG aldolase, contrary to mammalian aldolases, is a trimeric molecule made of identical protomers. Each subunit, shorter than those of mammalian aldolases, is constituted of a polypeptide chain of 225 amino acids. It does not present any sequence homology with fructose 1,6-diphosphate aldolase of muscle.

Structural data obtained by radiocrystallography shows that each protomer, for mammalian aldolases as well as for KDPG aldolase, presents a structure of type α/β similar to that of triose phosphate isomerase and pyruvate kinase. The motif is made of eight $\alpha\beta$ units organised through an 8-fold pseudo symmetry axis (Figs. 12.54 and 12.55 opposite). The α helices are approximately antiparallel to the β segments. As compared to related proteins, muscle aldolases contain four additional helices, three of them being located after segments a, b, and c. KDPG aldolase is shorter, but contains however an additional N-terminal helix like mammalian aldolases.

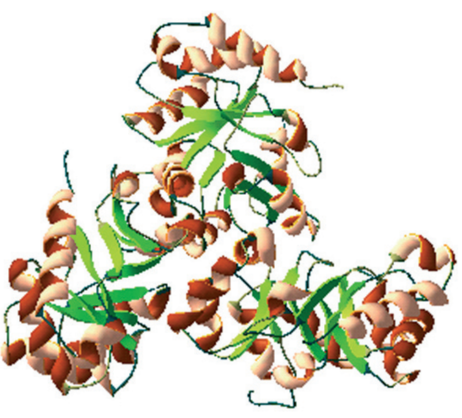


Fig. 12.54 Structure of the KDPG aldolase trimer (PDB: 1EUN)

The β barrel forms the centre of the protein (Fig. 12.55 opposite). It is constituted not only of hydrophobic amino acids but also of polar residues. In the muscle enzymes, residues Asp33, Lys107, Lys146, Glu187 and Lys229 are located at the interior of the cavity. The potentially charged parts of these residues are collinear, the acid and basic residues being alternatively arranged, a situation which can assure charge neutralisation and permits the refolding of these residues towards the interior of the β barrel. In KDPG aldolase, there are twelve pairs of ions important to the stability of the structure. Some charged groups are situated in the central cavity of the molecule (Glu56, Arg141, Asp166), the others at the interface between the subunits. Some of them form interprotomer salt bridges (Arg75-Asp223′; Asp207-Arg210′). These interactions contribute to maintaining the trimer structure.

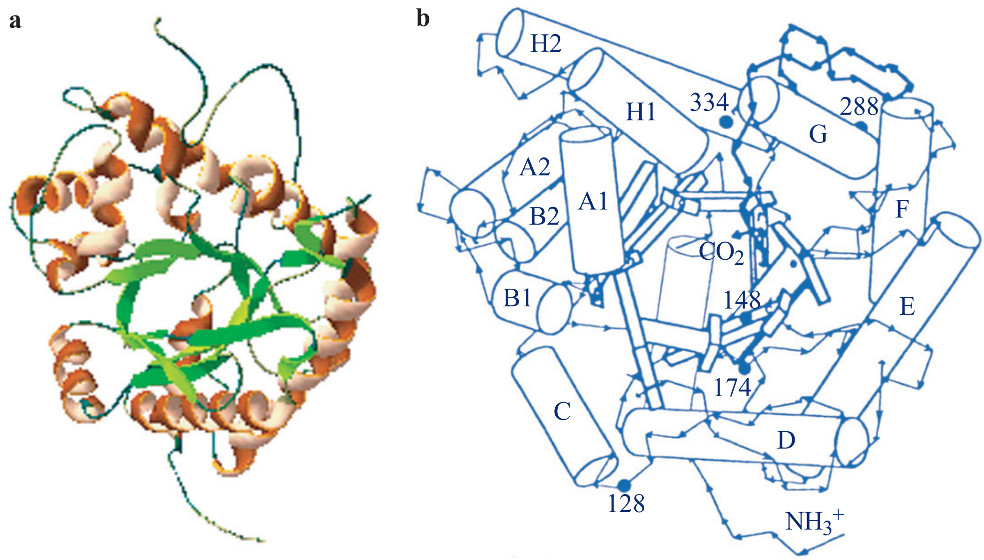


Fig. 12.55 (a) the protomer of rabbit muscle aldolase (PDB: 1ADO) (b) that of the human enzyme

(12.55b – Reprinted from *J. Mol. Biol.*, **219,** GAMBLIN S.J. et al., Activity and specificity of human aldolases, 174. © (1991) with permission from Elsevier)

In rabbit muscle aldolase, the C-terminal extremity presented an apparent disorder in the region 346–354. This region was identified in the human enzyme. The first nine residues (345–353) form a flexible loop at the surface of the enzyme which extends from the C-terminal extremity to helix H2 at the amino-terminal end of helix G. The side chain of serine 355 contracts a hydrogen bond with the nitrogen of the main chain at the extremity of the helix. This interaction orients the last C-terminal part that crosses the αβ motif; the three last residues are localised to the centre of the barrel (see Fig. 12.55b). This region of the protein seems to play a role in the catalytic activity. Indeed, its suppression by proteolysis reduces the enzymatic activity. It was suggested that the C-terminal peptide could modulate the enzymatic activity by assuring the binding and the alignment of the substrate during catalysis. This part of the molecule does not exist in KDPG aldolase. In contrast, it is present in enzymes of the parasites plasmodium and trypanosome.

In the quaternary structure of muscle enzyme, the subunits are arranged in a tetrahedral configuration. The regions of contact are essentially hydrophobic. The largest of the two interfaces is established by the interaction of side chains of helices E and F of adjacent subunits. The arrangement of subunits is such that the two pairs of helices bound by symmetry are approximately antiparallel. Helix E interacts with its equivalent E′ for practically all its length. Helices F and F′ enter into contact only by their C-terminal extremity. The interface with the weakest contact between subunits consists of the loop connecting the βc strand to helices C and D of the two subunits.

12.7.2. THE ACTIVE SITE

The residues implicated in the catalytic reaction are located in a pocket that extends from the surface of the subunit to the centre of the β barrel and includes the region where the charged groups are located. The localisation of the active site differs from that observed in enzymes that present a similar structure, like triose phosphate isomerase where the active site is situated in the region of loops at the COOH extremity of the β strands of the β barrel. In muscle aldolase, lysine 229 which is implicated in the formation of the SCHIFF base with C^2 of the substrate extends from the centre of the β barrel towards the middle of segment βf. Lysine 107 which interacts with phosphate in C^6 is at the surface of the substrate binding pocket. Residues Lys107, Lys146 and Lys229 form an alignment which seems to favor the binding of the triphosphate part of ATP or inositol triphosphate, powerful inhibitors of aldolase.

The inner walls of the active centre pocket are lined with hydrophobic residues coming from the β segments. The closest of the ε-amino group of Lys229 are Ile185, Ile77 and Phe144 near the N-terminal extremity of the β barrel. Residue Arg148, a probable candidate for the binding of phosphate in C^1, is situated at the surface of the substrate binding pocket on the side opposite to Lys107. The distance between the binding sites of the two phosphate groups is 11 Å, which is compatible with the interatomic distances of C^1 and C^6 in the non-cyclic configuration.

From the complete structure of the human enzyme including the C-terminal part of the protein, the group of WATSON carried out molecular modelling of the substrate 1,6-fructose diphosphate at the interior of the active centre. Figure 12.56 opposite illustrates the different enzyme-substrate interactions. The phosphate site in C^1 consists of residues Arg148 and Lys146. The phosphate in C^6 interacts with residues Lys41, Arg42 of helix A1 and Arg303 situated on the loop between the segment βh and helix H1. His361 is in close contact with Arg148 and at a distance permitting the establishment of a hydrogen bond with Tyr363. It could play a role in the positioning of this tyrosine during catalysis. In addition, Tyr363 is at a distance compatible with the establishment of a hydrogen bond with the protonated SCHIFF base made by lysine 229 with the C^3 of fructose 1,6-diphosphate. Site-directed mutagenesis experiments showed that Asp33 situated in the proximity of Lys229 is a residue essential to catalysis. It could act as a base catalyst to extract the proton of the hydroxyl group in C^4 [MORRIS & TOLAN, 1993].

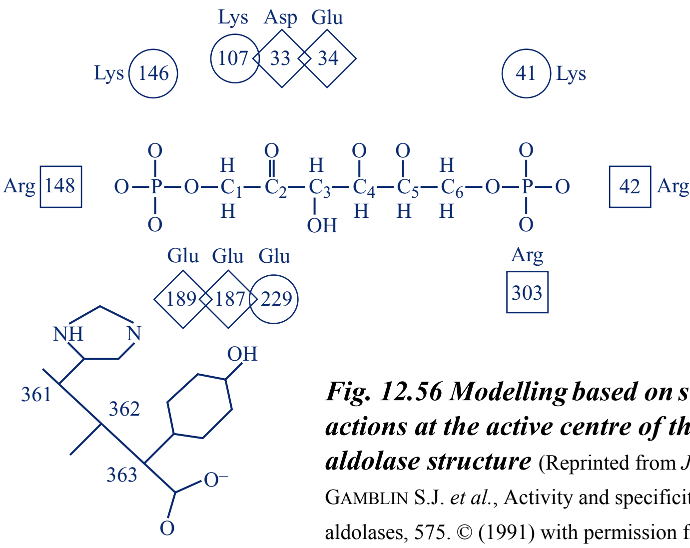


Fig. 12.56 Modelling based on substrate interactions at the active centre of the human aldolase structure (Reprinted from *J. Mol. Biol.*, **219**, GAMBLIN S.J. *et al.*, Activity and specificity of human aldolases, 575. © (1991) with permission from Elsevier)

All the amino acid residues of the active centre including the potentially charged groups of the β barrel were conserved over the course of evolution in all aldolases of class I. By comparison the residues in the vicinity of Lys144 of KDPG aldolase implicated in the formation of the SCHIFF base are indicated in Fig. 12.57.

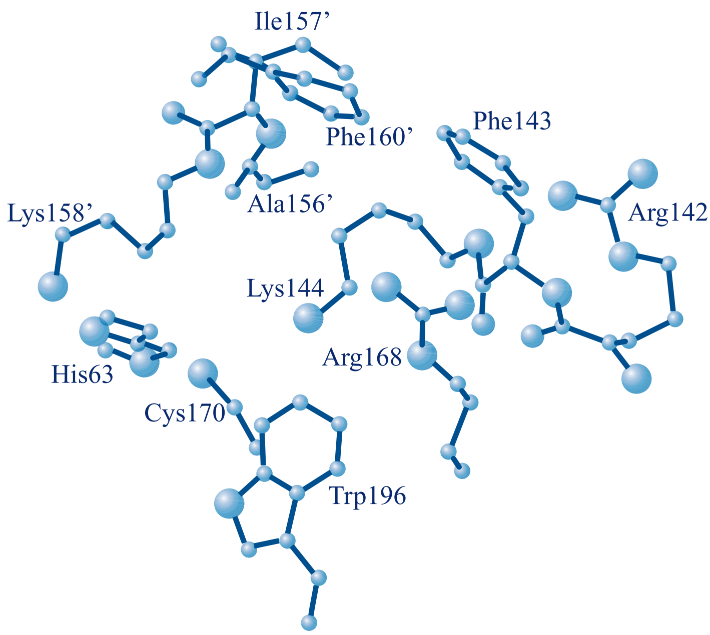


Fig. 12.57 Amino acid residues present in the vicinity of the catalytic lysine (Lys144) of KDPG aldolase (Reprinted from *J. Mol. Biol.*, **162**, MAVRIDIS I.M. et al., Structure of 2-keto-3-deoxy-6-phosphogluconate aldolase at 2.8 Å resolution, 441. © (1982) with permission from Elsevier)

This residue lies in a depression at the surface of the subunit. The entrance to the cavity is bordered by residues Leu145, Pro147, Phe169 and Pro171 of one subunit, and by residues Gly152′, Gly153′ and Ala155′ of another subunit. Like in the muscle enzyme, Lys144 is surrounded by a cluster of positively charged groups (Arg142, Arg168, Lys158′ and His63) and hydrophobic residues.

Although the residues around the active site are well conserved, only 4 of the 22 residues of the C-terminal region are found in the different aldolases, suggesting that this part of the molecule can be implicated in the variation in specificity and activity of these enzymes towards their substrates.

12.7.3. CATALYTIC MECHANISM

Mammalian aldolases are tetramers. But contrary to most oligomeric enzymes, the isolated monomer presents a significant enzymatic activity. The experiments of CHAN (1970, 1972) had shown that the subunits isolated from the rabbit enzyme bound to a solid matrix conserved 27% of the specific activity of the tetrameric aldolase. More recently, BEERMINK and TOLAN (1996) obtained a monomeric aldolase by a double mutation Q125D/E224A which destabilises the tetramer. This monomer presented 72% of the specific activity of the wild type aldolase.

A catalytic mechanism was proposed by HORECKER in 1972 for rabbit muscle fructose 1,6-diphosphate aldolase. The general mechanism involves a SCHIFF base intermediate. The structural data obtained since then, which concerns the reduced SCHIFF base intermediate formed by the transaldolase of *E. coli* and dihydroxyacetone phosphate, allowed to propose a catalytic mechanism. In the complex, the SCHIFF base formed with Lys132 (corresponding to Lys129 of the human enzyme) interacts by hydrogen bond with several residues as shown in Fig. 12.58a opposite. The active site is filled with water molecules; one of them is very close to the ε-amino group of the catalytic lysine and of the dihydroxyacetone part and forms hydrogen bonds with Glu96 and Thr156. This water molecule very likely plays a role in the catalytic mechanism. On the basis of these data a mechanism was proposed by JIA et al. (1997). It is represented in Fig. 12.58b. The first step of the reaction is a nucleophilic attack by the ε-amino group of lysine 132 on the C^2 carbon of carbonyl; the water molecule could act in the proton transfer from Lys132 to Glu96. The negative charge of oxygen in C^2 in the transition state could be stabilised by hydrogen bonds with the water molecule and the side chain of Thr156. The departure of a water molecule from carbaminolamine drives the formation of an imine with the enzyme. The cleavage of the imine and the release of the first product, glyceraldehyde-3-phosphate, are facilitated by the deprotonation of the hydroxyl group in C^4 which would be assisted by Asp17. The resulting SCHIFF base is stabilised by resonance until an aldose is bound to the active site. The carbanion of dihydroxyketone reacts afterwards with the carbonyl of the aldose in the reverse order from what occurs during transaldolisation. The proposed roles for Asp17 and Glu96 are based on structural data.

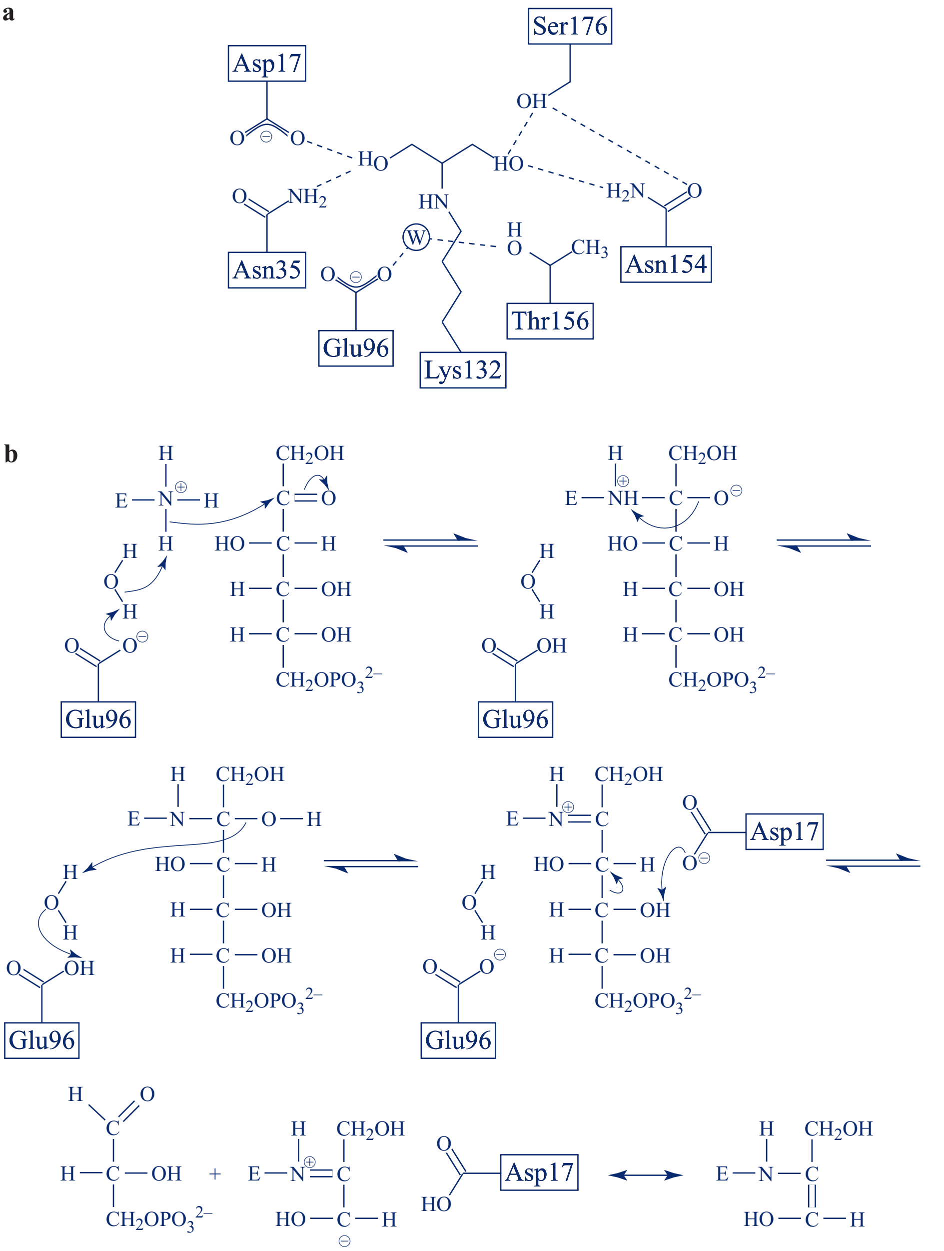


Fig. 12.58 (a) interactions of the reduced SCHIFF base intermediate at the active site of E. coli transaldolase – (b) catalytic mechanism proposed by JIA et al., 1997

(From *Protein Sci.*, **6**, No. 1, JIA J. et al., 1997, 119–124. © (1997 The Protein Society). Reprinted with permission of John Wiley & Sons, Inc.)

Despite the ensemble of structural and functional data obtained on these diverse enzymatic systems taken as examples since they are among the best known, many points still remain to be specified.

12.8. CONCLUSIONS AND PERSPECTIVES

The examples presented illustrate particularly well the contribution of precise structural data in the understanding of catalytic mechanisms implicated in enzymatic reactions. They show also their limits since, on the one hand these data remain static, and on the other hand the X-ray diffraction diagrams do not permit the localisation of protons. However the improvement of techniques that are currently developed in the domain of radiocrystallography, particularly the method of LAUE, should permit the rapid recording of diffraction data and foresee the possibility of carrying out the kinetic studies, meaning to follow the displacement of atoms during an enzymatic reaction. However, the application of this method to the study of enzymatic reactions still poses large problems, principally that of synchronisation of all the reacting molecules. Also it has only been applied to reactions triggered by a "flash" of light. In any case, the use of a plurality of methods is necessary to understand the diverse events which occur in the path of an enzymatic reaction. The use of NMR methods and neutron diffraction remains necessary to localise protons. Kinetic studies in solution and chemical labelling methods or analysis of isotope effects are indispensable. **It is clear that it is only by the use of a variety of methodologies relevant to different disciplines that one succeeds in understanding and mastering the ensemble of factors which cooperate in the catalytic efficiency of enzymes.** Despite all the progress made, there still are numerous enzymes for which the mechanism of action is not known, and even for well described systems, many problems still remain unresolved.

Throughout this chapter, the reader has realised the contribution of chemical labelling and more recently of site-directed mutagenesis in the study of catalytic mechanisms. The possibility of selectively labelling an amino acid residue or replacing this amino acid by another considerably improved our knowledge. But these methods do not only permit knowing, but also intervening and modifying the catalyst. Consequently they are used in new strategies with the goal of modulating or modifying the specificity and catalytic efficiency of an enzyme. The example of mutations carried out on trypsin showed that it was possible, by changing a single amino acid in a well determined position, to increase the specificity of the enzyme for peptide bonds near arginine or lysine. In 1988, WILKS et al. transformed lactate dehydrogenase of *B. stearothermophilus* in malate dehydrogenase by site-directed mutagenesis. Although these two enzymes which function with NAD^+ had practically no sequence homology, they present great structural analogies, in particular in the binding domain of NAD^+ binding domain. Lactate dehydrogenase reversibly catalyses the reduction of pyruvate in lactate by NADH:

$$\underset{CH_3—CO—COO^-}{\text{pyruvate}} + NADH + H^+ \longleftrightarrow \underset{CH_3—CHOH—COO^-}{\text{lactate}} + NAD^+$$

and malate dehydrogenase, that of oxaloacetate in malate by NADH :

$$\underset{^-OOC—CO—CH_2—COO^-}{\text{oxaloacetate}} + NADH + H^+ \longleftrightarrow \underset{^-OOC—CH_2—CHOH—COO^-}{\text{malate}} + NAD^+$$

Taking into account the position of atoms in the ternary complexes at the active centre of the two enzymes, mutants Asp197 ⟶ Asn, Thr246 ⟶ Gly and Gln102 ⟶ Arg of lactate dehydrogenase were constructed and analysed. Figure 12.59 shows the geometry of the active site in the ternary complex LDH-NADH-oxamate. Table 12.8 gives the values of the kinetic parameters of mutants and the wild type enzyme. The last mutation Gln102 ⟶ Arg gives rise to an enzyme very specific for malate with an excellent catalytic efficiency. The maximal rate obtained is twice larger than that of malate dehydrogenase of the same organism. Thus, engineered malate dehydrogenase, more active than the natural enzyme, was obtained by remodelling of another enzyme.

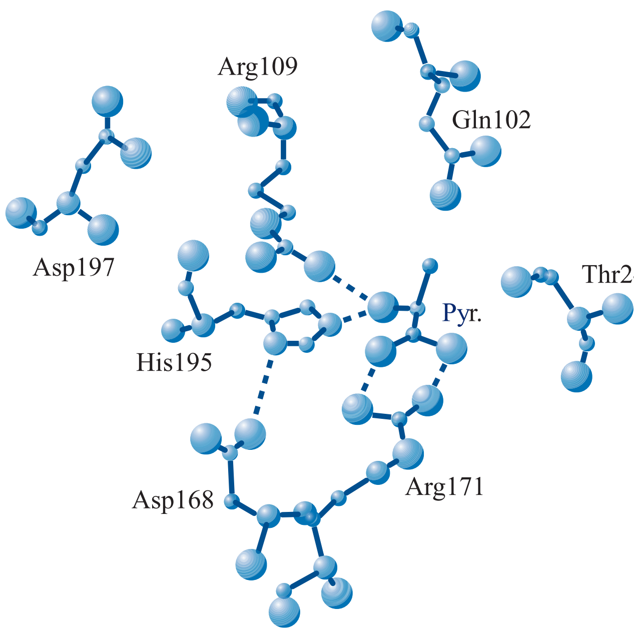


Fig. 12.59 Geometry of the active centre of lactate dehydrogenase (LDH) in the ternary complex LDH-NADH-oxamate (From *Science*, **242**, WILKS H.M. et al., A specific, highly active malate dehydrogenase by redesign of a lactate dehydrogenase Framework, 1541. © (1988) reprinted with permission from American Association for the Advancement of Sciences)

Table 12.8 Kinetic parameters of lactate dehydrogenase and its mutants

Enzyme	*Substrate*	$K_m(mM)$	$k_{cat}(s^{-1})$	$k_{cat}/K_m(M^{-1}.s^{-1})$
	Pyruvate			
Wild type		0.060	250	4.2×10^6
Asp197 ⟶ Asn		0.66	90	1.3×10^5
Thr246 ⟶ Gly		13	16.0	1.3×10^3
Gln102 ⟶ Arg		1.8	0.9	5.0×10^2
	Oxaloacetate			
Wild type		1.5	6.0	4.0×10^3
Asp197 ⟶ Asn		0.15	0.50	3.0×10^3
Thr246 ⟶ Gly		0.20	0.94	4.7×10^3
Gln102 ⟶ Arg		0.06	250	4.2×10^6

We have cited this example which was among the first well documented ones. Since then, other works have resulted in the modification of specificity obtained by mutation of several amino acids at the binding site of substrates after consideration of the structure.

Taking into account the successes of these approaches, research in molecular enzymology is oriented more and more towards the concept of new catalysts. Several strategies are currently being developed. The first uses the very strict specificity of abzymes (see Chap. 11). Results show that these catalysts are generally very selective. However, their catalytic efficiency, although higher than that of chemical catalysts, remains relatively weak compared to that of enzymes.

Another strategy, oriented by molecular modelling, utilises the genetic or even chemical approach (hemisynthesis) for the remodelling of enzymes or the conception of new enzymes. The goal of devising new proteins (*protein design)* is to obtain *de novo* polypeptides having a predetermined activity. The conception step consists of choosing a reaction mechanism, chemical groups and a spatial arrangement susceptible to giving rise to the researched activity. Then, one must imagine a polypep tide scaffold that will support these active groups in the correct orientation. Finally, there remains to be determined an amino acid sequence which will be conveniently folded to give a stable three-dimensional structure. Several approaches are possible. One consists of taking as a model a protein of known structure having properties close to those of the researched catalyst and locally remodelling it to create the active centre desired. This is the **local conception**. The other approach, or the **global conception,** conceives, starting from zero, a structure having one of the classical folds. In this case, the remodelling can be made by analogy using the data bank of known crystallographic structures. The last step consists of synthesising the corresponding gene coding for the sequence thus conceived.

Finally, the development of genetic engineering techniques now permit obtaining very numerous variants of a protein by methods of **in vitro *evolution*** associated with screens permitting the selection of the desired protein.

Despite the empirical success which was obtained so far, it is important to underline that significant progress will only be made when the folding code of proteins will be deciphered. Indeed, it is easy today to modify a protein sequence, even to conceive new ones, but to obtain an active enzyme, it is still necessary that this protein can acquire an adequate three-dimensional structure. The knowledge of mechanisms which drive a polypeptide chain to an operational spatial structure is an indispensable preliminary to the rational conception of new enzymes and represents an important challenge for biotechnology.

Whatever it may be, it is clear that molecular enzymology is long from falling obsolete; due to the utilisation of new tools, it has acquired a new youth and currently represents a domain in full development by both its fundamental aspects and its applications. With the success obtained in the sequencing of genomes, a large field of study is dedicated today to proteins and their structural and functional properties.

BIBLIOGRAPHY

BOOKS

CHRISTEN P. & METZLER D.E. eds –1985– *Transaminases,* John Wiley & Sons, New York.

FERSHT A. –1985– *Enzyme, structure and mechanisms*, 2nd ed., Freeman and Co, New York.

PELMONT J. –1996– *Enzymes : catalyseurs du vivant*, Collection Grenoble Sciences, EDP Sciences, Paris.

GENERAL REVIEWS

BACKER E.N. & DRENTH J. –1984– The thiol proteases: structure and mechanisms, in *Biological macromolecules and assemblies*, Vol. 3, F.A. JURNAK & A.M. MCPHERSON eds, John wiley & Sons, New York, 313–368.

BLOW D.M. –1971– The structure of chymotrypsin, in *The Enzymes*, 3rd ed., Vol. III, P.D. BOYER ed., Acad. Press, New York, 185–212.

BRAUNSTEIN A.E. –1985– Introduction: an historical survey in transamination and transaminases, in *Transaminases*, P. CHRISTEN & D.E. METZLER eds, John Wiley & Sons, New York, 1–35.

DALZIEL K. –1975– Kinetics and mechanism of nicotinamide nucleotide linked dehydrogenases, in *The Enzymes*, 3rd ed., Vol. III, P.D. BOYER ed., Acad. Press, New York, 1–60.

EKLUND H. & BRÁNDÉN C.I. –1984– Alcohol dehydrogenase, in *Biological macromolecules and assemblies*, Vol. 3, F.A. JURNAK & A.M. MCPHERSON eds, John wiley & Sons, New York, 73–142.

FRUTON J.S. –1976– The mechanism of the catalytic action of pepsin and related acid proteinases, in *Adv. Enzymol.* **44**, 1–36.

HARTLEY B.S. & SHOTTON D.D. –1971– Pancreatic elastase, in *The Enzymes*, 3rd ed., Vol. III, P.D. BOYER ed., Acad. Press, New York, 323–373.

HARTSUCK J.A. & LIPSCOMB W.N. –1971– Carboxypeptidase, in *The Enzymes*, 3rd ed., Vol. III, P.D. BOYER ed., Acad. Press, New York, 1–79.

HORECKER B.L., TSOLAS O. & LAI C.Y. –1972– Aldolases, in *The Enzymes*, 3rd ed., Vol. VII, P.D. BOYER ed., Acad. Press, New York, 213–258.

IMOTO T., JOHNSON L.N., NORTH A.C.T., PHILLIPS D.C. & RUPLEY J.A. –1972– Vertebrate lysozymes, in *The Enzymes*, 3rd ed., Vol. VII, P.D. BOYER ed., Acad. Press, New York, 666–868.

IVANOV V.I. & YA KARPEISKY M. –1969– Dynamic three-dimensional model for enzymic transamination, in *Adv. Enzymol.* **32**, 21–53.

JAMES M.N.G. & SIELECKI A.R. –1987– Aspartic proteinases and their catalytic pathway, in *Biological macromolecules and assemblies*, Vol. 3, F.A. JURNAK & A. MCPHERSON eds, John wiley & Sons, New York, 413–482.

JANSONIUS J.N. & VINCENT M.G. –1987– Structural basis for catalysis by aspartate aminotransferase, in *Biological macromolecules and assemblies*, Vol. 3, F.A. JURNAK & A. MCPHERSON eds, John wiley & Sons, New York, 187–285.

KRAUT J. –1971a– Chymotrypsinogen: X-ray structure, in *The Enzymes*, 3rd ed., Vol. III, P.D. BOYER ed., Acad. Press, New York, 165–183.

KRAUT J. –1971b– Subtilisin: X-ray structure, in *The Enzymes*, 3rd ed., Vol. III, P.D. BOYER ed., Acad. Press, New York, 547–560.

KRAUT J. –1977– Serine proteases: structure and mechanism of catalysis, in *Annu. Rev. Biochem.* **46**, 331–358.

MARTINEZ-CARRION M., HUBERT E., IRIARTE A., MATTINGLY J.R. & ZITO S.W. –1985– Mechanism of aminotransferase action, in *Transaminases*, P. CHRISTEN & D.E. METZLER eds, John Wiley & Sons, New York, 307–316.

METZLER D.E. –1979– Tautomerism in pyridoxal phosphate and in enzyme catalysis, in *Adv. Enzymol.* **50**, 1–40.

SCOPES R.K. –1973– 3-Phosphoglycerate kinase, in *The Enzymes*, 3rd ed., Vol. VIII, P.D. BOYER ed., Acad. Press, New York, 335–351.

SPECIALISED ARTICLES

ALBER T., BANNER P.W., BLOOMER A.C., PETSKO G.A., PHILLIPS D.C., RIVER P.S. & WILSON I.A. –1981– *Phil. Trans. Roy. Soc. London* **B293**, 159.

ANTONOV V.K., GINODMAN L.M., KAPITANNIKOV Y.V., BARSHEVSKAYA T.N., GUROVA A.G. & RUMSH L.D. –1978– *FEBS Lett.* **88**, 87.

ATLAS D. –1975– *J. Mol. Biol.* **93**, 39.

BANKS R.D., BLAKE C.C.F., HASER P.R., RICE D.W., HARDY G.W., MERRETT M. & PHILLIPS A.W. –1979– *Nature* **279**, 773.

BANNER D.W., BLOOMER A.C., PETSKO G.A., PHILLIPS D.C., PEGSON C.L., WILSON I.A., CONNAN P.H., FURTH A.J., MILMAN J.D., OFFORD R.D., PRIDDLE J.D. & WALEY S.G. –1975– *Nature* **255**, 609, London.

BASH P.A., FIELD M.J., DAVENPORT R.C., PETSKO G.A., RINGE D. & KARPLUS M. –1991– *Biochemistry* **30**, 5826.

BEERNINK P.T. & TOLAN D.R. –1996– *Proc. Natl Acad. Sci. USA* **93**, 5374.

BERGER A. & SCHECHTER I. –1970– *Phil. Trans. R. Soc. London B Biol. Ser.* **259**, 249.

BERNHARD S. ET AL. –1970– *J. Mol. Biol.* **49**, 85.

BERNSTEIN B.E. & HOL W.G.J. –1998– *Biochemistry* **37**, 4429.

BERNSTEIN B.E., MICHELS P.A. & HOL W.G.J. –1997– *Nature* **385**, 275.

BIRKOFT J.J. & BLOW D.M. –1972– *J. Mol. Biol.* **68**, 187.

BLAKE C.C.F., KOENIG P.F., MAIR G.A., NORTH A.C.T., PHILLIPS D.C. & SARMA V.P. –1965– *Nature* **206**, 757.

BLAKE C.C., JOHNSON L.N., MAIR G.A., NORTH A.C., PHILLIPS D.C. & SARMA D.R. –1967– *Proc. R. Soc. London B Biol. Sci.* **169**, 378.

BLAKE C.C.F.&.RICE D.W –1981– *Phil. Trans. Roy. Soc. London* **B293**, 93.

BLOW D.M., JANIN J. & SWEET R.M. –1974– *Nature* **249**, 54.

BODE W., FELDHAMMER H. & HUBER R. –1976– *J. Mol. Biol.* **106**, 325.

CANFIELD R.E. –1963– *J. Biol. Chem.* **238**, 2691 and 2698.

CAPEILLIAIRE-BLANDIN C., BRAY R.C., IWATSUBO M. & LABEYRIE F. –1975– *Eur. J. Biochem.* **54**, 549.

CHAN W.W. –1970– *Biochim. Biophys. Res. Com.* **41**, 1198.

CHAN W.W. –1972– *Arch. Biochem. Biophys.* **14**, 136.

CHAPUS C., KERFELEC B., FOGLIZZO E. & BONICEL J. –1987– *Eur. J. Biochem.* **166**, 379.

COHEN G.H., SILVERSTON E.W., MATTHEWS B.W., BRAXTON H. & DAVIES D.R. –1969– *J. Mol. Biol.* **44**, 129.

COLL M., GUASH A., AVILES F.X. & HUBER R. –1991– *Embo. J.* **10**, 1.

COOPER A.J. & MEISTER A. –1989– *Biochimie* **71**, 387.

CUI Q. & KARPLUS M. –2001– *J. Am. Chem. Soc.* **123**, 2284.

DAVENPORT R.C., BASH P.A., SEATON B.A., KARPLUS M, PETSKO G.A. & RINGE D. –1991– *Biochemistry* **30**, 5821 and 5826.

DAVIES G.J., GAMBLIN S.J., LITTLECHILD J. & WATSON H.C. –1993– *Proteins* **15**, 283.

DELBAERE L.J.T., HUTCHEON W.L.B., JAMES M.N.G. & THIESSEN W.E. –1975– *Nature* **257**, 758.

DRENTH J., HOL W.G., JANSONIUS J.N. & KOEKOEK R. –1972– *Eur. J. Biochem.* **26**, 177.

DUNN B.M. & FINK A.–1984– *Biochemistry* **23**, 5241.

EKLUND H., PLAPP B., SAMAMA J.M. & BRÁNDÉN C.I. –1982– *J. Biol. Chem.* **257**, 14349.

FELDHAMMER H. & BODE W. –1975– *J. Mol. Biol.* **98**, 683.

FINK A. & MEEHAN P. –1979– *Proc. Natl. Acad. Sci. USA* **76**, 1566.

FLEMING A. –1922– *Proc. Roy. Soc. London* **B93**, 306.

FRUTON J.S.–1970– *Adv. Enzymol. Rel. Areas Mol. Biol.* **33**, 401.

GALDEN A., AULD D.S. & VALLEE B.L. –1986– *Biochemistry* **25**, 646.

GAMBLIN S.J., DAVIES G.J., GRIMES J.M., JACKSON R.M., LITTLECHILD J.A. & WATSON H.C. –1991– *J. Mol. Biol.* **219**, 573.

GARDELL S.J., CRAIK C.S., HILVERT D., URDEA M.S. & RUTTER W.J. –1985– *Nature* **317**, 551.

GERTLER A., WALSH K.A. & NEURATH H. –1974– *Biochemistry* **13**, 1302.

GHELIS C. & YON J.M. –1979– *C.R. Acad. Sci. Paris* **D289**, 197.

GOMIS-RÜTH F.X., GOMEZ-ORTIZ M., VENDRELL J., VENTURA S., BODE W., HUBER R. & AVILES F.X. –1997– *J. Mol. Biol.* **269**, 861.

GUASH A., COLL M., AVILES F.X. & HUBER R –1992– *J. Mol. Biol.* **224**, 141.

GUIARD B. –1985– *Embo. J.* **4**, 3265.

GUIARD B., GROUDINSKI O. & LEDERER F. –1974– *Proc. Natl Acad. Sci. USA* **84**, 2629.

GUILBERT C., PERAHIA D. & MOUAWAD L. –1995– *Comput. Phys. Commun.* **91**, 263.

HARLOS K., VAS M. & BLAKE C.C. –1992– *Proteins* **12**, 133.

HARUYAMA K., NABAI T., MIYAHARA I., HIROTSU K., MIZYGUSHI H. & KAGAMIYAMA H. –2001– *Biochemistry* **40**, 4632.

HENDERSON R. –1970– *J. Mol. Biol.* **54**, 341.

HUBER R, KUKLA D., BODE D., SCHWAGER P., BARTELS P., DEISENHOFER J. & STEIGEMANN W. –1974– *J. Mol. Biol.* **89**, 73.

INOUE K., KARAMITSU S., OKAMOTO A., HIROTSU K., HIGUSHI T. & KAGAMIYAMA H. –1991– *Biochemistry* **30**, 7796.

JENKINS W.J., YPHANTIS D.A. & SINGER I.W. –1959– *J. Biol. Chem.* **234**, 51.

JIA J., SCHÖRKEN U., LINDVIST Y., SPENGER G.A. & SCHNEIDER G. –1997– Crystal structure of the reduced Schiff-base intermediate complex of transaldolase B from *Escherichia coli*: mechanistic implications for class I aldolases, in *Protein Sci.* **6**, 119.

JOLLES J., JAUREGUI-ADELL J., BERNIER I. & JOLLES P. –1963–*Biochim. Biophys. Acta* **78**, 668.

KAMPHUIS I.G., DRENTH J. & BAKES E.N. –1985– *J. Mol. Biol.* **182**, 317.

KAMPHUIS I.G., KALK K.H., SWART M.B. & DRENTH J. –1985– *J. Mol. Biol.* **179**, 233.

KARPEISKY M.Y., KHOMUTOV R.M., SEVERIN E.S. & BRENSOV Y.N. –1963– in *Chemical and biological aspects of pyridoxal catalysis*, A.E. BRAUNSTEIN, E.S. SEVERIN & Y.M. TORCHINSKY eds, Pergamon Press, Oxford, 323.

KELLY J.A., SIELECKI A.R., SYKES B.D., JAMES M.N.G. & PHILLIPS D.C. –1979– *Nature* **282**, 875.

KERR M.A., WALSH K.A. & NEURATH H. –1976– *Biochemistry* **15**, 5566.

KIRBY A.J. –2001– *Nat. Struct. Biol.* **8**, 737.

KLINMAN J.P. –1981– *C.R.C. Crit. Rev. Biochem.* **10**, 39.

KNOWLES J.R. –1970– *Phil. Trans. Roy. Soc. London* **B257**, 135.

KOMIVES E.A., CHANG L.C., LOLIS E., TILTON R.F., PETSKO G.A. & KNOWLES J.R. –1991– *Biochemistry* **30**, 3011.

KOSHLAND D.E. –1953– *Biol. Rev.* **28**, 416.

KOSSIAKOFF A.A. & SPENCER S.A. –1981– *Biochemistry* **20**, 6462.

KUO L.C., FUKUYAMA J.M. & MAKINEN M.W. –1983– *J. Mol. Biol.* **163**, 63.

LAI C.Y. –1975– *Arch. Biochim. Biophys.* **166**, 358.

LASHTCHENKO P. –1909– *Z. Hug. Infektionkrankh.* **64**, 419.

LEDERER F. –1987– in *Flavins and flavoproteins*, Walter de Gruyter and Co, Berlin, 513.

LEDERER F. & MATTHEWS F.S. –1987– in *Flavins and flavoproteins*, Walter DE GRUYTER and Co, Berlin, 133.

LEVITT M. –1974– in *Peptides, polypeptides and proteins*, E. BLOUT, F. BOVEY, M. GOODMAN & N. LOTAN eds, John Wiley pub., New York, 99.

LIPSCOMB W.N. –1970– *Accounts for Chem. Res.* **3**, 81.

LIPSCOMB W.N. –1980– *Proc. Natl Acad. Sci. USA* **77**, 3875.

LONSDALE-ECCLES J.D., NEURATH H. & WALSH K. –1978– *Biochemistry* **17**, 2805.

LONSDALE-ECCLES J.D., KERR M.A., NEURATH H. & WALSH K. –1979– *FEBS Lett.* **100**, 157.

MAKINEN M.W., YAMAMURA K. & KAYSER E.T. –1976– *Proc. Natl Acad. Sci. USA* **68**, 3882.

MATTHEWS D.A., ALDEN R.A., BIRKTOFT J.J. & KRAUT J. –1975– *J. Biol. Chem.* **250**, 7120.

MATTHEWS F.S. & ZIA Z.X. –1987– in *Flavins and flavoproteins*, Walter DE GRUYTER and Co, Berlin, 123.

MAVRIDIS I.M., HATADA M.H., TULINSKI A. & LEBRODA L. –1982– *J. Mol. Biol.* **162**, 419.

MAY A., VAS M., HARLOS K. & BLAKE C.C.F. –1996– *Proteins: Struct. Funct. Genet.* **24**, 292.

MCLACHLAN A.D. & SHOTTON D.H. –1971– *Nature* **229**, 202.

MCPHILLIPS T.M., HSU B.T., SHERMAN M.A., MAS M.T. & REES D.C. –1996– *Biochemistry* **35**, 4118.

MIZYGUCHI H., HAYASHI H., OKADA K., MIYAHARA I., HIROTSU K. & KAGAMIYAMA H. –2001– *Biochemistry* **40**, 353.

MORRIS A.J. & TOLAN D.R. –1993– *J. Biol. Chem.* **268**, 1095.

MOUAWAD L., DESMADRIL M., PERAHIA D., YON J.M. & BROCHON J.C. –1990– *Biopolymers* **30**, 1151.

NAVIA M.A., FITZGERALD P.M.D., MCKEEVER B.M., LEU C.T., HEIMBACH J.C., HERBER W.K., SIGAL I.S., DARKE P.L. & SPRINGER J.P. –1989– *Nature* **337**, 615.

NEURATH H., WALSH K.A. & WINTER W.P. –1967– *Science* **158**, 1638.

NICKBARG E.B. & KNOWLES J.R. –1988– *Biochemistry* **27**, 5939.

NICKBARG E.B., DAVENPORT R.C., PETSKO G.A. & KNOWLES J.R. –1988– *Biochemistry* **27**, 5948.

OUCHNINIKOV Y.A., EGOROV C.A., ALDANOVA N.A., FEIGINA M.Y., LIPKIN V.M., ABDULAEV N.G., GRISHIN E.V., KISELEV A.P., MODYANOV N.M., BRAUENSTEIN A.E., POLYANOVSKI O.L. & NISONOV V.V. –1973– *FEBS Lett.* **29**, 31.

PARSON S.M. & RAFTERY M.A. –1972– *Biochemistry* **11**, 1623.

PHILLIPS D.C., FLETTERICK R. & RUTTER W.J. –1990– *J. Biol. Chem.* **265**, 20692.

ROSSMANN M.G., MORAS D. & OLSEN K.W. –1974– *Nature* **250**, 194.

REES D.C., LEWIS M. & LIPSCOMB W.N. –1983– *J. Mol. Biol.* **168**, 367.

SCHINDLER M., ASSAF Y., SHARON N. & CHIPMAN D.M. –1977– *Biochemistry* **16**, 423.

SCHMIDT M.F. & HERRIOT J.R. –1976– *J. Mol. Biol.* **103**, 175.

SEYDOUX F., NÉMÉTHY G. & YON J.M. –1969– *Biochim. Biophys. Acta* **171**, 145.

SIELECKI A.R., FEDOROV A., BOODHOO A., ANDREVA N.S. & JAMES N.G. –1990– *J. Mol. Biol.* **214**, 143.

SIELECKI A.R., FUJINAGA M., READ R.J. & JAMES N.G. –1991– *J. Mol. Biol.* **219**, 671.

SHOTTON D.M. & WATSON H.C. –1970– *Nature* **225**, 811.

SNELL E.E. –1945– *J. Am. Chem. Soc.* **67**, 194.

STROUD R.M., KAY L.M. & DICKERSON R.E. –1974– *J. Mol. Biol.* **83**, 185.

SUZUKI N. & WOOD W.A. –1980– *J. Biol. Chem.* **255**, 3427.

SYGUSH J., BRANDY J., & ALLAIRE M. –1987– *Proc. Natl. Acad. Sci. USA* **84**, 7846.

TONEY M.D. & KIRSCH J.F. –1991– *Biochemistry* **30**, 7456.

TULINSKY A. & BLEVINS R.A. –1987– *J. Biol. Chem.* **262**, 7737.

UMEYAMA H., NAGAWA S. & KUDO T. –1981– *J. Mol. Biol.* **150**, 409.

URBAN P. & LEDERER F. –1985– *J. Biol. Chem.* **260**, 11115.

VALLEE B.L. & NEURATH H. –1955– *J. Biol. Chem.* **217**, 253.

VOCADIO D.J., DAVIES G.J., LAINE R. & WITHERS S.G. –2001– *Nature* **412**, 835.

WALTER J. & BODE W.–1983– *Hoppe Seylers Z. Physiol. Chem.* **364**, 949.

WATSON H.C., WALKER N.P.C., SHAW P.J., BRYANT T.N., WENDELL P.L., FOTHERGILL L.A., PERKINS R.E., CONROY S.C., DOBSON M.J., TUITE M.F., KINGSMAN A.J. & KINGSMAN S.M. –1982– *Embo. J.* **1**, 1635.

WEBB M.R. & TRENTHAM D.R. –1980– *J. Biol. Chem.* **255**, 1775.

WILKS H.M., HART K.W., FEENEY R., DUNN C.R., MUIRHEAD H., CHIA W.N., BARSTOW D.A., ATKINSON T., CLARKE A.R. & HOLBROOK J.J. –1988– *Science* **242**, 1541.

WRIGHT C.S., ALDEN R.A. & KRAUT J. –1970– *Nature* **221**, 235.

XIA Z.X., SHAMALA H., BETHGE P.H., LIM L.W., BELLAMY H.D., XUONG N.H., LEDERER F. & MATHEWS F.S. –1987– *Proc. Natl Acad. Sci. USA* **84**, 2629.