9 – TOPOLOGY OF THE ACTIVE CENTRE OF ENZYMES

For most enzymes, there is a disproportion between the size of the enzymatic molecule and the size of the substrate. Such an observation brought about very early the notion of the **active centre,** the fact that a very small proportion of the surface of the enzyme enters into contact with the substrate. However, the concept of the active centre remained for a long time rather poorly defined. For KOSHLAND, the active centre was constituted by all the enzyme atoms that are in contact fixed in the boundaries of the VAN DER WAALS radius with the substrate atoms, meaning those that stay at a minimal distance such that the electron clouds are not perturbed. KOSHLAND distinguishes thus the contact residues and the auxiliary residues, the latter being able to play a role in enzymatic activity (Fig. 9.1).

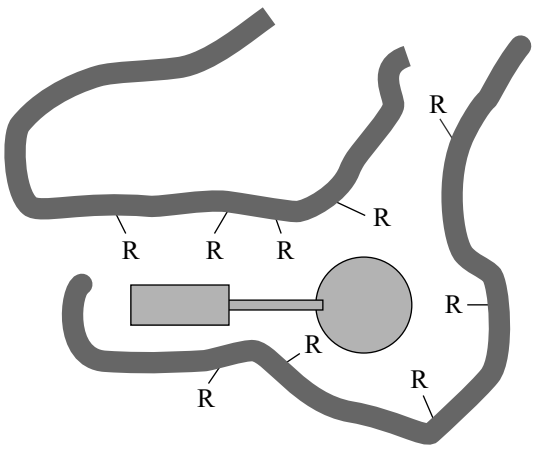


Fig. 9.1 Schematic representation of the active centre of an enzyme with the different residues R that interact with the substrate

Meanwhile, such an idea remains insufficient. It is purely spatial and does not take into account the functional aspect. Indeed, the residues of the enzyme that come into contact with the substrate can have very diverse roles, either interfering with the binding of the substrate or participating directly in the catalysis. Some do not play any role; they are present at the active centre by consequence of the polypeptide sequence and have direct influence neither on the enzyme-substrate association nor on the catalysis. Their modification or replacement by other amino acids in crossing from one species to another, for example, or replacement by site-directed mutagenesis, does not change their activity. On the contrary, certain amino acids which are not bound by VAN DER WAALS forces to the substrate can play an essen-

J. Yon-Kahn, G. Hervé, *Molecular and Cellular Enzymology*,
DOI 10.1007/978-3-642-01228-0_10, © Springer-Verlag Berlin Heidelberg 2010

tial role, favoring the binding of the substrate either by creating a microenvironment favorable to the functioning of the active centre or by maintaining its functional conformation.

Upon enzyme-substrate association, diverse events that bring about the displacement of certain enzyme atom groups can occur. For example, in carboxypeptidase, tyrosine 248 undergoes a rotation of 12 Å upon substrate binding; in the free enzyme this residue is situated near the molecule exterior far from the other residues of the active centre. For other enzymes, including hexokinase and phosphoglycerate kinase, there occurs a bringing together of domains upon substrate binding. It is important to specify these notions somewhat further. One must distinguish:

- the catalytic site that concerns the groups directly involved in the chemical act;
- the binding site that corresponds with enzyme groups establishing non-covalent interactions with the substrate;
- the conformational site that corresponds with specific residues important for maintaining an active conformation of the enzyme.

The active centre is constituted by the ensemble; it includes altogether the catalytic site, the binding site and the conformational site.

However, it is not always easy to know with precision the topology of the active centre. Several methods of investigation have been used for this aim. The kinetic approach consists of analysing variations in kinetic parameters as a function of pH; it has historically represented a first approach to the problem. The dependence of kinetic constants on pH is attributed to ionisation of enzymatic groups important for the activity. Nevertheless, as we will see, this method lets stand the most often an ambiguity concerning even the nature of these groups. It remains important for specifying their role but permits only to hint at their nature and not to specify it. The chemical methods to specifically label protein groups constitute precious instruments for determining the nature of essential groups. Genetic methods, in particular site-directed mutagenesis permitting the selective replacement of one amino acid by another, offer a very powerful tool complementary to chemical methods. The most precise topological information without any doubt is supplied by radiocristallographic studies and nuclear magnetic resonance. Structural determinations by X-ray diffraction require obtaining crystals of complexes that form an enzyme with substrate analogs. But even when the structure of the complexes has been solved, the functional studies are necessary to define the role of different residues at the active centre. Nuclear magnetic resonance studies have the advantage of being applicable to proteins in solution, but they are limited by the strong concentrations they require. Therefore, these different methods do not exclude each other but are complementary. These diverse approaches and the information that they bring are analysed successively.

9.1. KINETIC APPROACH – ANALYSIS OF pH PROFILES

This approach consists of analysing the influence of pH on enzymatic reactions. For a long time one had remarked that enzymes are active in a more or less limited pH range. In most cases, one defines even rather precisely an optimal pH (Fig. 9.2). The effect of pH on enzymatic reactions, like all pH effects, issues changes in the ionisation state of the system components as a function of pH. The existence of an optimum can result in different effects. Some of them are of a conformational order and are not connected to the ionisation of the active centre groups. One situation fairly often encountered results from the decrease in enzyme stability at extreme pH where a denaturation occurs that can be irreversible. This is relatively common with globular proteins. In other cases, reversible denaturations due to a variation of some charges of the molecule happen at extreme pH. In short, the more specific conformational effects are at times responsible for the variations in activity as a function of pH. In some cases, the ionisation of a single group or of a pair of groups plays a key role in the active conformation of an enzyme. Some examples will be given later.

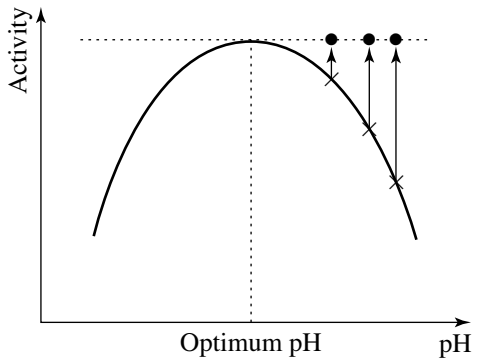


Fig. 9.2 Variation in enzymatic activity as a function of pH
The crosses represent the pH values of enzyme incubation; the points *indicate the activity after the return to optimal pH*

The action of pH on the simple enzyme-substrate association does not result in a variation in V_m, but modifies the parameter K_m. It can be assessed more directly by using a competitive inhibitor. The activity curve as a function of pH results also, at least in part, from the ionisation of catalytic groups. In these conditions, the effect of pH affects both parameters V_m and K_m.

Finally, if the substrate itself is susceptible to being ionised, and if its ionisation state moderates its binding, the enzyme affinity for the substrate will vary as a function of pH. It is convenient to distinguish experimentally these diverse effects. Some of them are easily analysed and must be approached at the beginning; first of all the conformational state of the protein, then the ionisation states of the substrate that are easy to distinguish separately. Moreover, it is important to analyse the role of ionisable enzyme groups in the association of the substrate and in the catalytic act itself.

9.1.1. EFFECT OF pH ON THE CONFORMATIONAL STATE OF THE PROTEIN

9.1.1.1. IRREVERSIBLE DENATURATION

Once the curve of global activity variation as a function of pH is established, it is easy to determine if the decrease in activity, either in alkaline or in acidic conditions, or in the two conditions, is due to an irreversible denaturation of the enzyme. The very simple experiment consists of pre-incubating the enzyme at these extreme pH values for a determined time, then placing it at the optimal pH to measure its activity. If, under these conditions, the activity stays constant whatever the pre-incubation pH and equal to the previously measured activity at optimal pH, an irreversible denaturation cannot be the cause of the decrease in activity at extreme pH. It is the same if the activity returns to a value superior to the value observed at the pH of the pre-incubation although inferior to the previously observed value at the optimal pH; one can be situated then in a case of slowly reversible denaturation (see Fig. 9.2).

9.1.1.2. REVERSIBLE DENATURATION

If the decrease in enzymatic activity at extreme pH is due to a reversible denaturation, the return of the activity to its initial value measured at optimal pH can be more or less slow. In many cases this phenomena is not immediate and it is possible to follow the kinetics. As renaturation is on the contrary a rapid phenomenon, the existence of a denaturation at extreme pH risks happening unnoticed. **In order to be exempt from this phenomenon, it is important to measure, as a function of pH, a physical parameter revealing the conformational state of the protein.**

9.1.1.3. IONISATION OF GROUPS THAT INTERFERE SPECIFICALLY WITH THE ACTIVE CONFORMATION OF THE ENZYME

As previously mentioned, in order to obtain a clear response on the role of ionisable groups that interfere with the activity of an enzyme, the conformational studies as a function of pH must be carried out in parallel with kinetic studies. Indeed, the active conformation of certain enzymes is maintained by a specific interaction between two charged groups. The changes in pH involving charge modification of one or another of these groups delete the interaction. It follows that either the substrate binding or the catalytic constant depends on the ionisation of these groups, although they interfere only in the conformation of the active site of the enzyme, and not in the catalytic act.

For several serine proteases, it has been shown that an ionic interaction secures the optimal conformation of the active centre. For chymotrypsin, this interaction is formed between the ⟨-amino group of isoleucine 16 at the N-terminal position of the enzyme in the NH_3^+ state and the carboxylate of aspartate 194 in the proximity of serine 195 of the catalytic site. For trypsin and elastase that are proteases of the same family, the same type of interaction exists. Each of these groups undergoes an

ionisation that induces a structural transition on both sides of the optimal pH. The amino group is ionised with an apparent pK of 10 for trypsin (CHEVALLIER & YON, 1966), of 8.9 for chymotrypsin (OPPENHEIMER et al., 1966); the carboxylate is ionised with a pK of 4 for trypsin and 3 for chymotrypsin. In alkaline medium, these enzymes exhibit therefore an equilibrium between a neutral form comprising the existence of a salt bridge that maintains the active conformation of the catalytic site and an alkaline form in which the salt bridge is broken. In acidic medium they are in equilibrium between the neutral form and the acidic form in which the interaction is equally broken (Fig. 9.3).

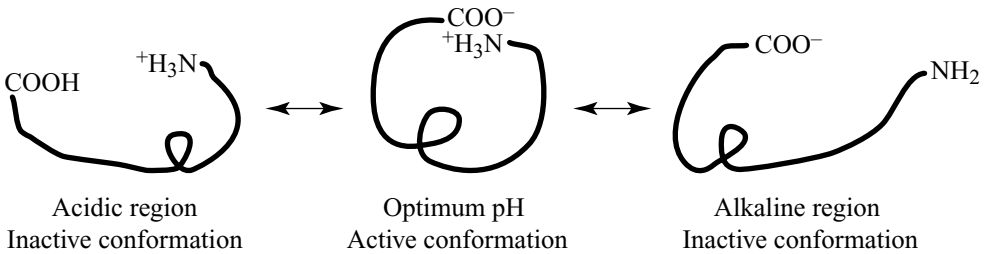


Fig. 9.3 Conformational equilibrium of serine proteases as a function of pH

The variations of diverse physical parameters such as rotatory power, fluorescence, and spectrum difference, as a function of pH reveal these two pK (Fig. 9.4). The results of crystallography studies have confirmed the existence of salt bridge in serine proteases. Figure 8.5 of the preceding chapter shows, according to the atomic coordinates, the salt bridge and the catalytic groups of chymotrypsin. The opening of the salt bridge drives the loss of the optimal conformation of the active centre; it results in important variations in the functional properties (see Part IV).

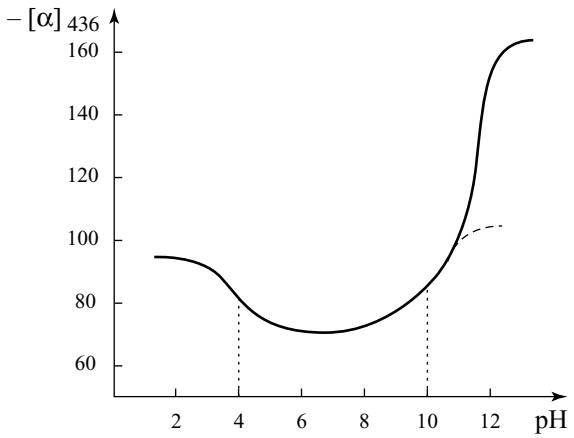


Fig. 9.4 The effect of pH on the conformational equilibrium of trypsin evaluated by variations in rotatory power at 436 nm
In the alkaline range, a more global denaturation prevails at the transition due to the deprotonation of the amino group

9.1.2. SUBSTRATE IONISATION STATES

The substrate itself can exist under different ionic forms in the enzyme activity range. Generally substrates are simpler molecules than enzymes and possess only a

limited number of ionisable groups. A direct titration permits the determination of the ionisation pK values independently of the presence of the enzyme.

◤ Let us consider the substrate:

$$SH_2^{++} \underset{}{\overset{K_{s1}}{\rightleftharpoons}} SH^{+} \overset{K_{s2}}{\rightleftharpoons} S$$

with its ionisation states. Several situations can arise. In one, the three forms are equally recognised by the enzyme without the charge interfering. In this case, the ionisation states of the substrate do not interfere with the activity curve as a function of pH. In another, a form is recognised preferentially by the enzyme, for example SH^+. The substrate fraction having the optimal affinity for the enzyme is in the form:

$$SH^+ = \frac{S_t}{1 + \frac{H^+}{K_{s1}} + \frac{K_{s2}}{H^+}}$$

The activity curve as a function of pH varies in exhibiting two values of pK, pK_{s1} and pK_{s2} which, if they have been identified by a direct titration, could be distinguished from pK values of enzyme group ionisation. When the affinity of the enzyme for the substrate depends on the ionisation state of the latter, this signifies that an electrostatic interaction interferes with the enzyme-substrate association and reflects itself in the variations in K_s as a function of pH. ◢

9.1.3. EFFECT OF pH ON ENZYME-SUBSTRATE ASSOCIATIONS

Certain enzyme-substrate associations involve some electrostatic interactions between one or several charges of the enzyme and the opposite charges of the substrate. Thus, the specific substrates of trypsin possess a positive charge on the ε-amino group of lysine or the guanidine group of arginine that interacts with the charge of aspartate 177 situated at the bottom of a hydrophobic pocket. The protonation of aspartate in acidic medium brings about a decrease in the enzyme affinity for its substrate.

◤ If one can know the parameter K_s (for example in the case of limiting acylation substrates for which $K_m = K_s$), the variations in K_s as a function of pH in the ionisation regions of this carboxyl group are represented by a sigmoidal curve as indicated in Fig. 9.5 opposite. In acidic medium the curve presents a plateau giving the value of $K_{s,max}$, that is in the absence of electrostatic interaction; in alkaline medium when the group is deprotonated one obtains $K_{s,min}$. The enzyme affinity for its substrate is maximal as it involves, in addition to some other interactions, the interaction between the charges of the enzyme and the substrate. Starting with these two values, it is easy to evaluate the electrostatic contribution to the energy of interaction. Indeed:

$$\Delta G_1 = RT \ln K_{s,max}$$

$$\Delta G_2 = RT \ln K_{s,min}$$

$$\Delta\Delta G = \Delta G_2 - \Delta G_1 = RT \ln K_{s,min}/K_{s,max}$$

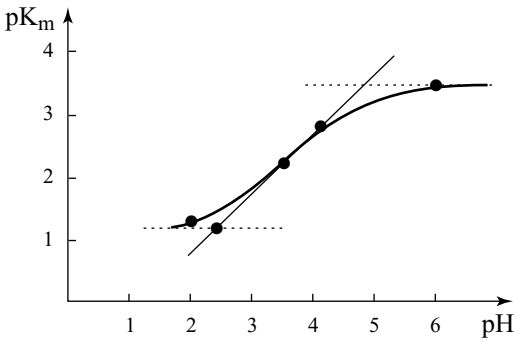


Fig. 9.5 Variations in K_m as a function of pH for a substrate of trypsin in which acylation is the limiting step in the reaction ($K_m = K_s$); DIXON plot

This value permits, in using COULOMB'S law, to calculate the approximate distance between the charges:

$$\Delta\Delta G = \frac{305 Z_1 Z_2}{d_e(6d_e - 7)} \times \frac{e^{-\kappa(r_0 - d_e)}}{1 + \kappa r_0}$$

In this expression Z_1 and Z_2 are the groups' charges that enter into the electrostatic interaction; d_e is the distance between these charges. The factor $(6d_e - 7)$ is a value approaching the dielectric constant. The exponential factor is a correction factor that takes into account the ionic atmosphere surrounding these groups, in which κ is the DEBYE and HÜCKEL constant equal to 0.7×10^7 for an ionic strength of 0.1; r_0 is the smallest ion approach distance estimated by a first approximation of 10.2 Å. This value includes the radius of the enzyme active group that is 1.7 Å; the radius of an ion included in the ionic atmosphere is 1.3 Å and the diameter of two molecules of water is 7.2 Å. ◢

It is generally preferable to resort to competitive inhibitor analogs of the substrate and to analyse the variations in pK_i as a function of pH (K_i, the inhibition constant being a true dissociation constant). In the case of trypsin inhibition by two specific competitive inhibitors, benzoyl arginine and tosyl arginine, the respective distances of 6.9 and 6.5 Å were determined in this manner (Fig. 9.6). For the inhibition of trypsin by benzamidine, the distance between the charges of the enzyme and of the substrate was evaluated to be 4.3 Å.

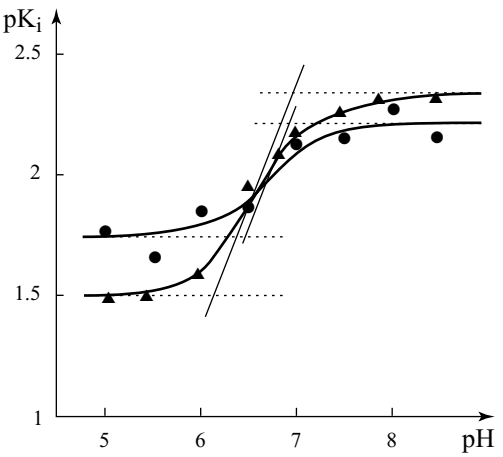


***Fig. 9.6 Variations in pK_i as a function of pH upon inhibition of trypsin hydrolysis of acetyl phenylalanine methyl ester by benzoyl argininamide* (triangles) *and tosyl argininamide* (circles)**

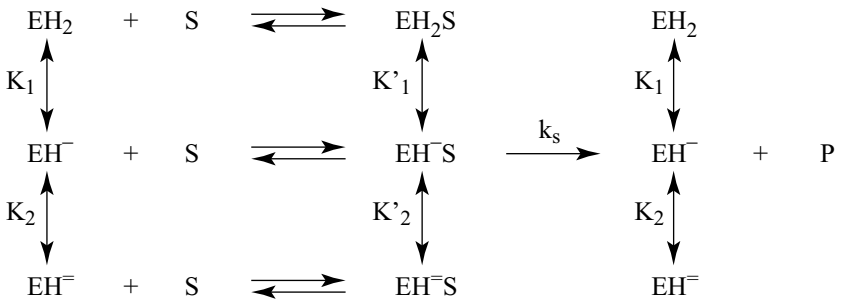
(Reprinted from *Biochim. Biophys. Acta*, **122**, BECHET J.J. *et al.*, Etude de l'inhibition de l'activité estérasique de la trypsine par des dérivés acides et amides des esters spécifiques, 101. © (1966) with permission from Elsevier)

9.1.4. EFFECTS OF pH ON THE IONISATION OF CATALYTIC GROUPS

Enzymes, like all proteins, contain a large quantity of ionisable groups. There exists a whole series of ionisation states of the enzyme and the distribution of the entire enzyme in its different ionic forms depends on the pH and the constants of ionisation of diverse groups. However, one had remarked for a while that the enzymatic activity is generally restrained over a fairly narrow range of pH, and one had deduced from this that only one of the ionic forms of the enzyme (or of a part of the enzyme) is catalytically active. Despite the size of the protein molecule, the number of ionisable groups interfering in the catalysis is generally restrained. In this paragraph, we will analyse the incidence of ionisation of these groups on the behaviour of kinetic parameters as a function of pH. Different authors including LAIDLER, ALBERTY, MASSEY, and DIXON have given the complete scheme of enzymatic reactions as a function of pH. To approach the methodology of such a study, we will consider first the simple case of an enzymatic reaction with a single substrate and involving a single enzyme-substrate intermediate of the MICHAELIS complex type; we will consider next the case of two intermediates, then that of several intermediates.

9.1.4.1. ENZYMATIC REACTIONS INVOLVING A SINGLE INTERMEDIATE

The general scheme of the reaction as a function of pH is written:



In such a scheme, only the ionic form EH^- of the enzyme and the form EH^-S of the complex are catalytically active. These forms are in equilibrium on the acid side with the forms EH_2 and EH_2S, respectively, and on the alkaline side with $E^=$ and $E^=S$.

The different ionic forms of the free enzyme are susceptible to associating with the substrate. K_1, K_2, K'_1, and K'_2 are the ionisation constants of the free enzyme and of the MICHAELIS complex. The reaction rate as a function of pH is in the general form:

$$v = \frac{k_s(pH)es}{K_m(pH)+s}$$

with:

$$k_s(pH) = k_{s,opt}\frac{1}{1+\frac{H^+}{K'_1}+\frac{K'_2}{H^+}}$$

and:
$$K_m(pH) = K_{m,opt} \frac{1 + \frac{H^+}{K_1} + \frac{K_2}{H^+}}{1 + \frac{H^+}{K'_1} + \frac{K'_2}{H^+}}$$

$k_{s,opt}$ and $K_{m,opt}$ are the kinetic parameters at optimal pH when the free enzyme is entirely under the form EH^- and the complex under the form EH^-S.

Study of variations of k_s as a function of pH

These expressions are useful under their logarithmic form. One has:

$$\log k_{s(pH)} = \log k_{s,opt} - \log\left[1 + \frac{H^+}{K'_1} + \frac{K'_2}{H^+}\right]$$

We generally obtain the profiles of pH of the form indicated in Fig. 9.7a.

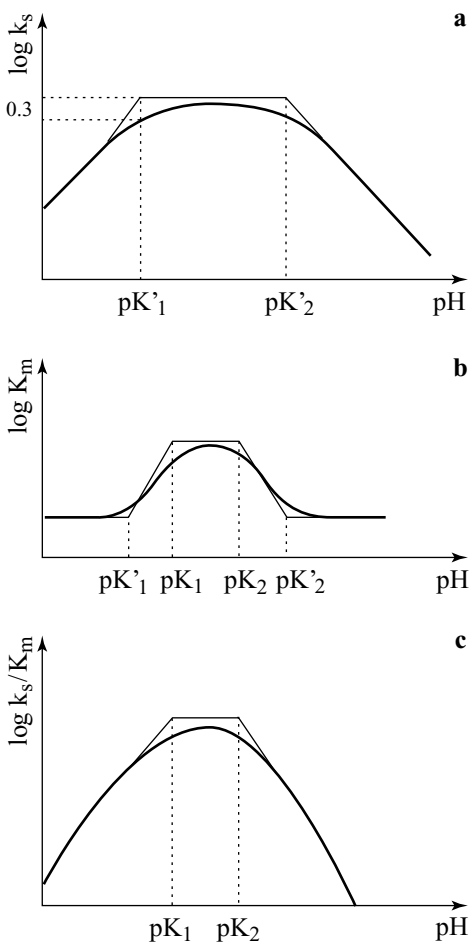


Fig. 9.7 DIXON plots
(a) *log k_s as a function of pH*
(b) *log K_m as a function of pH*
(c) *log k_s/K_m as a function of pH*

This curve permits obtaining the pK of the groups that are ionised in the MICHAELIS complex, meaning pK'_1 and pK'_2. It summons a certain number of observations known under the name of the DIXON rules.

The curve is formed by portions of straight lines joined by short curves, the curvature happens in the vicinity of the pK of ionisation. The slope of the lines is represented by integers, 1, 0, and –1 on the graph in Fig. 9.7a. The value of the slope corresponds with the number of protons implicated in the conversion of one ionic form to another. In the present case, a single proton is implicated in the passage of one ionisation state to another; the slope is equal to 1 or –1 following that the group is essential under deprotonated or protonated form. This results obviously from the preceding relationship. Indeed, at low pH $(H^+) > K'_1$ and *a fortiori* $(H^+) > K'_2$, so the equation simplifies into:

$$\log k_{s(pH)} = \log k_{s,opt} + pH - pK'_1$$

The line representing $\log k_{s(pH)}$ as a function of pH therefore has a slope of 1. At high pH $(H^+) < K'_2 < K'_1$ and one has:

$$\log k_{s(pH)} = \log k_{s,opt} - pH + pK'_2$$

The value of the pK of ionisation is given by the intersections of portions of lines. However, if the two values of pK'_1 and pK'_2 are not sufficiently distant, the plateau corresponding with the optimum is not determined with sufficient accuracy. Another imprecision occasionally comes from experimental difficulties in acidic or alkaline medium and it is not always easy to follow an insufficiency of experimental points and trace correctly the corresponding lines. In this case, the determination of the pK values meanwhile is possible. Indeed, in the vicinity of pK, the graph no longer follows a linear variation, but it curves with a downward concavity. This curve is situated at a vertical distance of 0.3 pH units below the intersection point of the lines if the ionisation of a single group interferes (see Fig. 9.7). Indeed, in the vicinity of pK'_1 for example, one has:

$$\log k_{s(pH)} = \log k_{s,opt} - \log\left(1 + \frac{H^+}{K'_1}\right)$$

At pK, $H^+ = K'_1$ and the equation becomes:

$$\begin{aligned} \log k_{s(pH)} &= \log k_{s,opt} - \log 2 \\ &= \log k_{s,opt} - 0.3 \end{aligned}$$

Variations in K_m as a function of pH

The expression of K_m is more complex as it yields the ionisation pK of catalytic groups of free enzyme and of the MICHAELIS complex. Its analysis follows nevertheless the same principles. The expression of $K_{m(pH)}$ in its logarithmic form is the following:

$$pK_{m(pH)} = pK_{m,opt} - \log\left(1 + \frac{H^+}{K_1} + \frac{K_2}{H^+}\right) + \log\left(1 + \frac{H^+}{K'_1} + \frac{K'_2}{H^+}\right)$$

It is interpreted by the curve given in Fig. 9.7b yielding four ionisation pKs. The rules of determining pK stated previously stay valid for the analysis of variations in pK_m as a function of pH. One supplementary rule occurs that derives from the equation above. In the vicinity of the pK of free enzyme, the graph presents a downward concavity; in the vicinity of pK of the complex, the concavity is upward. The analysis of variations in pK_m as a function of pH stays delicate. Indeed, the difference in the pK values between the free enzyme and the complex are often very weak and remain difficult to determine graphically. The pK′ values are given by the representation of log k_s as a function of pH.

There is a third type of graphical representation giving only the pK of the free enzyme. It results from studying variations in log (k_s/K_m) as a function of pH. Indeed, from preceding expressions of k_s and of K_m, it follows:

$$\left(\frac{k_s}{K_m}\right)_{pH} = \left(\frac{k_s}{K_m}\right)_{opt} \times \frac{1}{1 + \frac{H^+}{K_1} + \frac{K_2}{H^+}}$$

The corresponding curve is given in Fig. 9.7c.

Having determined the values of different ionisation pK that affect the enzymatic reaction, it is only possible to make some hypotheses on the nature of the groups implicated in the catalytic activity. One knows indeed the ionisation pK of principal groups titrated in the proteins. Table 9.1 recalls these data.

Table 9.1 Average values of pK and enthalpy of ionisation of protein side chain residues

Ionisable group	*pK*	*ΔH (kcal . mol^{-1})*
Carboxylate (Asp, Glu)	4–4.5	1.5
Imidazole (His)	6–6.5	6.9
α-amino	6–8	10–13
ε-amino (Lys)	9.2	10–12
Phenol (Tyr)	10	
Sulfhydryl (Cys)	8.5	
Guanidine (Arg)	12	12–13
Alcohol (Ser, Thr)	> 14	

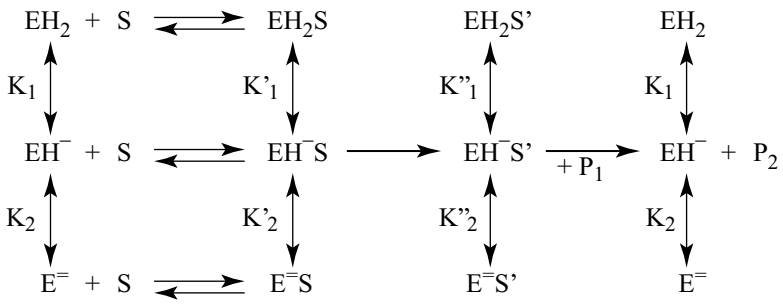
However, the kinetic study as a function of pH can only give an ambiguous response concerning the nature of the group; indeed, the pK of the ionisable groups in the proteins can be shifted depending on their environment. Hence for lactoglobulin, two carboxylate groups ionise with apparent pK neighboring neutrality; their

titration accompanies a conformational protein transition. For serine proteases, the apparent pK of ionisable groups participating in the establishment of the salt bridge that maintains the structure of the active site is displaced, as reported in Section 9.1.1.3 of this chapter. **The kinetic method gives mainly information on the role of the group.** It must always be accompanied by direct chemical determin ation of the implicated group or groups. It meanwhile permits making hypotheses that will direct the research of a reagent susceptible to labelling the essential residue. Moreover it requires, as we have already underlined, the use of physical methods to determine if the ionisable group is not implicated in a conformational event. The potentialities offered by site-directed mutagenesis techniques also permit such a determination.

9.1.4.2. ENZYMATIC REACTIONS IMPLICATING TWO INTERMEDIATE COMPLEXES

Like what was analysed in Part II, enzymatic reactions proceed via the formation of several intermediate complexes. The analysis of kinetic parameter variations as a function of pH then becomes more complex. In the classic case of two reaction intermediates and a single catalytic process ("bottle-neck mechanism"), like for example reactions catalysed by serine proteases that proceed via the formation of an acyl-enzyme intermediate, the analysis remains relatively simple.

◤ The reaction scheme as a function of pH can be written:



The constant K_1 and K_2 relate with ionisations of groups in the free enzyme, K'_1, K'_2 with those in the MICHAELIS complex and K''_1, K''_2 with those in the acyl-enzyme. In this scheme a single catalytic process is permitted, implicating the forms EH^-, EH^-S and EH^-S' that are the only catalytically active ionic forms; k_2 and k_3 are the constants of acylation and of deacylation, respectively.

Like what was presented in Part II (Chap. 5), the kinetic parameters of this scheme have complex values.

$$k_{cat} = \frac{k_2 k_3}{k_2 + k_3}$$

and:

$$K_{m,app} = K_m \frac{k_3}{k_2 + k_3}$$

The diverse constants depend on pH. Rather simple equations permit the expression of each value as a function of pH. Therefore k_2 and k_3 depend on the pK of the MICHAELIS constant and of the acyl-enzyme, respectively.

$$k_{2(pH)} = k_{2,opt} \frac{1}{1 + \frac{H^+}{K'_1} + \frac{K'_2}{H^+}}$$

$$k_{3(pH)} = k_{3,opt} \frac{1}{1 + \frac{H^+}{K''_1} + \frac{K''_2}{H^+}}$$

$$K_{m(pH)} = K_{m,opt} \frac{1 + \frac{H^+}{K_1} + \frac{K_2}{H^+}}{1 + \frac{H^+}{K'_1} + \frac{K'_2}{H^+}}$$

One knows that, whatever the pH is, there exists a simple relationship between the constants:

$$\frac{k_{cat}}{K_{m,app}} = \frac{k_2}{K_m}$$

It follows that the variations in this ratio as a function of pH permit the determination of the pK of groups that are ionised in the free enzyme:

$$\left(\frac{k_{cat}}{K_{m,app}}\right)_{pH} = \left(\frac{k_2}{K_m}\right)_{opt} \times \frac{1}{1 + \frac{H^+}{K_1} + \frac{K_2}{H^+}}$$

Whatever the rate-limiting step of the reaction is, it is therefore always possible to obtain the ionisation pK of groups in the free enzyme. In addition, with substrates for which k_2 is limiting, one can determine the ionisation pK for the MICHAELIS complex. For substrates in which k_3 is limiting, one can obtain the ionisation pK of catalytic groups in the acyl-enzyme. However, it is essential to ascertain that these stages remain limiting in all pH ranges studied.

9.1.4.3. ENZYMATIC REACTIONS INVOLVING SEVERAL INTERMEDIATE COMPLEXES

The preceding results can be generalised to n intermediates. When the kinetic scheme is such that there only exists a single catalytic process, the study of variations in the ratio k_{cat}/K_m as a function of pH permits the determination of ionisation pK of catalytic groups of the free enzyme. When certain steps are limiting, the variations in the rate constant of the limiting step as a function of pH permit the determination of ionisation pK of the complex upstream of the limiting step. If there is more than one catalytic process, i.e. if a catalytic group can be active under different ionisation states, the problem becomes much more complex. However, this situation is very unlikely.

9.2. CHEMICAL APPROACH TO STUDYING THE ACTIVE CENTRE OF ENZYMES

The kinetic and conformational approaches previously described drive to specify the role of ionisable groups, but concerning the nature of these groups the answer still remains ambiguous. In certain cases it can give some indications, yet it is necessary to provide a direct experimental proof.

9.2.1. PRINCIPLE OF CHEMICAL LABELLING

The side chains of a protein susceptible to interfering in catalysis and to be modified by a reagent are generally nucleophilic or electrophilic groups. It concerns essentially the imidazole group of histidine, the amine groups of lysine and arginine, the phenol group of tyrosine, the thiol group of cysteine, the sulfur of methionine, the carboxyl group of glutamate and of aspartate and, when it occurs at the active centre of enzymes, the functional alcohol group of serine and of threonine.

The choice of a reagent depends on the amino acid residue that one wants to modify. It can be directed by the results of kinetic studies performed at different pH values. It is important to know **that there do not exist specific reagents** that permit the modification of one type of residue at the exclusion of others, **but only selective reagents** bringing about the preferential modification of a category of groups in well-determined conditions. Besides, the protein milieu is heterogeneous with polar parts (essentially external) and apolar parts, creating locally a microenvironment that can considerably modify the reactivity of a group in the side chain. The surface of the protein, generally polar, interacting with the aqueous solvent, possesses very variable charge densities from one place to another. The protein interior is essentially constituted by non-polar residues forming a hydrophobic "core" with a weak dielectric constant. There is also inside a protein a large variety of local dielectric constants, electrostatic fields that make the same functional group able to present very different reactivities according to its localisation. All the effects of the microenvironment are of great importance for both the efficiency of enzymatic catalysis and the mechanism of action of the chemical reagents.

The particular microenvironment that exists at the active centre of an enzyme is responsible for the **hyperreactivity** of certain groups. For example, for serine proteases the catalytic serine is abnormally reactive, whereas the other serine residues do not present any reactivity. This hyperreactivity is created by the microenvironment particular to the active centre. Meanwhile, some other residues like those implicated in catalysis could present hyperreactivities with certain chemical reagents. Therefore, among the 19 tyrosines of papain, only tyrosine 123 reacts rapidly with diisopropyl phosphofluoridate. However, the catalytic SH group in the same enzyme and in other thiol proteases is not phosphorylated, but it is hyper-reactive

with cyanate. Glutamate 35 of lysozyme to which is attributed a catalytic role is the only carboxyl group of the enzyme that is resistant to chemical modifications.

The reactivity, the nucleophilic power, of a functional group is determined by the electron density of the implicated atoms, polarisability and steric factors that limit the access of the reagent. Hyperreactivity is therefore not linked to an exaltation of nucleophilic power itself, but to external factors of the functional group. All the external factors included in the effects of the microenvironment created by the protein matrix affect not only the reactivity of functional groups, but also those of the reagent. The principal effects are analysed successively.

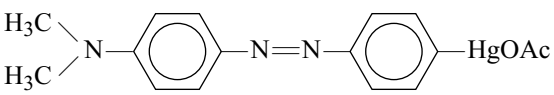
9.2.1.1. EFFECTS OF THE MICROENVIRONMENT ON THE PROTEIN FUNCTIONAL GROUPS

The reactivity of a functional group corresponds with its nucleophilic power that is generally but not always reliant on its basic properties, meaning its ionisation pK. There is a relationship between the nucleophilic power of a group and its pK of ionisation, the BRØNSTED relationship. Yet in a protein one observes important variations in the apparent pK of ionisable groups; these values are sometimes very different from those that are in the denatured protein or when these groups are carried by simple substituted amino acids. It is a question of "abnormal" pK. It results that the reactivity of a protein side chain toward a chemical reagent cannot be anticipated starting with data obtained from models.

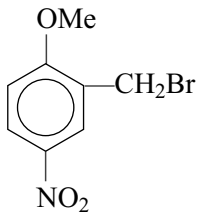
In some cases, the reactivity of a group was compared in the protein and in the model compound. Thus histidine at the catalytic site of ribonuclease is alkylated 500 times more quickly by bromoacetate at pH 5 than is N-acetylhistidine. The catalytic thiol group of papain is 3 000 times more reactive than cysteine is with cyanate and 15 000 times more with bromoacetate.

Polarity of the microenvironment

The polarity of the local environment affects reactivity in a way dependent on the reaction. The effect of an electric field in the vicinity of a charged group can modify the ionisation pK and reactivity of this group either in one direction or in the opposite direction according to the charges present. Diverse approaches have been proposed to evaluate the microenvironment polarity, in particular the use of reporter groups, using either ionisation pK titration data or spectral data of these molecules introduced as extrinsic probes of polarity. Hence the binding of azo-mercurials to the single SH of bovine serum albumin lowers the pK of the reporter group anilino nitrogen to 1.9 in the protein, whereas when the compound is coupled to cysteine it is 3.3.



It is possible to separate polarity effects from electrostatic effects by using certain reagents insensitive to pH effects but having chromatic properties that depend on the solvent polarity. In this way 2-methoxy-5-nitrobenzyl has been used by KOSHLAND in the case of chymotrypsin:



As a consequence of the presence of the aromatic nucleus, this compound binds to the binding site of chymotrypsin substrates and alkylates methionine 192. Its spectrum of absorption is very sensitive to the polarity of the medium; it presents a maximum at 288 nm in hexane and at 317 nm in dimethylsulfoxide. Molecules involving a spin label like a nitroxide radical have been also introduced as reporter groups. Studies in electron paramagnetic resonance then give information on the microenvironment and its variations over the course of an enzymatic reaction, or upon ligand binding for example.

Formation of hydrogen bonds

Some ionisable residues in proteins can form hydrogen bonds, which drive the displacements of ionisation pK and generally a decrease in reactivity. Different hydrogen bonds can be established between residues and side chains (Fig. 9.8).

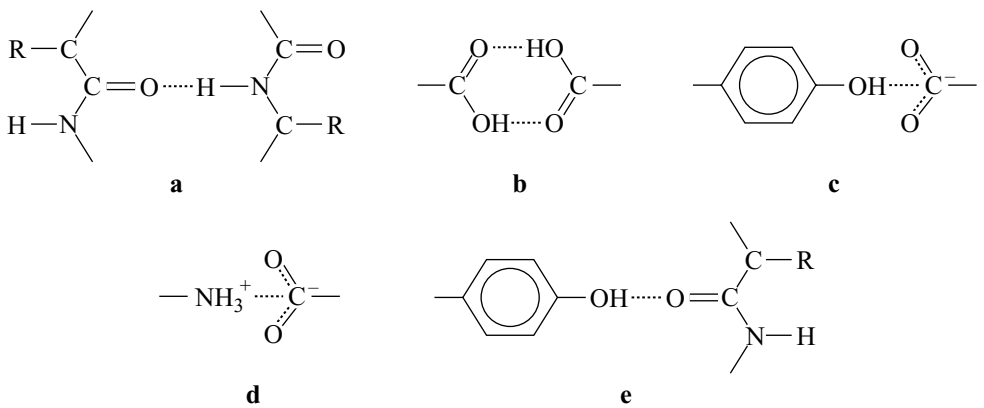


Fig. 9.8 Principal hydrogen bonds in proteins
*(**a**) interamide – (**b**) between two carboxylates – (**c**) between a charged residue and a protonated residue – (**d**) between two charged residues (ionic interaction) – (**e**) between the oxygen of the peptide bond carbonyl and a protonated residue*

Following that the implicated group acts as the donor or acceptor, its ionisation pK increases or decreases when it acquires a hydrogen bond. The equilibria are:

$$DH,A \overset{K_h}{\rightleftharpoons} DH\cdots A$$

$$DH \overset{K_1}{\rightleftharpoons} D^- + H^+$$

$$A + H^+ \overset{K_2}{\rightleftharpoons} AH^+$$

K_h, K_1, and K_2 being the constants of hydrogen bond formation and ionisation of groups of the donor DH and acceptor A, respectively. The apparent ionisation constant of the donor group is given by the equation:

$$K_{obs} = \frac{(D,A)(H^+)}{(DH,A)+(DH...A)}$$

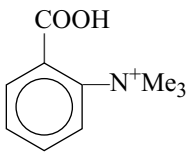
$$K_{obs} = \frac{K_1}{1+K_h}$$

(D,A) being the pair of donor and acceptor groups under their non-protonated form. The ionisation constant of the acceptor group is:

$$K_{obs} = \frac{K_2}{1+K_h}$$

Electrostatic effects

The presence of positive charges in the vicinity of an ionisable group disfavors its protonation, either by stabilisation of an ion pair (if the group is negatively charged), or by avoiding the formation of a new positive charge. Thus the pK of the carboxyl group in the following compound:



is 1.37 whereas it is 3.45 in the meta and para isomers.

The electrostatic effects can also create attractive or repulsive effects on the reagent.

Steric effects

Among the numerous effects due to the protein matrix, one notes sometimes a total steric protection rendering the group completely inaccessible whatever its size; only denaturation of the protein permits the group to react with a given reagent. Most often a partial protection is observed; the accessibility depends then on the size of the reagent. In some cases, the protein matrix limits freedom of rotation of side chains and therefore decreases their reactivity by preventing them from adopting an optimal orientation.

Effect of charge transfer

Spectroscopy techniques have revealed, in some cases, the participation of protein aromatic amino acids in charge transfer complexes with coenzymes, substrates or reagents. Charge transfers in systems with π electrons are susceptible to modifying reactivity in different directions.

Diverse effects

Diverse other factors susceptible to modifying the reactivity of protein amino acid side chains must still be mentioned. The formation of metallic complexes tends to weaken the reactivity of a nucleophilic group; indeed, the metal attractor has a tendency to diminish the electron density of the nucleophile. The establishment of reversible covalent bonds such as those driving the formation of a SCHIFF base or of a hemiacetal, tend to render the group provisionally non-functional or to diminish its reactivity.

9.2.1.2. EFFECT OF MICROENVIRONMENT ON THE REAGENT

The protein matrix has not only an effect on the intrinsic reactivity of amino acid residues, but it can also considerably modify the reactivity of the chemical compound used for the selective labelling in different manners and in a way that it is not always possible to foresee.

Selective adsorption of the reagent

Selective adsorption of the reagent on the enzyme before the reaction increases its reactivity by effect of proximity. This is equivalent to locally increasing its concentration. Formation of a complex between the protein and the reagent could induce a saturation effect on the reaction rate (to see later). This presents an advantage for the reaction that is particularly useful in chemical affinity labelling.

Steric effects

A partial steric protection of a residue by the protein matrix risks to notably modify the reactivity of the chemical compound; this protection depends on the size of the reagent. The presence of a substrate, an analog or a coenzyme can prevent the chemical compound from modifying a residue that usually is very reactive.

Electrostatic effects

Electrostatic interactions between the reagent and the charged protein groups or the electrostatic fields situated in the vicinity often create attractive or repulsive effects with the chemical compound following the nature of this last and the nature of the charges.

Proximity effects

The presence of other functional groups situated in the vicinity sometimes has the effect of increasing or on the contrary of diminishing the efficiency of the reagent.

Conformational effects

The presence of the reagent involves in some cases conformational restrictions limiting the rotational freedom of the residue and prohibiting it from taking the necessary conformation for the reaction to happen favorably.

The large number of factors capable of interfering makes the action of a reagent on a protein unforeseeable.

9.2.1.3. REQUIRED CONDITIONS FOR CONDUCTING A CHEMICAL MODIFICATION

Due to its complexity the protein structure makes the chemical approach delicate; also numerous controls are necessary for all the experimental steps. Taking into account all the implicated factors and the variety of proteins, there are not universal methods that permit deciding the choice of a reagent. In addition, when one wants to use a chemical reagent to label the active centre of an enzyme, it must not be forgotten that proteins have a relatively weak conformational stability and that they are susceptible to denaturation. However, chemical reactions on models are often carried out very efficiently in conditions that are hardly favorable to the protein structure. Between two reagents of a given group, one will choose that which is soluble in aqueous medium if it exists, or that which only necessitates a weak proportion of organic solvent; still it will be necessary to verify that this latter does not alter the conformation of the protein.

To be interpreted with a minimum of ambiguity, a chemical modification must only be carried out on a small number of groups, a single one if possible. The stoichiometric inactivation of an enzyme renders the interpretation much easier. *The labelling must be selective and limited.* It is important that the reaction conditions be the gentlest possible, both to limit the labelling and to preserve the native structure of the protein. The reaction can be oriented by different factors, in particular the pH. For example, numerous reagents of methionine that are non-selective in neutral medium become so in acidic conditions, because methionine does not ionise as the other residues that, in these conditions, are protonated and are no longer efficient nucleophiles. The variations in pH or in polarity of the medium can also reversibly modify the protein conformation and render accessible certain groups that are not so at optimal pH. Reversible stoichiometric labelling represents without any doubt, when it is possible, one of the most favorable situations.

The goal being to label functional residues of the active site and only those, the reagents used must be the most selective possible. In addition, it is important to find good conditions so that the labelling occurs selectively.

◤ The technical aspect of chemical modifications must be considered carefully. As a general rule, the study of the topology of an enzyme active centre can only be carried out with very pure proteins. It is preferable that at least the primary structure be known; this permits the identification not only of the nature of the modified residue, but also its position in the polypeptide sequence. This identification necessitates an important analytic work involving cleaving the protein into smaller fragments, the isolation of these peptides, the analysis of peptide maps, and the analysis of the peptide (or peptides) labelled by the reagent. For one uses radioactive reagents and one executes the analysis on the peptides thus labelled. The development of chromatography techniques, in particular the introduction of HPLC, has considerably increased resolution and the rapidity in separation of peptides. ◢

Knowledge of the 3-dimensional structure of a protein, allowing to "mimic" the chemical labelling by molecular modelling and use of energy minimisation methods, offers a remarkable tool for approaching a chemical modification as rationally as possible.

9.2.2. STRATEGY OF CHEMICAL MODIFICATIONS

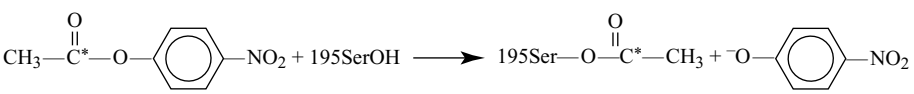
We will analyse successively different strategies used to attempt to selectively label the active centre of enzymes.

9.2.2.1. LABELLING BY A SUBSTRATE, A QUASI-SUBSTRATE OR A COENZYME

One of the reagents susceptible to binding at the active site and forming a covalent bond is the substrate itself or one of its less reactive analogs. Indeed, certain enzymatic reactions proceed via the formation of a covalent compound between one part of the substrate and a residue of the catalytic site (see Chap. 5). This intermediate is generally unstable in the conditions of optimal activity of the enzyme; its lifetime is very short. The labelling by a substrate is only possible in conditions, if they exist, that permit the accumulation of this intermediate.

◤ One of the most classical examples is the acylation of serine 195 of serine proteases by titrants of active sites, like paranitrophenyl acetate (pNPA) or paranitrophenyl guanido benzoate (NPGB) for trypsin. The use of these compounds for the titration of active sites of proteases has been presented previously (see Chap. 5). At pH lower than 5, the acetyl-enzyme formed during the hydrolysis of pNPA by trypsin or chymotrypsin is stable.

In using pNPA labelled with [^{14}C], one recovers the radioactivity on serine 195 after hydrolysis of the enzyme:



This labelling is therefore very selective since it only modifies serine 195. The enzyme thus labelled is totally inactive. Nevertheless its activity can be restored by alkalinisation to the optimal activity pH; in these conditions the acetyl-enzyme is hydrolysed and the active enzyme is regenerated.

Another example of the use of a quasi-substrate is the labelling of triose phosphate dehydrogenase by acetylphosphate. The physiological substrate of this enzyme is 1,3-diphosphoglycerate. Acetylphosphate is a quasi-substrate as paranitrophenyl acetate is for serine proteases.

In the case of enzymatic reactions that proceed *via* the formation of a phosphoryl-enzyme, the phosphorylated residue was identified. For phosphoglucomutase, the active centre serine implicated in the phosphorylation was labelled using glucose-6-phosphate ^{32}P as the substrate. The enzyme obtained after purification is already phosphorylated and catalyses the following reaction:

$$\text{G-1-P} \xleftarrow{\textit{phosphoglucomutase}} \text{G-6-P}$$

The addition of the substrate G-6-P favors the reaction in the reverse direction that can be thus split into:

$$\text{P-enzyme} + \text{G-6*P} \rightleftharpoons \text{enzyme} + \text{G-1,6*-diP}$$

$$\text{G-1,6*-diP} + \text{enzyme} \rightleftharpoons \text{*P-enzyme} + \text{G-1-P}$$

The implicated phosphorylserine was identified in this way (Fig. 9.9). The same type of direct labelling experiment by the phosphate of a specific substrate was achieved with phosphoglycerate mutase. There are some other examples of labelling by phosphorylation of a serine residue: the conversion of phosphorylase b in phosphorylase a by phosphorylase kinase in the presence of $\gamma[^{32}P]$ ATP. Alkaline phosphatase is phosphorylated during the catalytic cycle. These phosphoryl-enzymes are sufficiently stable so analysis of the group labelled by $[^{32}P]$ is possible.

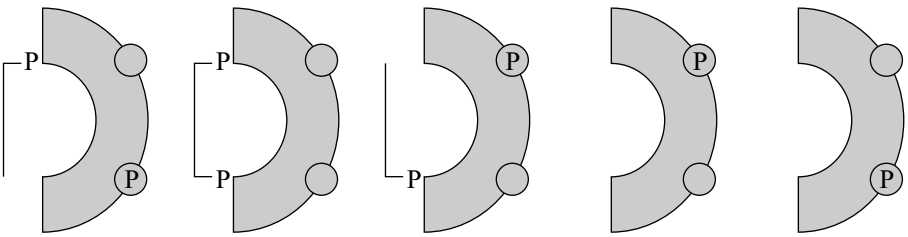
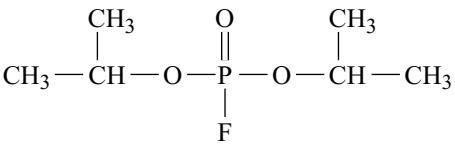


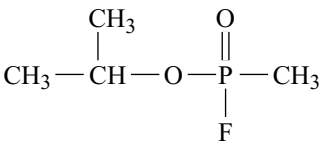
Fig. 9.9 Phosphorylation of the active site of phosphoglucomutase
(Adapted with permission from *Biochemistry*, **4**, HARSHMAN S. & NAJJAR V.A., 2526.

The labelling method by a quasi-substrate consists therefore of using a molecule that resembles a specific substrate. The preceding compounds are either true substrates or quasi-substrates that present the advantage of driving reactions that can be reversed. Some other quasi-substrate compounds of which the structure and reactivity differ from those of physiological substrates drive irreversible reactions. This is the case with certain phosphorylated derivatives that behave as powerful

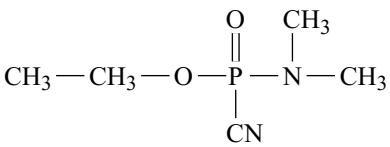
inhibitors of proteases and serine esterases; let us cite for example diisopropyl phosphofluoridate (DFP), sarin, tabun, diethyl p-nitrophenyl phosphate and tetraethyl pyrophosphate (TEPP) presented in Fig. 9.10. All these compounds form a phosphoryl bond with the single serine residue of the active centre of a great number of esterases. They are particularly powerful inhibitors of acetylcholinesterase, and in this way some of them were largely used as combat gas. There exist some cases where the covalent intermediate is unstable but may be stabilised by a supplementary reaction. Alkyl- and aryl-sulfonyl fluorides are powerful serine protease inhibitors. Phenylmethylsulfonyl fluoride (PMSF) is an inhibitor even more powerful than DFP. It is often used during enzyme purifications to avoid protein degradation by proteases present in the cell extracts.



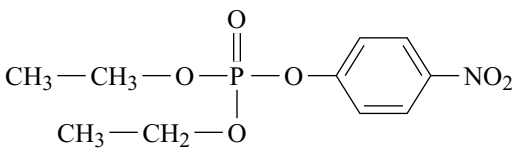
Diisopropylfluoro-phosphate (DFP)



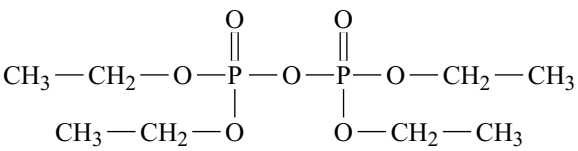
Sarin



Tabun



Diethyl p-nitrophenyl-phosphate



Tetraethyl pyrophosphate (TEPP)

Fig. 9.10 Several phospho derivatives, irreversible inhibitors of serine proteases and esterases

These inhibitors block the serine of the active site in an irreversible manner. Meanwhile to identify this, a supplementary chemical reaction must be used [GOLD, 1965]; phenyl methyl sulfonyl chymotrypsin is easily degradated with the loss of the phenyl methyl sulfonyl group. In the presence of a nucleophile such as 2-mercaptoethylamine in urea 8 M, the O-sulfonyl serine group is partially converted to S-amino ethyl cysteinyl which, being stable, was identified (Fig. 9.11 opposite). The stabilisation by an additional chemical reaction was used to identify the groups implicated in the catalytic site of enzymes which, over the course of the reaction, form a SCHIFF base with their substrate.

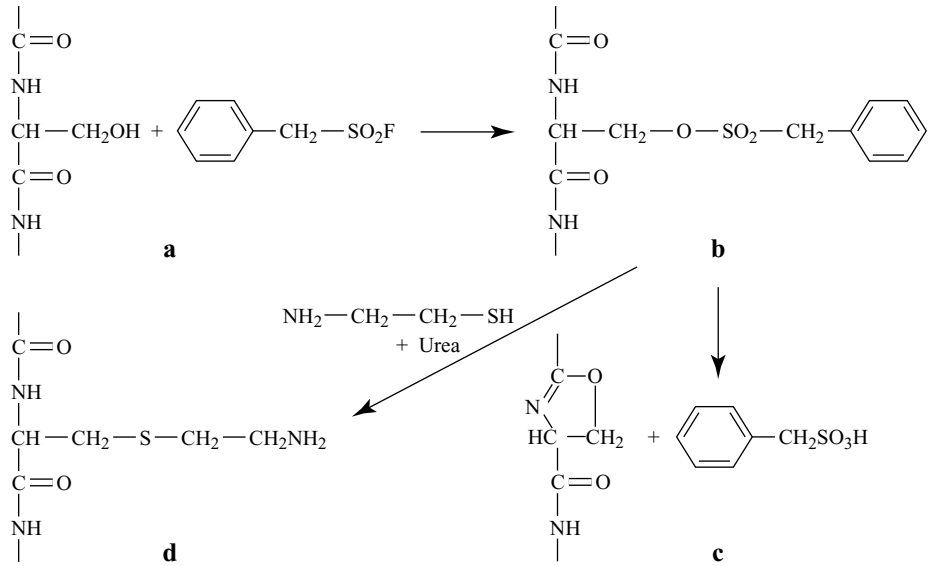
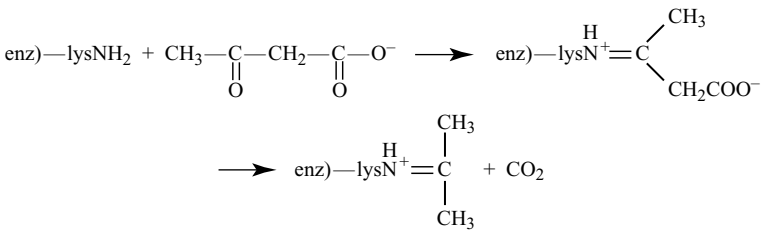


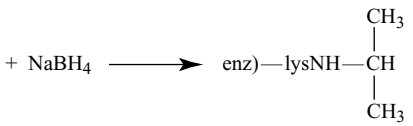
Fig. 9.11 Identification of serine labelled by PMSF

*(**a**) labelling reaction that leads to the compound (**b**), which can be easily degradated in (**c**) – in (**d**) use of 2-mercaptoethylamine in urea 8 M that transforms the O-sulfonyl serine group in S-aminoethyl cysteine*

◤ This is the case with acetoacetate decarboxylase that catalyses the decarboxylation of a keto acid, giving rise to a SCHIFF base following the reaction:



The unstable SCHIFF base can be stabilised by reaction with sodium borohydride:



The use of tritiated borohydride has permitted the isolation and identification of an isopropyl-ε-lysine showing the catalytic role of a lysine residue in this enzyme. The particular role played by a lysine at the catalytic site of aldolases and transaldolases was demonstrated in this manner.

Likewise the reduction by sodium borohydride has shown the role of a lysine in the binding of pyridoxal phosphate (PLP) in a certain number of enzymes, in particular in transaminases like aspartate amino transferase. The coenzyme PLP is bound to the protein by a lysine group as a SCHIFF base (Fig. 9.12 below). By an analogous ap-

proach, PLP allowed the identification of the active site of aspartate transcarbamylase even though it is not a cofactor of this enzyme. It behaves as a competitive inhibitor forming a SCHIFF base with a lysine residue localised at the active centre (KEMPE & STARK, 1975).

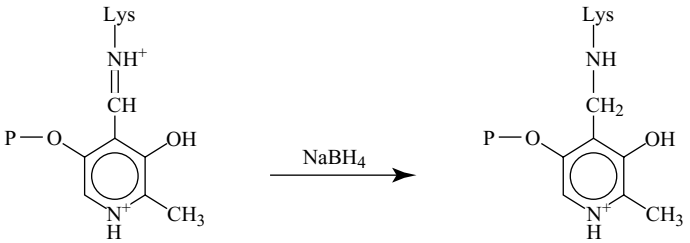


Fig. 9.12 Reduction by sodium borohydride of the SCHIFF base formed between a lysine of aspartate amino transferase and the pyridoxal phosphate ◢

However, there exist some cases where the use of a quasi-substrate does not result in labelling of a group implicated in the catalysis. Let us cite for example the labelling of chymotrypsin by bromoacetyl aminobutyrate. This reagent is supposed to acylate the single serine of the active site, which happens in a first step. Following this there exist two possible paths, either the simple deacylation like for an ordinary substrate, or the alkylation of another group, methionine 192, by the bromoethylketone with consecutive cleavage of the acyl-serine bond (Fig. 9.13 opposite). The labelled methionine is a non-catalytic residue; the serine is finally free and the enzyme thus modified presents still a weak catalytic activity.

9.2.2.2. AFFINITY LABELLING

Affinity labelling uses the structural characteristics of substrates to direct a reactive function (acylating or alkylating) to the active centre of an enzyme; these reagents are called *active-site directed reagents*. They require therefore a part possessing the structure of specific substrates on which is grafted a reactive function. The effective concentration of the reagent at the active centre is thus considerably increased.

In this case, like for quasi-substrates, the reaction proceeds *via* the formation of a complex between the reagent and the enzyme:

$$E + R \underset{}{\overset{K}{\rightleftharpoons}} ER \xrightarrow{k} E\text{-}R$$

R being the reagent and E-R the modified enzyme, K the dissociation constant of the complex ER and k the specific rate constant of the chemical reaction. The reaction rate is then given by a relationship implicating a Michaelian saturation:

$$v = \frac{ker}{K + r}$$

whereas usually one has a second order reaction:

$$v = ker$$

r being the concentration of the reagent and e that of the enzyme. If the reagent is in excess such that the variation in its concentration is negligible over the course of the chemical reaction, one has:

$$v = k'e \qquad \text{with } k' = kr = Ct$$

k′ is a constant of pseudo-first order (see Chap. 4).

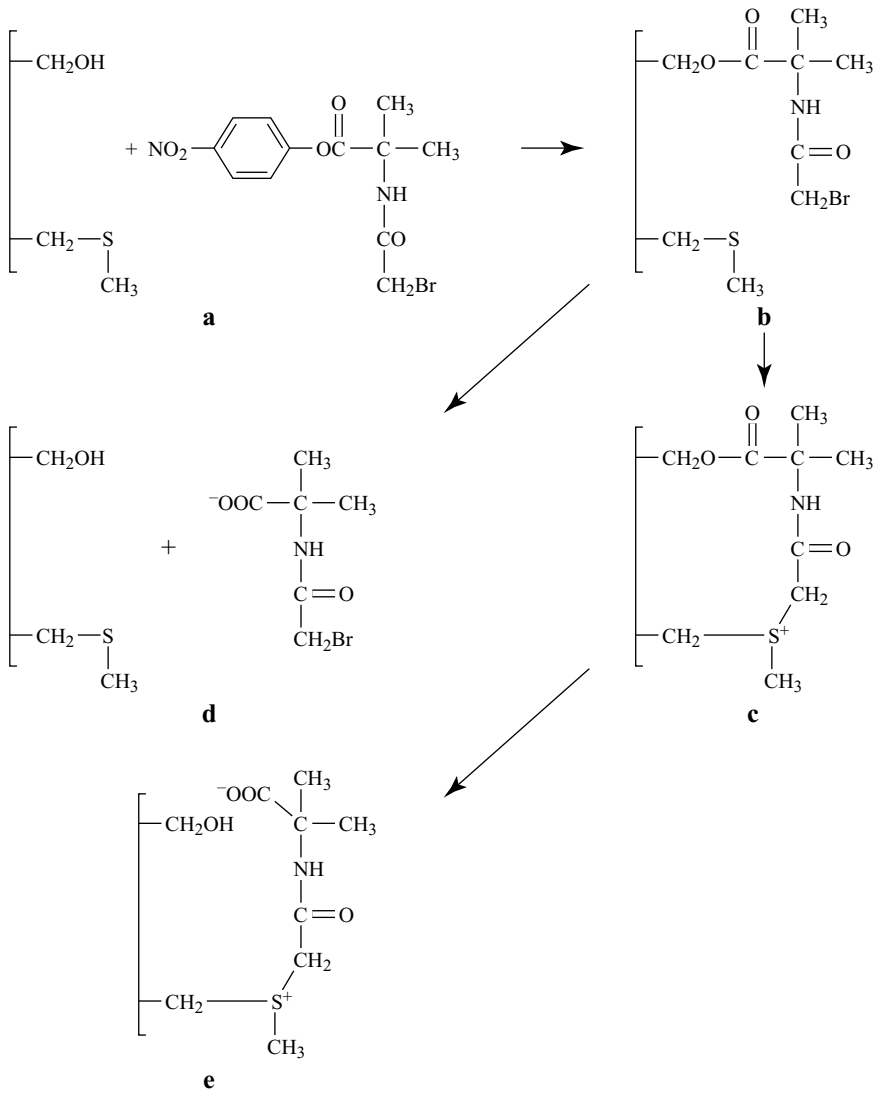


Fig. 9.13 Labelling of chymotrypsin by bromoacetyl amino isobutyrate

One of the first applications of this method permitted SCHOELLMANN and SHAW (1963) to demonstrate the catalytic role of His 57 in chymotrypsin. The reagent was tosyl phenylalanine chloromethyl ketone or TPCK. Figure 9.14 shows the analogy between a specific substrate, tosyl phenylalanine methyl ester (TPME) and TPCK. The latter provokes the progressive disappearance of the enzymatic activity with the correlated labelling of histidine 57. At the end of the reaction the inhibition is complete. The reaction is slowed by the presence of a competitive inhibitor like phenylpropionate (Fig. 9.14).

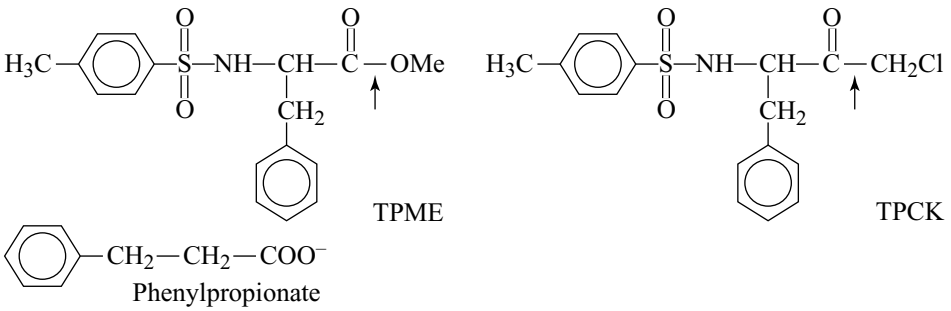


Fig. 9.14 TPCK, affinity reagent of chymotrypsin

TPCK does not bring about any inhibition of the activity of trypsin. On the other hand SHAW et al. (1965) showed that its homolog tosyl lysine chloro methyl ketone (TLCK) is specific for trypsin and does not inhibit chymotrypsin (Fig. 9.15). These compounds permit therefore a selective inhibition of one or the other enzyme. Thus some traces of trypsin present in a commercial preparation of chymotrypsin can be specifically inhibited by TLCK, and reciprocally TPCK is used to inhibit all possible traces of chymotrypsin in a preparation of trypsin.

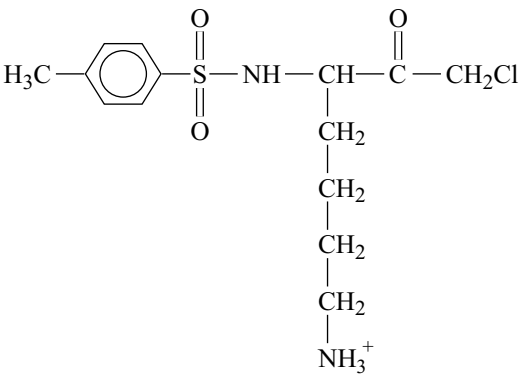


Fig. 9.15 Structure of TLCK, affinity reagent of trypsin

◤ Proteolytic enzymes were not only the first but also the most studied for directing the synthesis and use of affinity reagents. This strategy was progressively developed and such compounds were introduced to study other enzymes. Thus, at the same time, despite the difficulty to obtain pure acetylcholinesterase in sufficient quantity,

some affinity reagents of this enzyme were synthesised. Esters of methyl sulfonate possessing a quaternary ammonium such as the compound (a) (Fig. 9.16) inactivate the enzyme irreversibly. We have seen in the preceding paragraph that the quasi-substrates like organophosphates are powerful inhibitors of acetylcholinesterase. Certain carbamates like neostigmine (compound (c), Fig. 9.16) provoke their inhibition by carbamylation of the serine of the active centre.

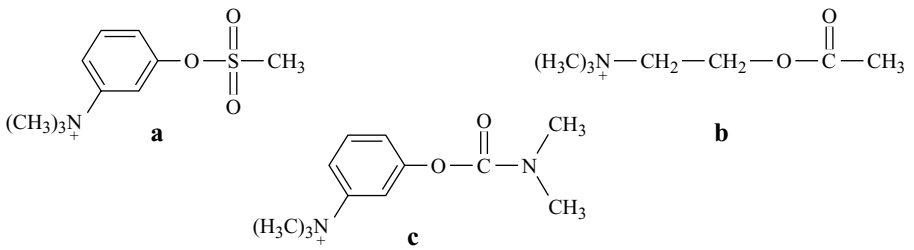


Fig. 9.16 Affinity reagents of acetylcholinesterase (a) and (c); the compound (b) is the substrate

Some active site-directed reagents of adenosine deaminase in which the bromoacetamide chain was branched in the *ortho, para* or *meta* position of 9-benzyl adenine (Fig. 9.17) were synthesised. The *ortho* and *para* isomers completely inhibit the enzyme whereas the substituted *meta* compound only provokes a partial loss of activity (80%); it is likely that, in this last case, the alkylation site is different from that of the *ortho* and *para* compounds. **Not only is it important that the reagent presents a specific part for directing it to the active centre, but also the orientation of the chemical function susceptible to label a catalytic group must be favorable.**

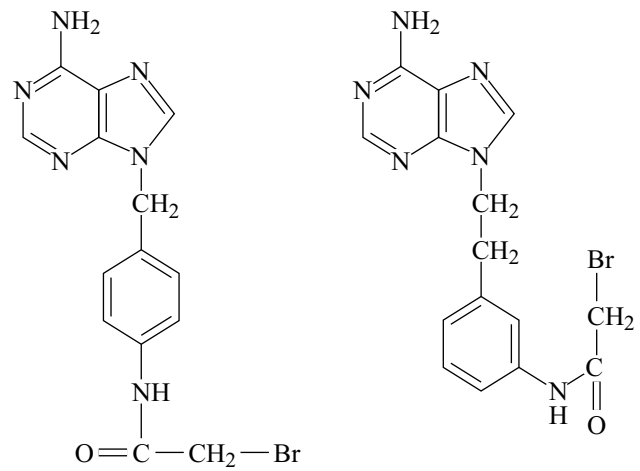


Fig. 9.17 Affinity reagents of adenosine deaminase

◤ Numerous active site-directed reagents were synthesised to inhibit glycosidases (see the review by LALLEGERIE et al., 1982). On some of them an epoxide function is grafted, which is the case for inhibitors of α- and β-glucosidases (compounds (a) and (b), respectively, in Fig. 9.18) and for an inhibitor of lysozyme on which has been grafted β-epoxy-propyl aglycone (compound c). In all these cases the labelled group was a carboxyl. Certain isothiocyanate derivatives, like β-D-gluco pyranosyl isothiocyanate (compound d) are powerful inhibitors of β-glucosidases.

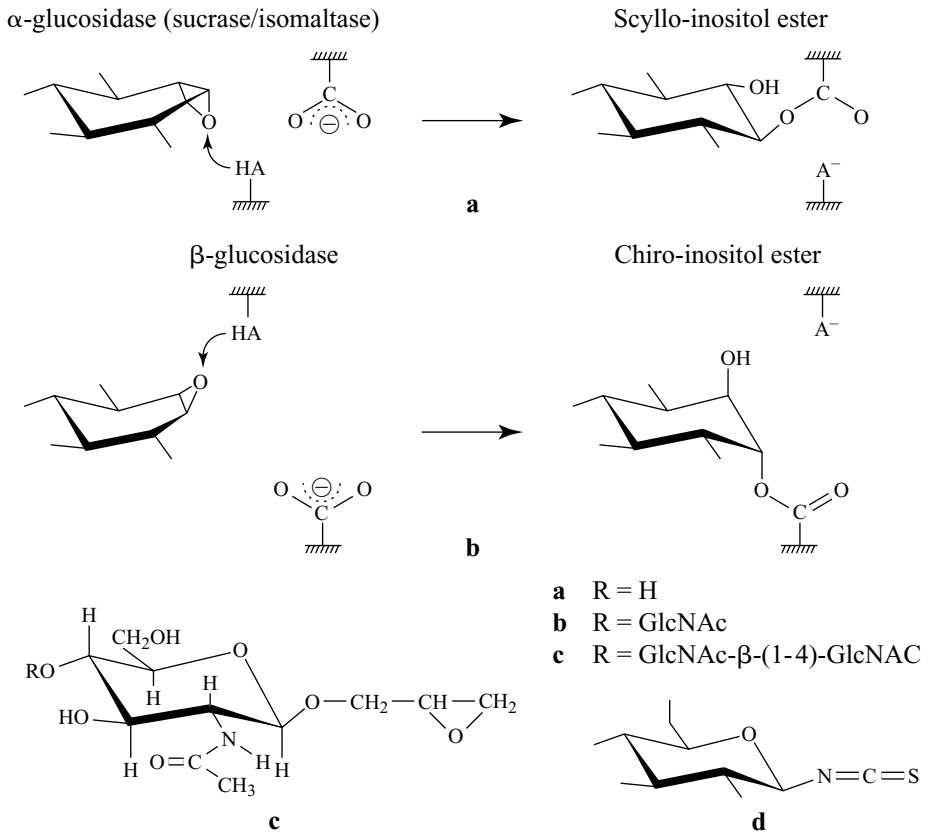


Fig. 9.18 Affinity reagents of several glycosidases ◢

Although they bind to the binding site of substrates, affinity reagents do not necessarily react with a catalytic group.

◤ Some reagents of chymotrypsin (Fig. 9.19 opposite) like benzyl bromide (a), and phenacyl bromide (b) modify methionine 192 which is near the active site but is not implicated in the catalysis, and the labelling only provokes 50% inactivation. The presence of specific substrates or inhibitors meanwhile brings about a protection of the enzyme with respect to the chemical labelling. **It is important therefore to interpret the results with caution.**

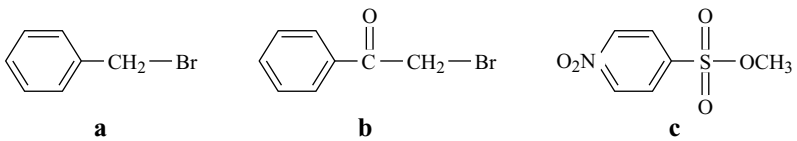


Fig. 9.19 Affinity reagents of chymotrypsin reacting with methionine 192

N-bromoacetyl-β-D-galactosylamine inhibits β-galactosidase in an irreversible way by alkylating methionine 500. Nevertheless this residue is not implicated in the catalysis. Bacteria grown in the absence of methionine, this being replaced by selenomethionine, synthesise a fully active β-galactosidase. ◢

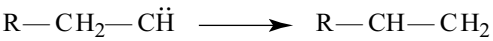
The use of affinity reagents is not limited only to the determination of catalytic groups. In some cases, it permits at the active site the introduction of a fluorescent reporter group or a spin label, comprising for example a nitroxide radical. Let us cite for example 2,2,5,5-tetramethyl-3-carboxypyrolidine-1-oxide phenyl ester that binds at the active site of chymotrypsin. This label allows following, by variation in fluorescence or by electron paramagnetic resonance (EPR), variations in the environment of the active centre under the influence of diverse effectors.

9.2.2.3. PHOTOAFFINITY LABELLING

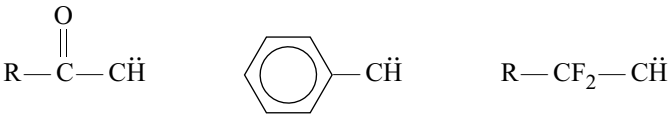
The principle of photoaffinity labelling resembles that of affinity labelling in the sense that the reagent possesses a specific part that permits it to interact with the active centre of the enzyme, thus serving as a vector to direct a chemical function near the active site. But the difference from affinity labels is that the chemical function is not initially reactive; it is susceptible to being activated by photolysis. The activation must be carried out in gentle conditions and at wavelengths such that the protein does not undergo photochemical damage. The derivative activated by photolysis must be highly reactive, not suffer molecular rearrangements producing a less reactive compound and have a very short lifetime. These reagents are capable of chemically modifying inert groups like methyls and methylenes in side chains of hydrophobic amino acids alanine, valine, leucine, isoleucine etc.

There exist different types of photoactive groups; these are precursors of carbenes and nitrenes. Irradiation of the precursors brings about the formation of excited electron states driving the appearance of reactive species.

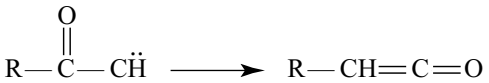
Carbenes react very rapidly with a great variety of chemical functions, either with a nucleophile group to give a carbanion, or by addition of multiple bonds including those of aromatic compounds, by insertion in single bonds including C—H bonds, or by abstraction of a hydrogen. If the adjacent carbon carries a hydrogen, this latter migrates easily bringing about the formation of a non-reactive compound:



Therefore, the adjacent atom must not carry a hydrogen. One generally chooses compounds of the type:

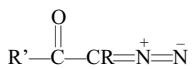
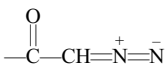
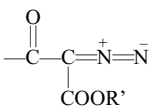
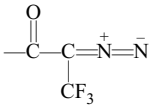
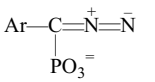
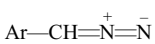
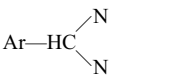


Acetocarbenes can undergo an intramolecular rearrangement or a WOLFF rearrangement driving a ketene:



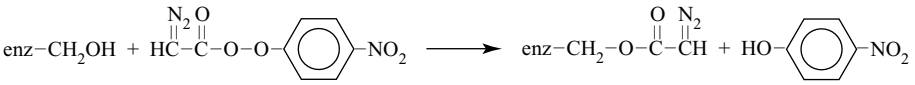
The ketene thus formed is reactive but can only interfere with its acylating function that reacts with a nucleophile group whereas carbenes present multiple reactivities. Precursors that produce carbenes by a photochemical reaction are indicated in Table 9.2.

Table 9.2 Principal precursors of carbenes (Reprinted from *Methods in Enzymol.*, **46**, BAYLEY H. & KNOWLES J.R., Photoaffinity labeling, 69. © (1977) with permission from Elsevier)

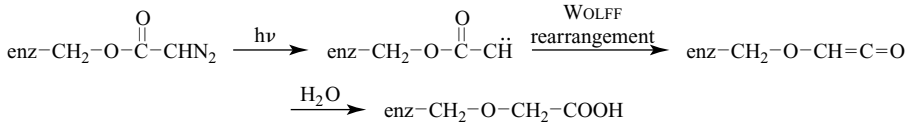
Type of reagent	*Formula*	*Stability in the dark in neutral solution*	*Sensitivity to intramolecular rearrangement after photolysis*
α-diazocetones		limited	high
α-diazocetyl-		limited	high
α-diazomalonyl		fair	reasonable
Trifluorobenzyl-phosphate		good	rather weak
α-diazobenzyl-phosphate		weak	—
Aryldiazomethane		weak	weak
Aryryldiazirine		good	weak

◤ Diazoacetyl derivatives give carbenes by photolysis at 370 nm. The first example of the application of this method by WESTHEIMER (1962) was carried out on the labelling of

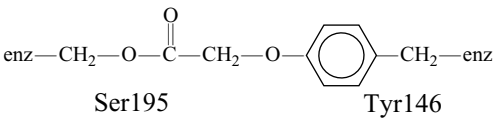
chymotrypsin. In fact, the authors used p-nitrophenyl-2-^{14}C-diazoacetate that reacts with serine of the active centre of chymotrypsin following the reaction:



At pH 6.2 diazoacetyl serine is stable for at least 48 hours. Its stoichiometric formation drives the inactivation of the enzyme. Photolysis of ^{14}C diazoacetyl chymotrypsin provokes incorporation of 20–25% of the radioactive reagent in the protein. Among amino acids carrying radioactivity, serine 195, tyrosine 146 and histidine 57 were identified. O-carboxymethylserine is formed after a WOLFF rearrangement following the reactions:

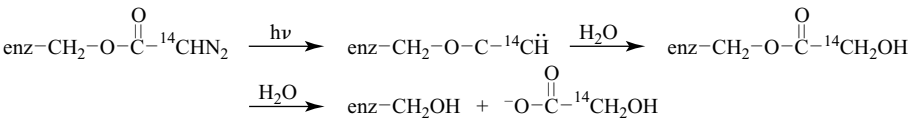


O-carboxymethyl tyrosine is formed by insertion of carbene on the hydroxyl group of tyrosine 146 of another molecule of chymotrypsin thus forming a dimer:



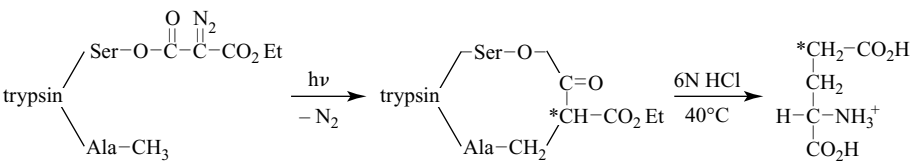
Ser195 Tyr146

One part of the enzyme is regenerated with the formation of glycolic acid according to the reactions:



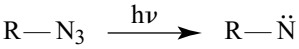
The diazoacetyl derivatives were also used as affinity reagents directed to the active centre of subtilisine and glyceraldehyde-3-phosphate dehydrogenase. Diazomalonyl derivatives were used with success in the case of chymotrypsin. They are more stable than the diazoacetyl derivatives and are less susceptible to undergoing WOLFF rearrangements.

Photolysis of diazomalonyl chymotrypsin as well as that of diazomalonyl trypsin leads to the same products as diazoacetyl derivatives, except for the acquisition of a glutamic acid with diazomalonyl trypsin following the reaction:



Arylazidomethanes are generally too reactive to be used with proteins. Aryldiazirines, more stable in the dark, are more interesting compounds than the preceding precursors.

Nitrenes participate also in a great variety of reactions, but their reactivity is weaker than that of carbenes. Nitrenes also result in the activation of a precursor by light:



There are several categories of nitrene precursors, acylazides, phosphorylazides, sulfonylazides, arylazides and nitroarylazides. The rearrangements that happen in the case of nitrenes are of the same type as those observed with carbenes. Therefore irradiation of 2-azido-biphenyl (Fig. 9.20) drives the intermediate of nitrene to the formation of carbazole.

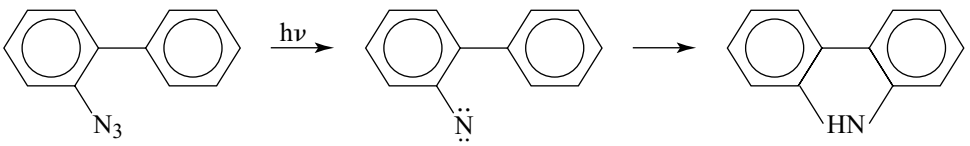


Fig. 9.20 Reaction of 2-azido biphenyl under the effect of irradiation

Among the nitrene precursors used to label enzymes, the most common are bound to nucleotides, such as 8-azido-cAMP, 8-azido-ATP, diazido-ATP (Fig. 9.21), and β-azido-NAD. They are used to label the active sites of kinases, ATPases and concerning the last reagent, NAD^+ dehydrogenases. Such compounds serve also in the study of receptors or membrane transporters, for example the receptors of nucleotides and cyclic AMP.

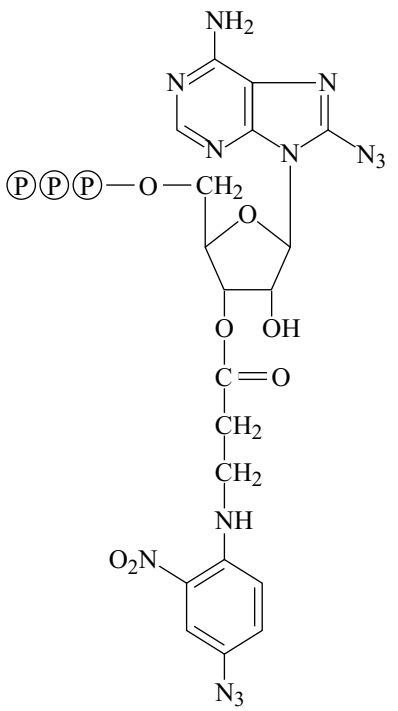


Fig. 9.21 Structure of diazido-ATP

In order to avoid non-specific labelling of the protein, one often resorts to scavengers added to the medium that trap the reagent molecules not bound to the active centre. For this purpose, p-amino benzoic acid, p-aminophenylalanine, dithiothreitol, ®-mercaptoethanol, tris or even amino acids are used.

9.2.2.4. SUICIDE REAGENTS

"Suicide" reagents are substrate analogs that are recognised by the enzyme, and the first transformation steps are the same as in the normal enzymatic reaction, followed by the generally covalent formation of an intermediate complex, EX. The reagent function is generated by the formation of this complex; it is only potential in the initial form S′ of the substrate. At this stage, the complex EX can evolve following two paths, either the formation of the product P′ with regeneration of the enzyme, or the irreversible formation of an inactive complex EI according to the scheme:

$$E + S' \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES' \xrightarrow{k_2} EX \begin{array}{l} \xrightarrow{k_3} E + P' \\ \xrightarrow{k_4} E\text{—}I \end{array}$$

The partition coefficient is given by the ratio of the rate constants k_3/k_4.

◤ Suicide substrates were used to study diverse enzymes. In the case of serine proteases, YON collaborators (1973) showed that 3,4-dihydro 3,4-dibromo methyl coumarine VII inhibits α-chymotrypsin by blocking both serine 195 and histidine 57 following the mechanism that is indicated in Fig. 9.22.

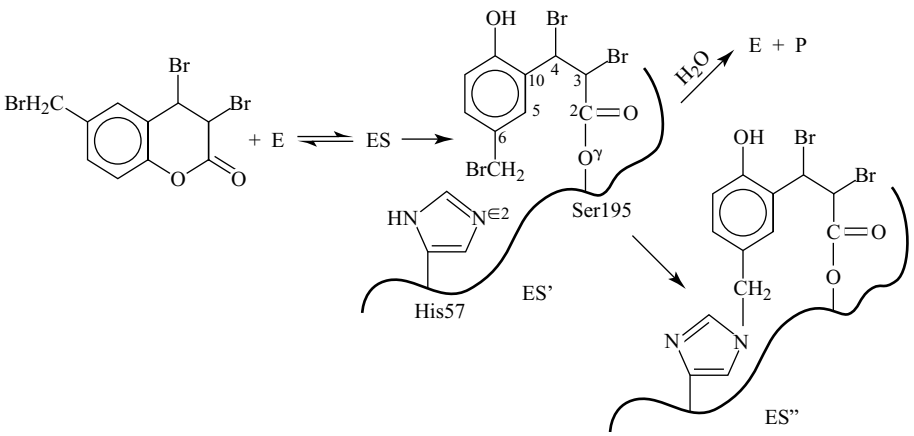


Fig. 9.22 Inhibition of chymotrypsin by a suicide substrate

Sulbactame was used as an inhibitor of β-lactamase. It behaves also like a suicide substrate. The mechanism of action was demonstrated by using the deuterated compound, 6,6-D_2-sulbactame. The different paths of transformation of this compound in the presence of the enzyme are represented in Fig. 9.23 below.

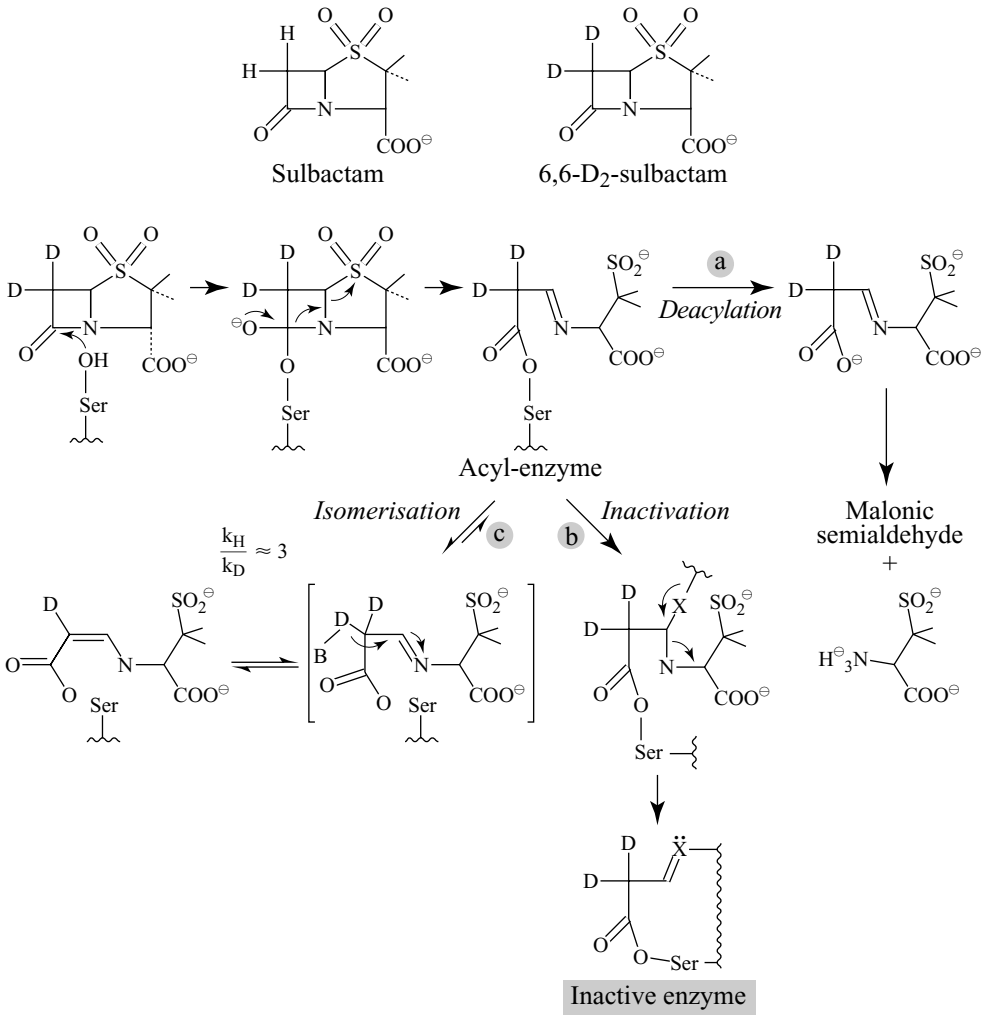


Fig. 9.23 Inhibition of β-lactamase by sulbactame ◢

The use of these reagents that form inactive compounds by blocking two of the catalytic groups permits not only the identification of these groups but also the evaluation of their distance according to the size of the reagents. **The use of "suicide" reagents is the object of important applications in molecular pharmacology.**

9.2.2.5. DIRECT LABELLING BY SELECTIVE REAGENTS

One does not always have reagents susceptible to being directed by affinity to the active centre of enzymes. Research and synthesis of these diverse types of compounds, affinity reagents, photoaffinity and suicide reagents is often laborious, and the examples we have cited represent favorable cases. The enzymologist often has no other recourse except the use of much less selective labelling. Direct labelling

accepts as selectivity criteria the preferential reaction of residues of the active site towards appropriate chemical reagents. Labelling procedures used consist of making reagents act, as selectively as possible, on the enzyme in solution. The reactivity of residues depends greatly on the pH and polarity of the solvent. Direct labelling must be accompanied by an analysis of modified residues, the determination of their position in the polypeptide chain and their importance in catalysis (to see later). In Sect. 9.2.2.7, some principal reagents of the amino acid side chains in proteins are presented.

9.2.2.6. DIFFERENTIAL LABELLING

In order to obtain a greater selectivity, one often resorts to differential labelling introduced by COHEN and WARRINGA (1953), then by KOSHLAND and collaborators (1959). Covalent labelling is carried out in two stages. In the first stage, often called the "pre-labelling stage", the protein is incubated in the presence of a large excess of substrate or a non-reactive analog of the substrate (a totally reversible competitive inhibitor); one chooses conditions such that the enzyme has a great affinity for the ligand. The goal of this stage is to protect the active centre against the reagent used for pre-labelling. Of course, the protective effect must be verified before undertaking the chemical modification. The enzyme thus protected is subjected to the action of the chemical reagent in great excess in order to modify all the residues of the same nature located away from the active site. The mixture is dialysed exhaustively or subjected to an exclusion chromatography in order to drive away the reagent in excess and to displace the inhibitor of the complex. In the second stage, the pre-labelled enzyme cleared of its inhibitor, with its active site becoming accessible, is subjected to the same chemical reaction as previously, but in the presence of the chemical reagent labelled by a radioactive isotope. All experimental conditions except the presence of the ligand must be identical to those in the first stage. It is necessary afterwards to identify the residue that carries the radioactivity. Each stage requires controls.

A quantitative description of the protection by the specific ligand in the differential labelling was presented by SINGER (1967). The rate of modification of a residue Y by a reagent R is given by the equation:

$$\frac{-dY}{dt} = k(Y)(R)$$

k is the second order rate constant, (Y) the concentration of unlabelled sites at time t, and (R) the concentration of the reagent in excess as to assure conditions of pseudo-first order. If x represents the fraction of residues Y that are modified, (1 – x) is the fraction of non-modified groups and one has:

$$\frac{dx}{dt} = k(1-x)(R)$$

The rate of modification in the presence of a protector ligand is the sum of the reactions of R with the free sites and the protected sites:

$$\frac{dY_p}{dt} = k(Y)(R) + fk(YP)(R)$$

The factor f is a measure of the degree of protection. The protection is total if f = 0; it is absent if f = 1. The binding equation leads to:

$$(YP) = K_P(P)(Y)$$

K_P being the association constant of P with Y. As previously mentioned, the labelling rate of the complex is the following:

$$\frac{dx_p}{dt} = \frac{k(1-x)(R)}{1+K_p(P)} \times [1+fK_p(P)]$$

The ratio of the chemical modification rates in the absence and presence of ligands is a function of the affinity of the enzyme for the ligand:

$$\frac{\frac{dx}{dt}}{\frac{dx_p}{dt}} = \frac{[1+K_p(P)]}{[1+fK_p(P)]}$$

9.2.2.7. PRINCIPAL REACTIONS OF AMINO ACID SIDE CHAINS

We present the most common reactions of different functional groups in the proteins. One cannot avoid here a slightly tedious enumeration; hopefully the reader pardons us. Recapitulative tables are given for the very practical goal of helping the experimenter in search of the adequate reagent to label a given amino acid.

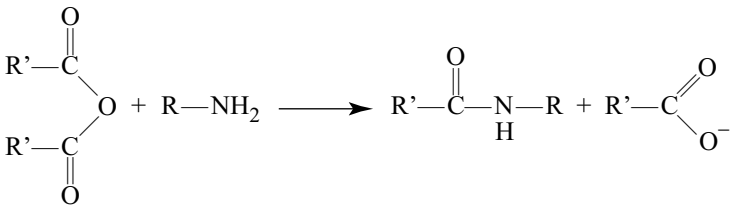
α- and ε-amino groups

The reagents of α- and ε-amino groups are the same. Sometimes however, it is possible to preferentially label the α-amino group by achieving the labelling in weak-ly basic conditions; the ε-amino groups react then more slowly than the α-amino groups as a consequence of their higher pK. Nevertheless there are exceptions when the α-amino residue is buried or participates in an interaction; this is the classical example of the NH_3^+ N-terminal group that, in serine proteases, participates in the formation of a salt bridge. Consequently its pK is shifted towards higher pH.

The principal reactions in which the α- and ε-amino groups participate are acylation, arylation, alkylation, the addition on a double bond and deamination.

◤ *Acylation*

Acylation is the attack of a trigonal sp^2 carbon by a nucleophile group, the nitrogen doublet of the amino group driving the formation of a bond of type R′–CO–NH–R. For this reaction, acid anhydrides are generally used:

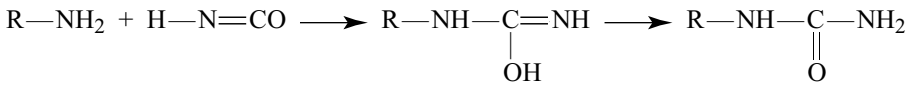


Depending on the nature of the radical R′, it is possible to conserve the positive charge (by using N-carboxyanhydride of α-amino acid), to introduce a negative charge (by using succinic, maleic, citraconic, or tetrafluoro succinic anhydrides), or even to neutralise the charge (by using acetic anhydride). These reagents are given in Table 9.3 below.

The amino group can be regenerated in different conditions according to the stability of the bond formed. The modification by maleic anhydride is easily reversed in acidic medium where the hydrolysis of the maleylamide bond is catalysed by the non-ionised carboxyl group. The half life of the maleyl-lysine at pH 3.5 and at 37ºC is around 10 h. Citraconylation is also reversible in acidic medium. Acylation by tetrafluorosuccinic anhydride gives a stable derivative in acidic medium, but it easily decomposes at pH 9.5.

These reagents are poorly selective; they are susceptible to reacting with other residues like cysteine, histidine and tyrosine. However these amino acids give unstable esters in alkaline medium.

Carbamylation by cyanate is a fairly selective reaction in alkaline medium:



Generally, carbamylation is carried out with a reagent of the form R—N═C═O. The reaction is performed around pH 8. Cysteine and tyrosine can also react reversibly at lower pH (pH 5-6).

Phenylisothiocyanate (PITC) drives an analogous reaction giving rise to the appearance of phenylthiocarbamyl with the ⟨-amino N-terminal group; this derivative cyclises in thiazolinone, then in phenyl thiohydantoin (PTH), bringing about a sequential degradation of proteins. This reaction is used in the EDMAN method that permits the sequence analysis of a polypeptide chain starting from the N-terminal extremity (see Table 9.3). Today one generally uses for microsequencing dimethylaminoazobenzene isothiocyanate (DABITC) that leads, by the same principle as previously, to the formation of a colored dimethylaminobenzene thiohydantoine (DABTH) derivative. The use of this reagent permits a very great sensitivity of detection and renders possible sequence determinations with less than a nanomole of polypeptide.

Guanidylation by action of O-methyl thiourea is performed principally on amino groups of lysine residues following the reaction:

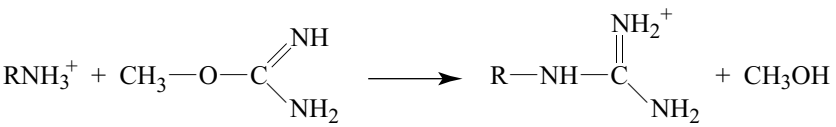
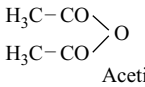
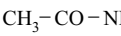
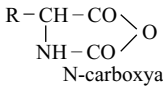
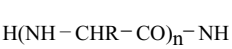
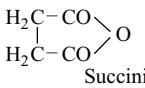
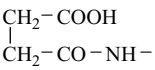
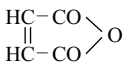
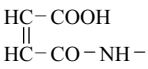
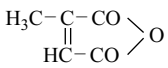
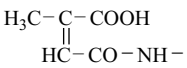
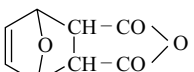
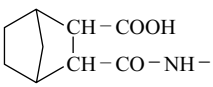
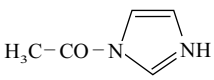
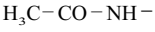
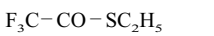
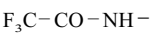
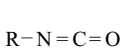
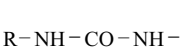
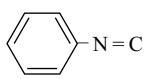
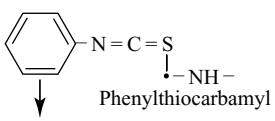
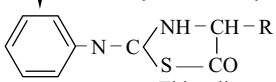
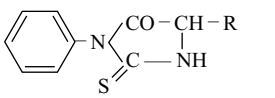
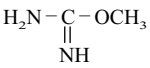
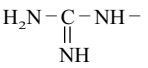
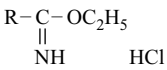
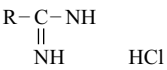
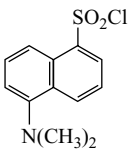
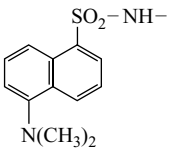
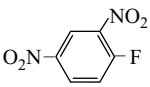
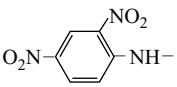
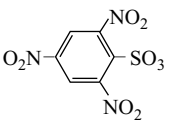
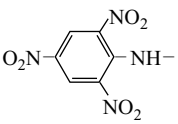
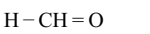
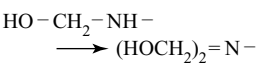
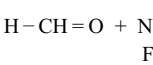
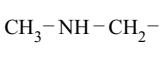
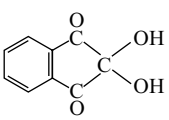
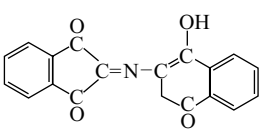
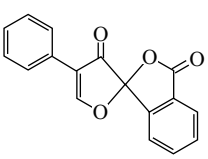
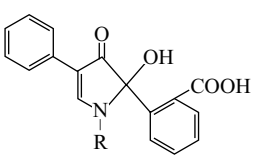
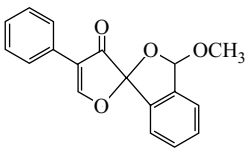
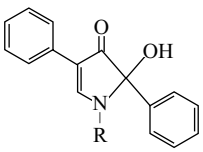
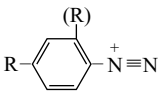
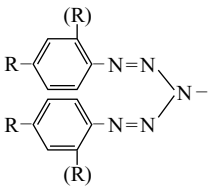


Table 9.3 Reactions of α- and ε-amino groups

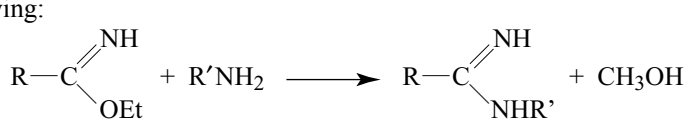
Reagent	*Product*	*Use*	*Comments*
 Acetic anhydride			
 N-carboxyamino acid anhydride		Introduction of polyR amino acids into the peptide	
 Succinic anhydride		Replacement of the positive charge of NH_2 by the negative charge of COOH	Stabilisation of insoluble polypeptides, limited trypsine cleavage next arginine
 Maleic anhydride		–	
 Citraconic anhydride		–	Regeneration of NH_2 in acetic media. Reagents differing in the stability and specificity of the bond
 Exo-cis-3,6-endoxo-Δ_4-tetrahydrophtalic anhydride		–	
 Acetylimidazole			
 Ethylthiotrifluoroacetate			
 Isocyanate		Reversible blocking of lysine ε-NH_2	Trypsin only cleaves at the side of arginine residues
 Phenylisothiocyantate (PITC)	 Phenylthiocarbamyl	Determination of the N-terminal	
	 Thiazolinone	Sequential degradation of proteins (EDMAN)	
	 Phenylthiohydantoin		

Reagent	*Product*	*Use*	*Comments*
		Replacement of NH_2 by the guanidyl group	
		Substitution of NH_2 without changing the charge	
 Dansyl chloride (DNSCl)	 (DNS−)	Determination of the N-terminal group and the ε-amines of lysine	Very stable fluorescent dansyl amino acids; high sensitivity of detection
 Dinitrofluorobenzene	 Dinitrophenyl− (DNP−)	Determination of the N-terminal group	Hydrolysis by HCl gives stable DNP amino acids (yellow)
 2,4,6-trinitrobenzene sulfonate (TNBS)	 Trinitrophenyl− (TNP−)	Colorimetry of α- and ε-amines	Hydrolysis by HCl gives TNP amino acids and regenerates amino acids
 Formaldehyde		Preparation of anatoxins	
 Formaldehyde + Sodium borohydride		Methylation of α- and ε-amines; introduction of ^3H or ^{14}C	
 Ninhidrine		Detection and colorimetry of NH_2	
 Fluorescamine		Detection and highly sensitive fluorimetry of NH_2	

Reagent	*Product*	*Use*	*Comments*
 2-methoxy-2,4-diphenyl-3(2H)-furanone (MDFP)			
$CH_2=CH-CN$ Acrylonitrile	$R-NH-CH_2-CH_2-CN$		
 Aryldiazo salts			Colorless product formed from free NH_2, contrary to colored products formed from histidine or tyrosine
$O=N-OH$ Nitrous acid	$-NH_2 \longrightarrow -OH + N_2$	Deamination	

It occurs around pH 10–11 in a very slow reaction. The positive charge of lysines is not lost by this modification. There is an increase in the length of the side chain that is transformed in homoarginine. Due to its weak reactivity and weaker nucleophile ability of its α-amino group, this reagent reacts selectively with ε-amino groups. More reactive S-methyl isourea shows little difference in reactivity towards α- and ε-amino groups.

Certain esters, thioesters or imidoesters are very reactive. Imidoesters are easily attacked by nucleophilic reagents and react selectively with amino groups of proteins between pH 8.5 and 9.5 to form amidines. **Amidination** leads to complete blockage of amino groups if relatively strong reagent concentrations are used. The reaction is the following:



Lysine residues thus blocked can be regenerated by incubation of the amidine derivative in the presence of ammoniac or hydrazine. Imidoesters in a longer chain such as dimethyl suberimidate or adipimidate are used as cross-linking reagents.

Sulfonylation is obtained by action of aliphatic or aromatic sulfonic acid halogenides on the amino groups of proteins. The most used reagent is S-dimethylamino-naphthalene-1-sulfonic acid chloride or dansyl chloride. Dansylation of amino groups gives fluorescent derivatives. For this reason, this reaction is used to introduce an extrinsic fluorescent label into a protein for conformational studies (see Table 9.3).

Alkylation and arylation

These reactions give rise to compounds of the type:

–NH–C or NH–Ar

They are brought about by action of aryl halogenides of which the most classic is dinitrofluorobenzene (DNFB) introduced by SANGER in 1945. The order of reactivity of the leaving group is:

$$F > Cl \sim Br > SO_3H$$

DNFB gives dinitrophenyl derivatives with α- and ε-amino groups. The reaction requires that the nucleophilic group not be protonated; at pH 8.5 and at lower pH, the ε-amino groups react more slowly than the α-amino groups. In ribonuclease therefore, dinitrophenyl lysine 41 is preferentially formed following the great reactivity of this particular group. Dinitrophenyl derivatives are stable in alkaline conditions and easily measurable by spectrophotometry. The compound formed with an amine absorbs in the visible spectrum with a maximum at 365 nm.

DNFB also reacts with other side chains in proteins, in particular the SH group of cysteines, the phenol group of tyrosines and the imidazole group of histidines. These amino acids substituted by the dinitrophenyl group can be regenerated by β-mercaptoethanol which displaces the reagent. This procedure therefore permits a selective labelling.

2,4,5-trinitrobenzene sulfonate (TNBS) also reacts with amino groups and is used as a reagent in colorimetric measurements. The absorption spectra of TNP lysine and TNP amino acids present a maximum around 345 nm with a shoulder at 420 nm. TNBS reacts little with the imidazole nitrogen or with the hydroxyls of tyrosine, serine and threonine; it does not react with arginine. It forms labile derivatives with SH groups.

There exist other reagents of amino groups used essentially for the spectroscopic detection of these residues; they give rise to colored derivatives like ninhydrine or fluorescent derivatives like fluorescamine or 2-methoxy-2,4-diphenyl-3H-furanone (MDPF) that allow a highly sensitive detection by fluorimetry.

Aliphatic aldehydes and ketones react rapidly and reversibly with protein amino groups. The products formed can be reduced to give stable alkylamino groups. In lightly alkaline conditions, residues other than the amino groups do not give stable derivatives. Formaldehyde reacts with lysine in the presence of sodium borohydride to give dimethyl lysine. If one uses tritiated borohydride, one introduces thus a radioactive labelling that permits the identification of the implicated lysine.

Addition reactions on a double bond

Nucleophilic addition of acrylonitrile was used to alkylate SH groups. The latter react in lightly alkaline conditions. At higher pH (9.5) the reaction with the amino groups becomes significant. At pH 9.2, all the lysine residues of ribonuclease are labelled by the reagent. The reaction is the following:

$$R–NH_2 + CH_2{=}CH{-}CN \longrightarrow R–NH–CH_2–CH_2–CN$$

Attack by an electrophilic reagent

Aryl azonium salts react with amino groups. The reaction proceeds rapidly to form a non-colored compound. In weakly alkaline conditions, the reaction of amino groups can become faster than those of tyrosine or histidine. These two last amino acids form colored compounds. Diazonium salts were extensively used for labelling antibodies.

Deamination by nitrous acid is one of the oldest methods of chemical modification of proteins. Although this reagent exhibits both nucleophilic and electrophilic properties; the latter is rarely observed with proteins.

Histidine residues

Histidine does not possess truly selective reagents. The best reagents of histidine in proteins are affinity reagents. Therefore, in conditions of well-determined pH, it is possible to operate a selective labelling with appropriate reagents, in particular diethylpyrocarbonate.

Acylation

Ethoxyformic anhydride or diethylpyrocarbonate reacts with imidazole; the derivative that results presents a characteristic absorption between 230 and 240 nm permitting to follow the reaction. At pH 4, ethoxyformic anhydride reacts principally with amino groups and the imidazole of histidine.

Arylation

Arylation by dinitrofluorobenzene (FDNB) gives an unstable derivative in acidic medium. This modification is very weakly selective.

Alkylation

Alkylation reactions by halo acids and amides at neutral pH represent the best approach to modifying histidines in native proteins. Alkylation by **iodoacetate or bromoacetate** brings about the formation of derivatives substituted in position N_1 or N_3 of imidazole or even bi-substituted compounds. With iodoacetate one obtains either 1-carboxymethyl histidine, 3-carboxymethyl histidine, or even 1,3-dicarboxymethyl histidine. This reaction is rarely selective except in the case of particularly reactive histidine residues. Thus, in ribonuclease, it was possible to label two histidines in the active centre (His 12 and His 119) and by working at distinct pH values to even selectively label either one or the other of these groups (CRESTFIELD et al., 1963). In carbonic anhydrase, the action of iodoacetate brings about a selective labelling of a single catalytic histidine. In these different cases, a single monocarboxymethyl derivative is formed.

Attack by an electrophilic reagent

Diazotetrazole reacts with histidine to give a colored bis-azo derivative that, as we have mentioned above, absorbs with a maximum at 480 nm, which permits proceeding to a spectrophotometric measurement. Similarly, diazosulfanilic acid or Pauly reagent gives a colored compound with histidine, permitting the detection.

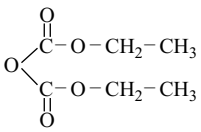
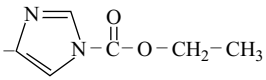
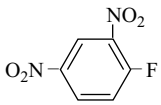
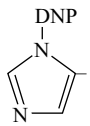
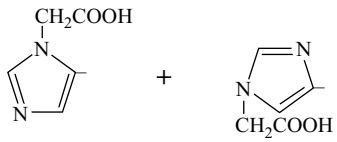
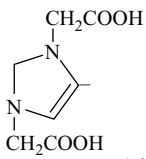
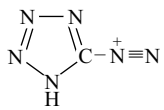
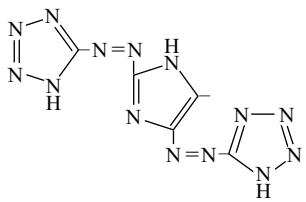
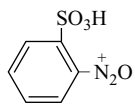
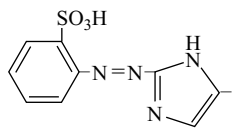
Iodation of histidine occurs generally with that of tyrosine. One can observe the formation of mono and diiodo histidine derivatives.

Photooxidation

Photooxidation of histidine results in irradiation of the protein in the presence of photosensitive compounds like methylene blue or Bengal rose. It is relatively selective in neutral medium and in rather gentle conditions.

Table 9.4 summarises these different reactions of histidine.

Table 9.4 Reactions of the imidazole of histidine

Reagent	*Product*	*Use*
 Ethoxyformic anhydride diethylpyrocarbonate		
 Dinitrofluorobenzene (DNFB)	 DNP-His	
ICH_2COOH Iodoacetic acid	 1-carboxy-methyl-His + 3-carboxy-methyl-His	
	 1,3-dicarboxymethyl-His	
 Diazotretrazole		Histidine colorimetry
 Diazo sulfanilic acid		Detection of histidine (PAULI reagent)
Light + methylene blue	Destruction of imidazole nucleus	Photo-oxydation

Tyrosine residues

The phenol group of tyrosine participates in the same types of reactions as amino groups, but in different conditions.

Acylation

Modification of tyrosines by **acetic anhydride** is reversible; the ester thus formed is unstable in alkaline medium, allowing, as we have previously mentioned, to selectively label amino groups at high pH. Acetyl-imidazole reacts with tyrosines with normal pK, giving an unstable ester in alkaline medium.

Arylation

Cyanuric fluoride (CNF), which presents a very high reactivity but possesses a weak selectivity, is hardly useful. It forms an ester with tyrosine that is stable in alkaline medium and of which the absorption spectrum is shifted towards blue as compared to that of tyrosine.

Fluorodinitrobenzene (FDNB) reacts with tyrosine to give the compound O-dinitrophenyl tyrosine.

Alkylation

Alkylation by **dimethylsulfate,** $SO_4(CH_3)_2$ brings about the formation of O-methyl tyrosine.

Action of electrophilic reagents

Nitration by **tetranitromethane** (TNM) is one of the most selective reactions with tyrosine. At pH 7-8 and with limited reaction times, nitration occurs in *ortho* of the OH group. One does not observe the formation of 3,5-dinitrotyrosine. The o-nitrophenolate ion is stable in acid hydrolysis. It has a maximum absorption at 428 nm, which permits following the reaction by spectrophotometry. The spectrum of absorption of 3-nitrotyrosine is very sensitive to the solvent polarity. The action of the reagent is dependent on the degree of tyrosine exposure. By selective reduction of ortho-nitrotyrosine by dithionite or sodium hydrosulfite, it is possible to convert tyrosine into 3-amino tyrosine. However there are parasitic reactions like the formation of intermolecular bridges, oxidation of cysteine and methionine, and the modification of tryptophan and histidine. Electrophoresis on a polyacrylamide gel in the presence of dodecyl sodium sulfate rapidly permits the detection of possible intermolecular bridges. Despite these inconveniences, the use of tetranitromethane knew much success for selectively labelling tyrosines, which occurs more rapidly than for other residues

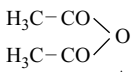
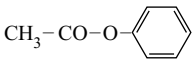
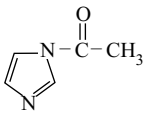
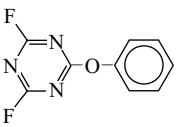
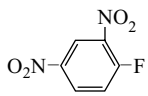
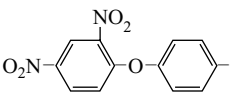
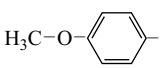
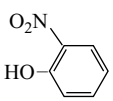
Iodation by iodine or iodic chloride is also a very useful reaction. The labelling occurs in ortho of the phenol group driving mono or diiodo (3-iodo or 3,5-diiodotyrosine) derivatives. Very reactive sulfhydryl groups are occasionally affected forming sulfenyl iodides. Outside of these substitution reactions, oxidative reactions implicating the methionine, cysteine, and tryptophan residues can occur. However, oxidations are predominant in acidic medium whereas substitution reactions are predominant in alkaline medium. Iodation can be brought about by an enzymatic reaction specific to tyrosines, the action of sodium iodide catalysed by peroxydase. Ioda-

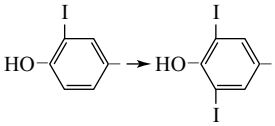
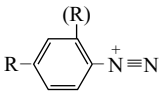
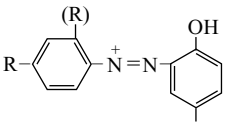
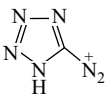
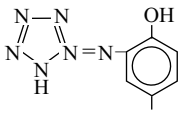
tion of tyrosines is very useful in labelling proteins with ^{125}I, in particular for radioimmunologic assays or for studies of receptors.

Aryl diazonium *salts* react with tyrosine to form derivatives of which the coloration is used in the detection of this amino acid. **Diazonium-1H-tetrazole** (DHT) reacts with tyrosine and also with histidine forming corresponding azoderivatives that absorb at 550 and 480 nm, respectively. The simultaneous spectrophotometric titration of two derivatives is therefore possible without ambiguity. The reaction rate increases with pH. Due to its instability, diazonium-1H-tetrazole is potentially explosive and must be manipulated with care. Concentrations greater than 0.2 M are considered dangerous. ◢

The different reactions of tyrosine are given in Table 9.5.

Table 9.5 Reactions of tyrosine

Reagent	*Product*	*Use*	*Comments*
 Acetic anhydride	 O-acetyl-tyrosine		
 N-acetylimidazole			
 Cyanuric fluoride			
 Dinitrofluorobenzene (DNFB)	 O-dinitrophényl tyrosine		Hardly selective
$SO_4(CH_3)_2$	 O-methyl tyrosine		
$C(NO_2)_4$	 3-NO_2 tyrosine	Specific nitration of tyrosine residues	Yellow 3-NO_2 tyrosine is resistant to HCI hydrolysis

Reagent	*Product*	*Use*	*Comments*
ICl_3 Iodine trichloride INa + lactoperoxydase		Introduction of ^{125}I into proteins	The enzymatic reaction is specific for tyrosines. It permits the selective labelling of cell surfaces proteins.
 Aryl diazonium salts		Detection of tyrosine	
 Diazonium-1H-tetrazole			

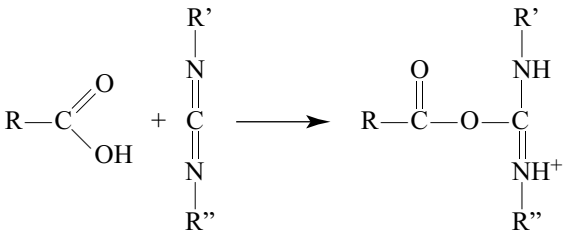
Carboxyl groups

The carboxylate ion is stabilised by resonance and is not very reactive. The reactive form is the protonated carboxyl COOH. The most common reactions used to modify carboxyl groups are, on the one hand esterification catalysed by an acid, and on the other hand coupling with a nucleophilic compound after activation by a carbodiimide soluble in water.

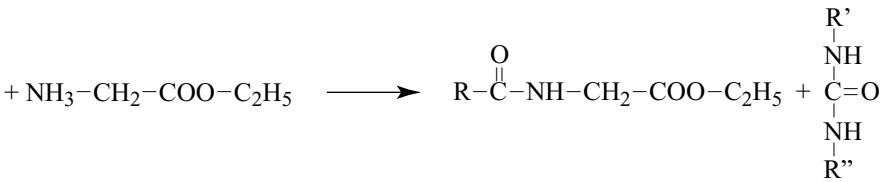
Carboxyl groups of aspartic and glutamic acids have pK values between 4.5 and 5. They can be esterified in relatively gentle conditions, either by methanol in diluted chlorhydric acid, or by diazo acid or diazo alcane derivatives. Diazomethane is a highly reactive compound and was used to esterify carboxylic acids. It is sometimes used to esterify carboxylic protein groups, but the reactions are often incomplete and other groups can also be modified. Diazo-acetate esters are more stable. The modification of SH groups constitutes a parasite reaction (WILCOX, 1967). Esterification can be reversed in alkaline medium.

◤ Carboxylic groups react with diazonium salts, isoxazolium salts, and triethyl oxonium fluoroborate. This last reagent was used to label carboxylic groups of the active centre of lysozyme by PARSON et al. (1969). The ester thus formed is labile, and the reaction is easily reversed around pH 7. Activation of carboxylic groups by N-alkyl-5-phenyl isoxazolium salts was described by the group of SHAW (1969). Therefore N-methyl-5-phenyl isoxazolium fluoroborate was used in the case of trypsin. It brings about complete inactivation of the enzyme; the activity can be recovered by treatment with hydroxylamine.

However, the most utilised reaction for labelling carboxylic residues is coupling with nucleophilic compounds, generally the ester or amide of an amino acid *via* the formation of a soluble carbodiimide intermediate in the presence of which a carboxylic group in aqueous milieu gives an O-acyl-isourea derivative:

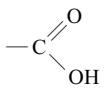
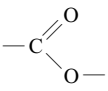
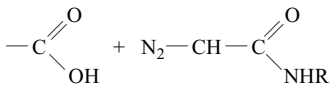
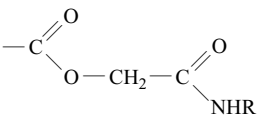
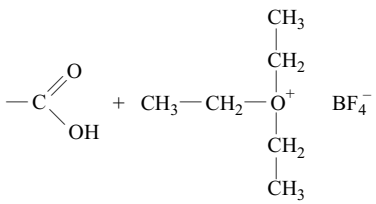
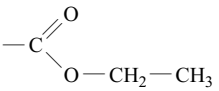
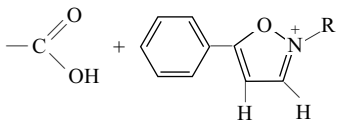
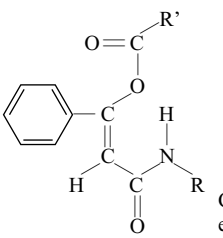
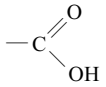


which with ethyl ester glycine gives an amine bond:



However, the carbodiimides are capable of reacting with other groups, in particular cysteines and tyrosines. Labelling of tyrosine can be reversed by treatment with hydroxylamine at pH 7 and 25ºC. These different reactions are regrouped in Table 9.6.

Table 9.6 Reagents of aspartate and glutamate carboxylate

Reagent	*Product*
 + CH_3OH Methanol	
 Diazoacetate	
 Triethyloxonium fluoroborate	
 N-methyl-5-phenyl-oxazoline fluoroborate	 Corresponding enolic ester
 + $R'-N=C=N-R''$ + $NH_2-CH_2-COOC_2H_5$ Carbodiimide + Glycine ethyl ester	$R-CO-NH-CH_2-COO-C_2H_5$

Hydroxyl groups of serine and threonine

Alcohol groups are usually hardly reactive due to their very high pK. Serine can become very reactive in a particular environment, such as in the active centre of serine proteases. For chemical modification one uses pseudo-substrates or affinity reagents, as previously mentioned. When it is reactive, serine can be acylated or phosphorylated. O-phosphoserine is naturally present in some enzymes. Occasionally, it is formed at the active site in an intermediate stage of the catalysed reaction; this is the case with phosphoglucomutase previously described. O-phosphothreonine is also found, although much less frequently.

Sulfhydril groups

Thiol groups of cysteines are very reactive under their form S^-. A great number of compounds are susceptible to react with cysteine; as we have underlined it, the labelling of cysteine often interacts as a parasite reaction and labels other residues at the same time. Since thiol groups are easily modifiable, certain reactions in which they participate are reversible, and their participation at the active site of some enzymes was revealed very early. In this chapter, the diverse reactions of thiols are rapidly reviewed to stress more particularly the reactions the most useful to enzymologists; their use depends of course on the goal that one has set.

◤ *Acylation and arylation*

Thiol groups react rapidly with acid anhydrides to give unstable derivatives in alkaline medium. The reaction of thiols with aromatic compounds such as chlorodinitrobenzene and bromo-nitro-imidazole is done more rapidly than the reactions of acylation, even in more acidic conditions.

Alkylation

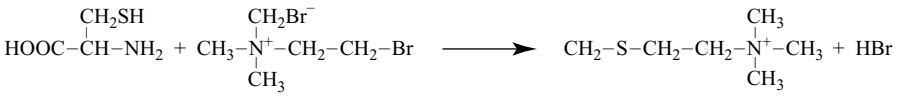
Carboxymethylation by acid halogenides is a very classical modification of thiols that are much more reactive than other residues that can react with these compounds. Reaction with iodoacetate gives rise to negatively charged carboxymethyl cysteine, following the reaction:

$$RS^- + I—CH_2—COO^- \longrightarrow R—S—CH_2—COO^- + I^-$$

Iodoacetamide is also highly used; its reaction with thiols brings about a non-charged carboxamido methyl cysteine derivative. The same reactions can be carried out with corresponding bromine derivatives.

N-ethylenimine reacts with proteins to converting cysteine residues in S(2-aminoethyl) cysteine. The reaction occurs in lightly alkaline conditions (RAFTERY & COLE, 1963). No other group reacts at this pH; in acidic medium, methionine reacts slowly to give the ion S(2-aminoethyl) methionine sulfonium (SCHROEDER *et al.*, 1967). This modification is interesting in the sense that S(2-aminoethyl) cysteine resembles lysine (with S in the place of CH_2), introducing into a protein a new trypsin cleavage site (WANG & CARPENTER, 1968).

Alkylation by **trimethyl ammonium bromide** converts cysteine to a derivative strongly basic, 4-thialamine, according to the reaction:



This trimethylaminoethyl compound is very stable and resists acid hydrolysis, permitting analyses driving the identification of cysteines that reacted.

Azobenzene-2-sulfenyl bromide, soluble in water, reacts selectively and rapidly with protein thiols in aqueous solution. Mixed disulphide thus formed is easily reduced by thiols (2-mercaptoethanol or dithiothreitol) or by sodium borohydride.

Addition reactions

Thiol groups participate in addition reactions with several reagents. Cyanoethylation by **acrylonitrile** is rapid and selective at pH 8; the reaction with amino groups occurs at higher pH. The reaction is carried out as follows:

$$-S^- + CH_2{=}CHCN \longrightarrow -S-CH_2-CH_2-CN$$

The most commonly used reagent is N-ethyl maleimide (NEM) that gives S-ethyl succinimido cysteine which is soluble in water and stable. NEM reacts rapidly with cysteine at neutral pH. The reaction can be followed by the decrease in absorbance at 305 nm. Indeed, NEM absorbs at this wavelength ($\varepsilon_M = 620$) whereas S-ethyl-succinimido-cysteine does not. In addition, S-ethylsuccinimido-cysteine gives by hydrolysis S-succinyl-cysteine that is stable in acidic medium therefore permitting the identification of the labelled cysteines. The reaction of cysteines with maleic anhydride also leads to S-succinyl-cysteine. NEM is a very selective thiol reagent that leads to an irreversible labelling.

The colored derivative **N-dimethylamino-3,5-dinitrophenyl maleimide** (DDPM) introduced by WITTER and TUPPY (1960) was used for the labelling of cysteines of many proteins.

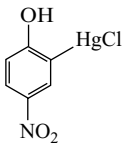
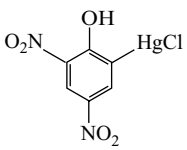
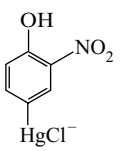
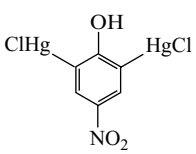
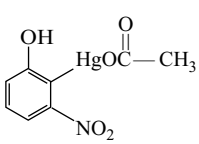
The addition reaction with 4-vinyl pyridine that brings about S-pyridyl ethyl cysteine is often used for identification tasks because the labelling is stable and resists acid.

Reaction with organomercurials

Reaction of thiols with organomercurials brings about the formation of mercaptides. One of the most common reagents is **p-chloro mercuribenzoate** (PCMB). The reaction can be easily followed by spectrophotometry. At pH 4.6, PCMB absorbs with a maximum at 234 nm and an extinction coefficient ($\varepsilon = 1.74 \times 10^4$ mol^{-1} . cm^{-1}). At 255 nm and at the same pH, the value of ε_M between the residues of PMB-cysteine and PCMB is on the order of 6 000–8 000 mol^{-1} . cm^{-1} according to the solvent used.

The mercuri-nitrophenol derivatives have the advantage of absorbing in the visible region of the spectrum (Table 9.7 below). They were frequently used as probes for sulfhydryl group environments in proteins as well as in the preparation of mercury derivatives to analyse proteins by X-ray diffraction.

Table 9.7 Spectral properties of some mercuri-nitrophenols
(from MCMURRAY & TRENTHAM, 1968)

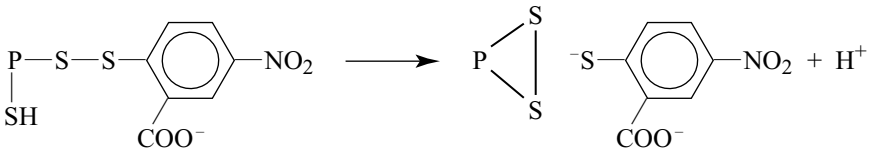
Compound	*Structure*	λ_{max} *(mμ)*	ε *(in 0.1 M KOH)* $(mol^{-1}.cm^{-1})$
2-chloromercuri-4-nitrophenol		405	1.74×10^4
2-chloromercuri-4,6-dinitrophenol		371	1.57×10^4
4-chloromercuri-2-nitrophenol		416	4.1×10^3
2,6-dichloromercuri-4-nitrophenol		410	1.74×10^4
2-acetoxymercuri-3-nitrophenol		410	1.7×10^3 *

* *In 0.1 M NaOH*

Exchange reactions with disulphide groups

One of the most classical reactions with thiol is their ease of exchange with disulphide groups to give mixed disulphides; these reactions are reversible. Introduced by ELLMAN (1959), one of the most familiar reagents is **5,5′-dithiobis (2-nitrobenzoate) or DTNB, or NbS2** in the official nomenclature. The reaction gives rise to a mixed disulphide with the appearance of a thionitrobenzoate ion that can be followed by variation in absorbance at 412 nm. One must take into account a light spontaneous hydrolysis of the reagent. Thionitrobenzoate ion (TNB^-) absorbs at 412 nm ($\varepsilon = 1.36 \times 10^4$ $mol^{-1}.cm^{-1}$ at pH 8), the mixed disulphide enzyme-thionitrobenzoate absorbs at 310 nm with an average extinction coefficient of 2.6×10^3 $mol^{-1}.cm^{-1}$, whereas DTNB presents a peak at 323 nm ($\varepsilon = 1.66 \times 10^4$ $mol^{-1}.cm^{-1}$). However,

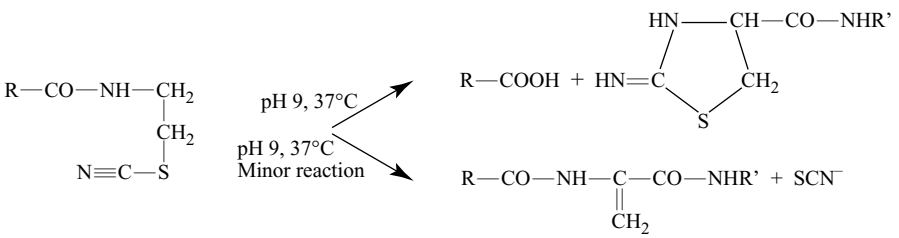
in proteins the presence of other thiols within the proximity of mixed disulphide can bring about the appearance of a second TNB^- ion with the formation of an internal disulphide bridge, bringing about an error in the titration:



- **4,4′-dithiopyridine** reacts with SH groups in a manner analogous to DTNB releasing one molecule of thiopyridine that can be measured by spectrophotometry. The reagent absorbs with a maximum at 247 nM ($\varepsilon_{247} = 1.63 \times 10^4$ mol^{-1} . cm^{-1}) and 4-thiopyridone at 324 nm ($\varepsilon_{324} = 1.98 \times 10^4$ mol^{-1} . cm^{-1}). There are other reagents of the same type; only the most common are presented here.

Cyanylation

Cyanylation of SH can be carried out by reaction with **2-nitro-5-thiocyanobenzoate** (NTCB) introduced by DEGANI et al. (1970), and DEGANI and PATCHORNIK (1974). Thiocyanoalanine is obtained. The substitution which is very specific leads to a derivative without steric hindrance; that is not the case for other thiol reagents. The reagent is obtained by treatment of DTNB by cyanogene bromide at pH 8. The reaction with thiols is followed by the variation in absorbance at 412 nm due to the thionitrobenzoate ion. JACOBSON et al. (1973) proposed a method using this reagent to obtain a selective cleavage of peptide bonds at the level of cysteines. The reaction is the following:



The cleavage is not carried out to 100%; a weak proportion is converted to a non-cleaved peptide that contains dihydroalanine.

Oxidation reactions

The action of an oxidising agent such as hydrogen peroxide or performic acid brings about the appearance of cysteic acid. Sodium sulfite transforms the cyteine residues into S-sulfonate derivatives (R–S–SO_3^-).

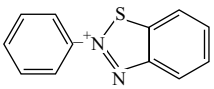
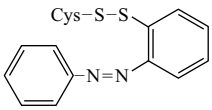
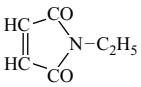
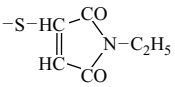
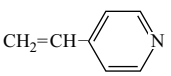
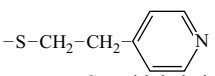
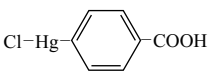
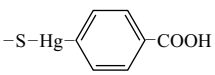
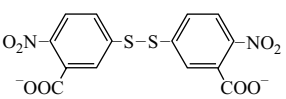
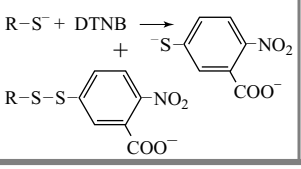
Tetrathionate gives a sulfonate S-sulfenyl derivative following the reversible reaction:

$$P\text{–}SH + SO_4O_6^{=} \longrightarrow P\text{–}S\text{–}S\text{–}SO_3^- + S_2O_3^{=} + H^+$$

often used to protect cysteines during chemical modification of other residues.

Table 9.8 below summarises the principal reactions of SH groups in proteins.

Table 9.8 Reactions of cysteine and cystine

Reagent	*Product*	*Use*	*Comments*
ICH_2COOH Iodoacetic acid	$-S-CH_2COOH$ Carboxymethyl cysteine (CM-Cys)	Substitution with introduction of a négative charge	CM-cys is resistant to acid hydrolysis
ICH_2CONH_2 Iodoacetamide	$-S-CH_2-CONH_2$	Introduction of a positive charge	Amino-ethyl cysteine is recognised by trypsine
H_2C(NH)H_2C Ethylene diamine	$-S-CH_2-CH_2-NH_2$		
H_2C(O)H_2C Ethylene oxide	$-S-CH_2-CH_2-OH$	Stabilisation without changing the charge	
$CH_3-N^+(CH_2-Br)(CH_3)-CH_2-CH_2-Br$ Trimethyl ammonium (2-bromoethyl) bromide	$-S-CH_2-CH_2-N^+(CH_3)_3$ 4-thiolaminine	Introduction of a positive charge	Very stable during HCl hydrolysis
 Azobenzene-2-sulfenylbromide		ε = 1,000 à 353 nm	
$CH_2=CH-C\equiv N$ Acrylonitrile	$Cys-S-CH_2=CH_2-CN$		
 N-ethylmaleimide (NEM)	 S-ethylsuccinimido cysteine	Spectrophotometric titration	
 Maleic anhydride	$Cys-S-CH(CH_2-CO)-CO-N-C_2H_5$ S-succinyl-Cys		
 4-vinyl pyridine	 S-pyridylethyl cysteine	Stabilisation and introduction of absorption in UV	The product is resistant to HCl hydrolysis
 p-chloromercuribenzoic acid (PCMB)		Formation of mercaptides	Cysteine can be regenerated by reduction
$R_1-S-S-R_1$ R_1 disulfide	$R-SH + R_1-S-S-R_1$ $\rightleftharpoons R-S-S-R_1$ (mixed disulfide) $\rightleftharpoons R-S-S-R$ (R disulfide)		
 5,5'-dithiobis (2-nitrobenzoate) (DTNB)		Thiols titration (ELLMAN reagent)	

Methionine

Methionine participates in numerous reactions with other amino acids. However, one can still obtain relatively selective labelling by placing it in acidic media (Table 9.9). Thus, alkylation by iodoacetate and iodoacetamide is carried out fairly selectively at low pH.

Table 9.9 Reactions of methionine

Reagent	*Product*	*Use*	*Comments*
$I-CH_2-COOH$ Iodoacetic acid	$-\overset{+}{S}(CH_3)-CH_2-COOH$	Substitution with introduction of a negative charge	
$I-CH_2-CONH_2$ Iodoacetamide	$-\overset{+}{S}(CH_3)-CH_2-CONH_2$	Substitution without changing charge	
Oxidation	$-S(=O)-CH_3$ Met-sulfoxide		Unstable in acid hydrolysis
Photo-oxidation	$-S(=O)_2-CH_3$ Met-sulfone (Met-SO_2)		Met-SO_2 is stable in acid hydrolysis
$BrC\equiv N$ Cyanogen bromide	$-AA_1-Met-AA_2-$ → $AA_1-HSL + H_2N-AA_2$ HSL = homoserine lactone → $-NH-CH(CH_2-CH_2-OH)-COOH$ Homoserine	Specific cleavage at the level of the methionine carboxyl	Homoserine lactone (HSL) gives homoserine by hydrolysis

◤ Oxidation by hydrogen peroxide occurs easily in acidic media; this yields a methionine sulfoxide. Photooxidation in acidic media or in the presence of a photosensitiser gives methionine sulfones. ◢

The reaction of methionine with cyanogene bromide permits the selective cleavage of the polypeptide chain at the level of the methionine carboxyl which is transformed into homoserine lactone, which by hydrolysis produces homoserine.

Arginine

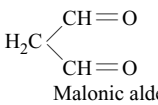
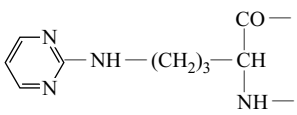
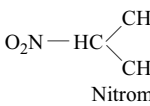
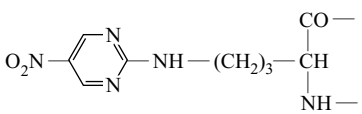
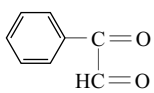
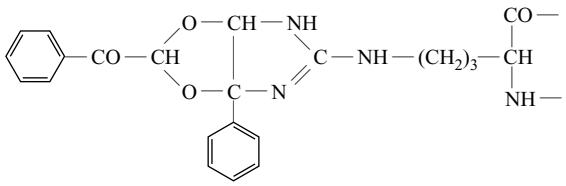
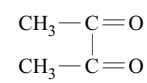
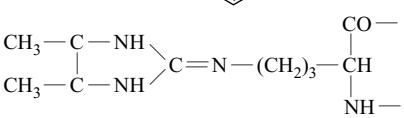
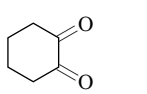
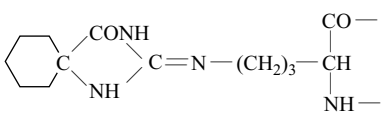
The arginyl residue possesses a high pK and is stabilised by resonance; it differs from other protein basic groups. The chemical modification of this residue is therefore very difficult. Acylation and nitration can only be carried out in extreme conditions, incompatible with protein stability. The only type of reaction practically used is condensation with dicarboxyl compounds that is still carried out at extreme pH values.

- Among the most useful reagents, one must cite **malonic or nitromalonic dialdehyde.** With nitromalonic aldehyde, a nitropyrimidine derivative is obtained; however the reaction is carried out at high pH (between pH 12 and 14), therefore only on denatured proteins.

  One of the very utilised reagents of arginine is **phenylglyoxal** introduced by TAKAHASHI (1968) that reacts rapidly at less extreme pH (pH 8). The derivative obtained is stable in acidic media and labile in alkaline media. Among the other arginine reagents, butanedione, its dimer or its trimer were used at pH values on the order of 8 to modify proteins. **1,2-cyclohexane dione** is used at higher pH.

  Table 9.10 indicates some of the reactions with arginine.

Table 9.10 Reactions of arginine

Reagent	*Product*	*Use*
 Malonic aldehyde		Arginine modified by this reactions is no longer accessible to trypsin.
 Nitromalonic dialdehyde		
 Phenylglyoxal		
 Butanedione		
 Cyclohexane dione		

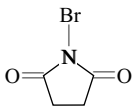
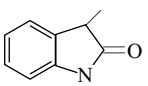
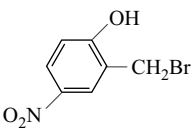
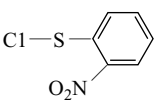
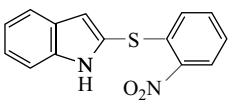
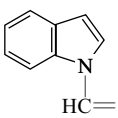
Tryptophan

Most of the reagents utilised to modify tryptophan react also with thiol groups which therefore must be protected beforehand. Oxidation of the indole nucleus is frequently used.

- Oxidation can be obtained by action of hydrogen peroxide, by ozonolysis in performic media in the presence of resorcinol; however this last reaction also modifies methionines. **N-bromosuccinimide** reacts rapidly with tryptophan to give an oxindole derivative that absorbs at 250 nm; although very utilised, this reaction is hardly selective. **2(2-nitrophenyl-sulfenyl)-3-methyl-3-bromoindolenine** (BNPS skatole) replaced N-bromosuccinimide; this compound also converts tryptophan into an oxindole derivative; it also reacts with methionine to give the corresponding sulfoxide derivative.

  Certain nitrobenzyl halogenides are very reactive towards tryptophan. **2-hydroxy-5-nitrobenzene bromide** is capable of alkylating tryptophan in gentle enough conditions. The compound formed has an absorption maximum at 410 nm that permits following its appearance. This reagent gives with methionine a derivative that spontaneously decomposes. Sulfenic halogenides such as **2-nitro or 2,4-dinitro-sulfenyl chloride** react in acidic media (pH < 3.5); they form compounds having a characteristic spectrum of absorption (Tab. 9.11). This reagent is also selective with tryptophans when the SH groups have been protected.

Table 9.11 Reactions of tryptophan

Reagent	*Product*	*Use*	*Comments*
Oxidation by H_2O_2 Light + methylene blue	Destruction of the indole nucleus	Photo-oxidation	
 N-bromosuccinimide	 Oxyindole		Generally brings about the cleavage of the peptide bond
 2-hydroxy-5-nitrobenzyl bromide	Complex reaction	High specific substitution for tryptophan	
 2-nitrophenylsulfenyl chloride (NPSCI)		Specific substitution of tryptophan in the absence of SH	
HCOOH + HCl Formic acid		Reversible substitution of tryptophan	

Another interesting modification is the formylation of tryptophan by formic acid in the presence of HCl; this modification is reversible in alkaline media. ◢

9.2.3. CRITERIA USED TO INTERPRET RESULTS

In order to determine if an amino acid residue is essential to the catalytic activity, it is necessary to define a certain number of criteria permitting the evaluation of the importance and role of a modified residue. **The diverse criteria used have their limits and one must not be satisfied with just one.** The principal criteria are stoichiometric inactivation, protection by the substrate, criteria following from kinetic analysis and of course, when it is possible to delete the chemical modification, the return of the enzymatic activity.

9.2.3.1. STOICHIOMETRIC INACTIVATION

When one modifies a residue essential to the active centre of an enzyme, one must observe a total inactivation for a reagent molecule bound to the active site; the inactivation must be stoichiometric. It will be convenient therefore to determine first of all the number of reagent molecules incorporated into the active site. In addition, it must be assured that there is a total loss of activity. Indeed, it is frequently observed that the labelling of a residue brings about a loss of activity down to a very weak value, sometimes less than 1%; in this case, the labelling of an essential residue cannot be concluded. To verify this, it is necessary to carry out a measure of activity with concentrations of the labelled enzyme two to three orders of magnitude greater than those used with the active enzyme.

If the loss of activity is not complete, the labelled group is not a catalytic group. It can be implicated in the ligand binding, the labelling bringing about an inactivation resulting from steric hindrance. The loss of activity can also result from a conformational variation, even weak, induced by the chemical modification. The binding of the reagent can occur on a non-catalytic residue located in the proximity of the catalytic residue in the three-dimensional structure, but which can be very distant in the sequence, rendering the interpretation difficult when the three-dimensional structure is not known.

9.2.3.2. SPECIFIC PROTECTION AGAINST INACTIVATION

Another criteria used correlatively with the preceding is protection by a competitive inhibitor or a mildly reactive substrate that blocks the entry to the active site. If the reagent labels a catalytic residue, one must observe a decrease in the labelling rate. However this criteria is very relative. The presence of the inhibitor will bring about a comparable effect if it concerns the labelling of groups neighboring the catalytic residue.

9.2.3.3. KINETIC ANALYSIS OF RESULTS

An indispensable and much more reliable criteria is that which results from the kinetic analysis of labelling and the loss of activity for each individual group.

Kinetic analysis for pseudo-first order reactions

Diagnostic rules permitting the determination of the role of a group labelled by a reagent were proposed by RAY and KOSHLAND (1961) in the case of pseudo-first order reactions, when the enzyme is treated by a large excess of reagent. The method developed consists of comparing the rate constants of the incorporation of the reagent and of the decrease in enzymatic activity. To illustrate the principle of analysis, the authors gave the example of a protein that possesses two histidine residues equally reactive towards a chemical compound. If the reaction is carried out to 50% of labelling, the distribution of molecules is that presented in Fig. 9.24; that is 25% of the molecules are not labelled by the reagent, 25% are labelled on residue 1, 25% on residue 2 and 25% on the two residues. Following the role of each of these two residues in the catalytic activity, the results observed will be different. If residue number 1 is essential, but not residue number 2, only the molecules E_1 and $E_{1,2}$ will be inactive. One will observe a 50% loss of activity; likewise if only residue 2 is essential. If the two residues are essential, the loss of activity will be 75%. If one of the two residues can substitute for the other, the loss of activity will only be 25%.

Fig. 9.24 Distribution of species having labelled histidine residues

In the case where there is a modification of a single essential group, the kinetics of labelling must correspond to the kinetics of activity loss. Indeed, if the protein is composed of diverse reactive groups, X_1, X_2, … X_n, the residues being modified with pseudo-first order constants k_1, k_2, … k_n, the fractions of residues not yet modified at time t will be: $x_1/x_{1,0}$, $x_2/x_{2,0}$, … $x_n/x_{n,0}$, respectively. The kinetics of labelling are given by the following equations:

$$x_1/x_{1,0} = e^{-k}1^t \quad ; \quad x_2/x_{2,0} = e^{-k}2^t \quad ; \quad \ldots\ x_n/x_{n,0} = e^{-k}n^t$$

The kinetic of disappearance of enzymatic activity is of the form:

$$A/A_0 = e^{-k_A t}$$

A being the activity remaining at time t, A_0 the initial activity, and k_A the rate constant. If a residue X_1 has been labelled with a constant k_1 and if $k_A = k_1$, the labelled residue is essential. If the two residues X_1 and X_2 have been labelled and if $k_A = k_1 + k_2$, the two residues are essential; the labelling of one or the other brings about a loss of activity. If the residue X_1 has reacted with the constant k_1 and if $k_A > k_1$, this residue is not a catalytic group (Fig. 9.25 below).

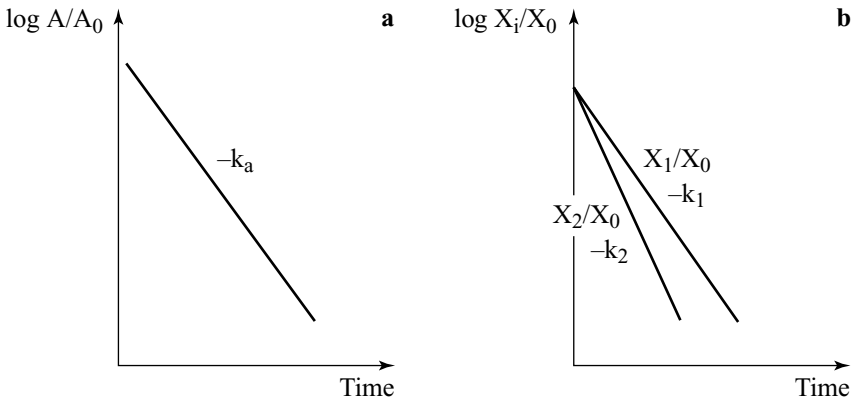
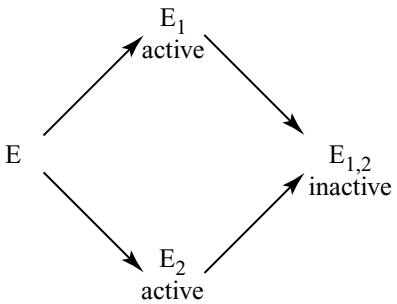


Fig. 9.25 Kinetics of residue labelling
*(**a**) kinetics of disappearance of enzymatic activity – (**b**) pseudo-first order kinetics for labelling of residues X_1 and X_2.*

RAY and KOSHLAND also considered two other cases, first that in which the enzyme would involve two residues X_1 and X_2 that can interfere in a complementary manner in the catalysis; this is the *either/or* (one or the other) case symbolised by the scheme:



The loss of activity as a function of time is then reliant on the rate of disappearance of groups X_1 and X_2 by the relationship:

$$A/A_0 = e^{-k_A t} = e^{-k_1 t} + e^{-k_2 t} + e^{-(k_1+k_2)\,t}$$

It seems very improbable that this case occurs at the active centre of enzymes.

The other case, more realistic, is that of the progressive denaturation of the enzyme bringing about the labelling of a non-essential residue:

$$\underset{\text{active}}{E} \xrightarrow{k_1} \underset{\text{active}}{E_1} \xrightarrow{k_2} \underset{\text{inactive}}{E_{1,2}}$$

In this case, the modification of the non-essential residue X_1 does not affect the activity of the enzyme but its conformation, so that the residue X_2 that is essential and non-reactive in the initial form becomes reactive in the form E_1. The kinetics of the reaction are:

$$A/A_0 = [k'_2/(k'_2 - k_1)]e^{-k_1 t} - [k_1/(k'_2 - k_1)]e^{-k_2 t}$$

Analysis of results in the most general case

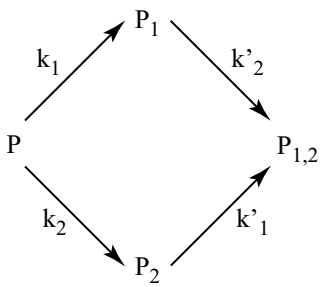
The analysis of RAY and KOSHLAND constitutes an important approach; however, it is limited to chemical modifications that are performed according to a pseudo-first order reaction, meaning in the presence of a large excess of reagent that constitutes a limitation. These conditions are present in a great number of cases, but for the others they are difficult to perform and control. TSOU (1962) proposed a more general treatment permitting the achievement of an analysis which applies whatever concentration of the reagent relative to titratable groups; this treatment applies to all cases and presents therefore a great interest when the treatment of RAY and KOSHLAND cannot apply.

TSOU considers different cases that are practically the same as those treated by RAY and KOSHLAND; he subdivides them following the relative reactivities of diverse categories of groups and the distribution of essential residues in these categories. He introduces a distinction based on the nature of the residues (His, Tyr, Lys etc.). The kinetic analysis however does not distinguish the nature of the groups, but only their reaction rate. TSOU does not make any hypothesis on the order of the reactions and considers, following the different cases treated, the variations in the residual activity fraction as a function of the fraction of non-modified essential groups. The case I which is the most classical relates to a single essential group; all the groups of the same type react with the same rate. If the number of groups essential to the activity is i, the activity fraction staying a, and the fraction of non-modified essential groups x_e, the following relationship must be verified:

$$a^{\frac{1}{i}} = x_e$$

The analysis consists of representing either log a as a function of log x, or a, $\sqrt{a}$, $\sqrt[3]{a}$, as a function of x until one obtains a linear graph.

◤ A generalisation of the treatment of TSOU was proposed by TENU and YON (unpublished results) of which the essential is described below. Different possibilities derive from a general model difficult to resolve in the case of n groups, but easy to represent and to resolve in the case of two residues. The scheme can be thus represented:



P_1 and P_2 represent the labelled species on the residues 1 and 2 respectively. In $P_{1,2}$, the two residues 1 and 2 are blocked by the reagent. One can be in the presence of two types of reactions, either a random mechanism with interactions ($k_1 \neq k'_1$; $k_2 \neq k'_2$), or without interactions ($k_1 = k'_1$; $k_2 = k'_2$), or an ordered mechanism ($k_2 = 0$ or $k_2 = k'_1 = 0$).

Random mechanisms

In the most general case of n residues, if x_i is the probability of the residue i being intact and y_i the probability that it is modified, these values are related by the expression:

$$x_i + y_i = 1$$

The number of species blocked on i residues is given by the relationship:

$$C_n^i = \frac{n!}{i!(n-i)!}$$

When all the rate constants are equal, the probabilities of formation of diverse species are:

$$x^n,\ C_n x^{n-1} y,\ \ldots\ C_n^i x^{n-1} y_i,\ \ldots$$

and:

$$(x+y)^n = x^n + \sum_{i=1}^{i=n} C_n^i x^{n-1} y_i$$

The chemical modification produces 2^n species. The conserved activity fraction corresponds with the sum of probabilities of diverse active species. Aside from this formulation one can find again different cases treated by RAY and KOSHLAND, and by TSOU.

Let us consider first the modification of essential residues, diverse residues reacting with whatever rates. This situation corresponds with the first case of RAY and KOSHLAND and with the third case of TSOU. One can show that:

$$a = \prod_i x_i^{b_i}$$

n_i being the number of residues modified with the constant k_i, x_i the fraction of intact residues n_i, b_i the number of essential residues among the n_i and r the number of residues blocked per enzyme molecule.

The residues are modified with the rate:

$$\frac{dx_i}{dt} = -k_i R^i x_i \quad ; \quad \frac{dx_j}{dt} = -k_j R^j x_j$$

R being the concentration of the reagent. The pseudo-first order conditions are fulfilled for i = j = 1. In the general case, one has:

$$\frac{\dfrac{dx_i}{x_i}}{\dfrac{dx_j}{x_j}} = \frac{k_i R^i}{k_j R^j}$$

There is a relationship between r and a:

$$r = \sum n_i - \sum n_i a^{k_i / \sum b_i k_i}$$

An interesting piece of information is given by the analysis of tangents to the origin; for a = 1, r = 0 of the curve a = f[r]. In the general case:

$$\left(\frac{dr}{da}\right)_{a=1} = -\frac{\sum n_i k_i}{\sum b_i k_i}$$

$$\left(\frac{da}{dr}\right)_{r=0} = -\frac{\sum b_i k_i}{\sum n_i k_i}$$

If the ratio $r/\Sigma n_i$ is not known with accuracy because there is an imprecision in a parameter (extinction coefficient of the protein or specific radioactivity of the reagent), it is nevertheless possible to determine the number of residues implicated Σb_i. Indeed, the representation $a^{\alpha} = f[r]$ is linear if $\alpha = 1/\Sigma b_i$ is independent of adopted parameters.

When all the residues react with the same rate $k_i = k_1$ and the relationship becomes:

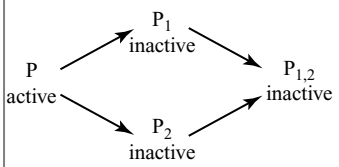
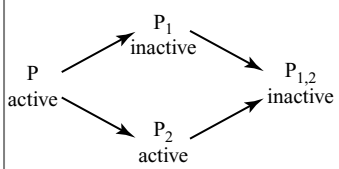
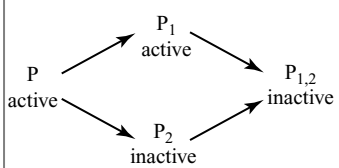
$$r = \sum n_i - \sum n_i a^{1/\sum b_i}$$

which can also be written: $$a^{1/\sum b_i} = 1 - r/\sum n_i$$

This is the first case treated by TSOU.

To illustrate this analysis, the simplest case of two residues is detailed below. One assumes that the two residues react with the rate constants k_1 and k_2 such that $k_1 < k_2$. It concerns a random mechanism without interactions between the reacting residues. In this type of mechanism, three possibilities must be considered, that the two residues are essential to the activity or only one, either residue 1 or residue 2 (Table 9.12).

Table 9.12 Case I: case I can lead to three possibilities

		Residue 1	*Residue 2*	*Activity*	
I.1	The two residues are implicated in the activity.	1 1 0 0	1 0 1 0	1 0 0 0	
I.2	Residue 1 is rapidly modified and is implicated in the activity.	1 1 0 0	1 0 1 0	1 1 0 0	
I.3	Residue 2 is slowly modified and is implicated in the activity.	1 1 0 0	1 0 1 0	1 0 1 0	

In the first case, one can show that:

$$r = 2 - a^{k_1/(k_1+k_2)} - a^{k_2/(k_1+k_2)}$$

if $k_1 = k_2$, the expression becomes: $a = (1 - r/2)^2$

r being the number of residues modified and a the residual activity. This expression corresponds with the first case of TSOU.

When residue 2 which reacts rapidly is the only group essential to the activity, the relationship between r and a becomes:

$$r = 2 - a - a^{k_1/k_2}$$

On the contrary, if the essential residue is that which reacts the most slowly, one has the following expression:

$$r = 2 - a - a^{k_2/k_1}$$

If $k_1 = k_2$, the two preceding expressions simplify into:

$$r = 2 - 2a$$

Figure 9.26 represents the variations in a as a function of r in these diverse cases and gives the value of the slope of initial and final tangents of the curve.

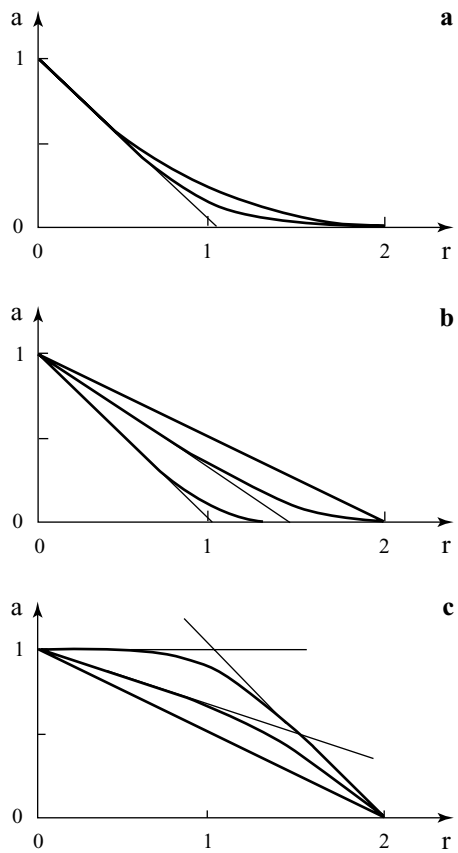


Fig. 9.26 Variations in a as a function of r in the random mechanism of labelling two residues
*Figure (**a**), (**b**) and (**c**) correspond respectively to the three cases of Table 9.12; the lines of Fig. (**b**) and (**c**) correspond to the case where $k_1 = k_2$*

In the case where the modification of all the essential groups is required for the loss of activity, which corresponds to the *either/or* case of RAY and KOSHLAND, the expression of a as a function of r becomes:

$$a = 1 - \frac{r^2}{4}$$

Ordered mechanisms

This type of mechanism can be symbolised by the following scheme:

$$\underset{\text{active}}{E} \xrightarrow{k_1} \underset{\text{active}}{E_1} \xrightarrow{k'_2} \underset{\text{inactive}}{E_{1,2}}$$

which is identical to the case of progressive denaturation treated by RAY and KOSHLAND in which the species $E_{1,2}$ is inactive. Yet the mathematical treatment is possible whatever the order of the reaction is. One defines two relations, one giving the expression of a, the other giving the expression of r. These expressions do not depend on the concentration of the reagent, but on the enzyme fraction (E/E_t) which reacted:

$$a = \frac{k_1}{k_1 - k'_2}\left[\left(\frac{E}{E_t}\right)^{k'_2/k_1} - \frac{k'_2}{k_1} \times \left(\frac{E}{E_t}\right)\right]$$

$$r = 2 + \left(\frac{2k'_2 - k_1}{k_1 - k'_2}\right) \times \frac{E}{E_t} - \frac{k_1}{k_1 - k'_2} \times \left(\frac{E}{E_t}\right)^{k'_2/k_1}$$

In these expressions, it is not possible to eliminate (E/E_t). Nevertheless the test of the slope of the tangents to the curve $a = f(r)$ brings about important conclusions. The tangent of the curve da/dr is given by the equation:

$$\frac{da}{dr} = \frac{k'_2\left[\left(\frac{E}{E_t}\right)^{(k'_2 - k_1)/k_1} - 1\right]}{2k'_2 - 1 - k'_2\left(\frac{E}{E_t}\right)^{(k'_2 - k_1)}}$$

This slope takes characteristic values for $r = 0$, when no residue has been blocked by the reagent, and for $r = 2$, when two residues are labelled.

If $k'_2 > k_1$, one has $\left(\frac{da}{dr}\right)_{r=0} = 0$ and $\left(\frac{da}{dr}\right)_{r=2} = -\frac{k'_2 - k_1}{2k'_2}$

If $k'_2 < k_1$ $\left(\frac{da}{dr}\right)_{r=0} = 0$ and $\left(\frac{da}{dr}\right)_{r=2} = -1$

Table 9.13 below summarises all these different cases.

Table 9.13 Compared analysis of the different cases for n = 2
(from TENU, Thèse Orsay, 1978)

Case	*Function a = f[r] or r = g[a]*	$\left(\frac{da}{dr}\right)_{r=0}$	$\left(\frac{da}{dr}\right)_{r=2}$
Random: I.1 Residues 1 and 2 are essential	$r = 2 - a^{k_1/(k_1+k_2)} - a^{k_2/(k_1+k_2)}$ $a = \left(1 - \frac{r}{2}\right)^2$ if $k_1 = k_2$	-1	0
Random: I.2 Residue 2 is blocked more rapidly than residue 1 and is essential	$r = 2 - a - a^{k_1/k_2}$ $a = 1 - \frac{r}{2}$ if $k_1 = k_2$	$-\frac{k_2}{k_1+k_2}$	0
Random: I.3 Residue 1 is blocked more slowly than residue 2 and is essential	$r = 2 - a - a^{k_2/k_1}$ $a = 1 - \frac{r}{2}$ if $k_1 = k_2$	$-\frac{k_1}{k_1+k_2}$	-1
Either/or *contribution of two residues:* The activity is intact when the two residues are not blocked	$a = 1 - \frac{r^2}{4}$ if $k_1 = k_2$	0	-1
Sequential scheme: $k'_2 > k_1$		0	$-\frac{k'_2 - k_1}{2k'_2}$
Sequential scheme: $k'_2 \leq k_1$		0	-1

For example, Fig. 9.27 is a representation of a as a function of r for the modification of histidine residues by diethyl pyrocarbonate in pyruvate kinase according to BORNMAN et al. (1974).

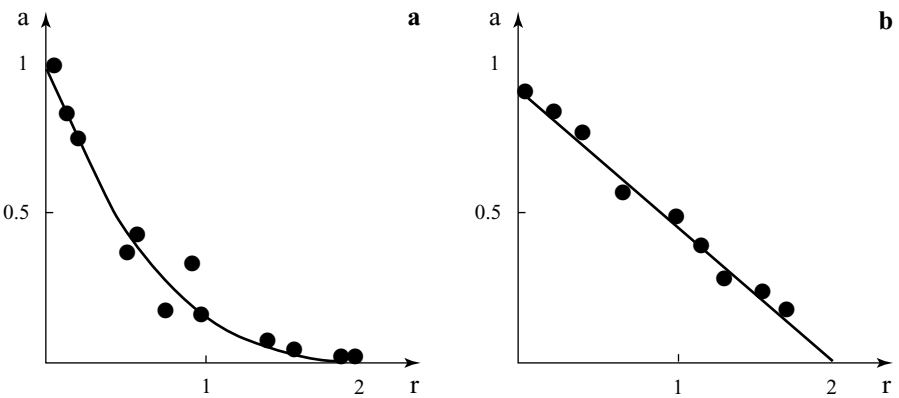


Fig. 9.27 Examples of analysis of results from chemical labelling of lysine residues of glucose-6-phosphate dehydrogenase (a) and of histidine residues of pyruvate kinase (b)

The results indicate that an essential histidine was modified by the reagent. In the case of glucose-6-phosphate dehydrogenase, two lysine residues were modified upon binding of pyridoxal phosphate (from MILHAUSEN & LEVY, 1975). The representation a = f[r] gives a curve, the tangent at the origin $(da/dr)_{r=0} = -1$ and $(da/dr)_{r=2}$ is null. This situation corresponds with the first case where the two residues are essential for the activity. The authors meanwhile had concluded a single essential residue since, when a residue is modified, more than 90% of the enzyme activity is lost. This example illustrates well the importance of rigorous analysis in the determination of residues essential to the activity of an enzyme. ◢

9.2.3.4. REVERSIBILITY OF THE CHEMICAL MODIFICATION AND THE LOSS OF ACTIVITY

In the case where the chemical modification can be reversed or more generally in the case where the reagent blocking an essential residue can be eliminated, one can recover the initial enzymatic activity. For example, in the case of serine acetylation by paranitrophenyl acetate in serine proteases such as α-chymotrypsin or trypsin, the chemical modification is stable in acidic media, but the increase in pH provokes the departure of the acetyl group; there is deacylation and the enzyme activity is restored. In the case of thiol proteases, the reversible labelling of the essential cysteine is possible and the enzyme recovers its activity after elimination of the chemical reagent.

However, the elimination of the chemical reagent employed to label a catalytic residue is not always possible. When it is, the criteria for reversibility constitutes an important element in the interpretation of the chemical modification.

9.3. USE OF MUTAGENESIS METHODS TO STUDY THE ACTIVE CENTRE OF ENZYMES

Introduced later, genetic methods, in particular site-directed mutagenesis techniques, offer another way of approaching the determination of groups essential to the enzyme activity. Site-directed mutagenesis techniques arose from the meeting of an American researcher, HUTCHINSON, and a Canadian researcher SMITH who were carrying out a year of sabbatical in the laboratory of SANGER at Cambridge at the moment where the sequencing of DNA was coming into focus (1977). These techniques permit the replacement of any amino acid of a protein by another and are perfectly adapted to the determination of amino acids implicated in enzymatic catalysis.

It is not our purpose to expand on the different methods of mutagenesis that are classical today and we redirect the reader to the specialised work: *Directed mutagenesis: a practical approach* (edited by MC PHERSON, Oxford University Press). We simply give the principle of one of these methods.

9.3.1. METHODOLOGY

The methodology is well developed today and represents a current technique concerning the modification of DNA. There exist different methods for replacing by genetic methods a determined amino acid in a protein. Figure 9.28 illustrates the principle of one of these methods, which utilises oligonucleotide synthesis.

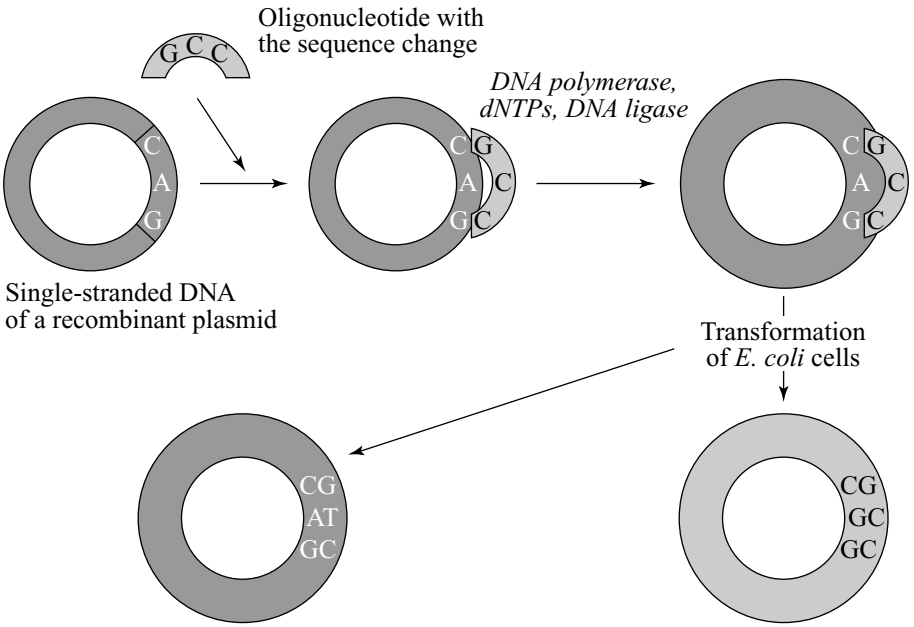


Fig. 9.28 Principle of the directed mutagenesis technique

- The DNA fragment corresponding to the protein gene that one wishes to mutate is cloned on a bacterial vector, generally a plasmid that permits obtaining a single-stranded form.
- The DNA fragment sequence is established and one defines the modifications that one wishes to introduce.
- The complementary oligonucleotide of the chosen sequence including the modification is synthesised by chemical methods.
- The oligonucleotide can be hybridised in vitro with a single-stranded vector. The hybrid is stable at a temperature less than the transition temperature T_m.
- DNA polymerase I recognises this hybrid as a substrate for replication in vitro, the oligonucleotide playing the role of primer and the single-stranded vector that of the matrix (template).
- This DNA is introduced into cells for transformation (*E. coli*, yeast…). It is replicated and gives rise to a mixed population, one part of the vectors carrying the

mutation, and not the other. Over the course of generations, there occurs a segregation.

- There remains selecting the colonies of mutant cells from colonies of wild type cells. One provides a simple method based on the following property: the hybrid formed between mutated DNA and the oligonucleotide has a greater thermostability than that formed by wild type DNA and the oligonucleotide.

There exist variations of this technique. An important condition for the application of the method is that the gene corresponding to the mutated protein expresses correctly and that the protein is not degradated by proteases. Although the modification produced in the DNA constitutes a well tested technique, the conditions of expression and degradation occasionally pose serious problems.

There currently exist variable genetic methods that are the object of numerous works. Aside from rational methods like site-directed mutagenesis, other methods proceed in a different manner like random mutagenesis and the techniques of directed evolution with selection of mutants as a function of their properties or the absence of them and further analysis of the mutation. These techniques however are essentially utilised for the remodelling of proteins in order to generate new properties. The selection is performed from bound ligands and uses methods like presentation on phage (phage display), or on ribosomes (ribosome display) or on RNA (RNA display). These methods are the subject of numerous reviews (TOBIN et al., 2000; O'NEIL & HOESS, 1995; HANES & PLÜKTHUN, 1997; WILSON & SZOSTAK, 1998). ◢

9.3.2. STRATEGY

The strategy used involves a preliminary, the choice of the protein modification. This is evidently dictated by the question which one wants to answer. If it concerns understanding the role of a group of a side chain in the catalysis or in the binding of the substrate, it will be convenient to consider first of all the three-dimensional structure of the enzyme studied and the position of the substrate or of an analog at the active centre. This strategy is equally valid for using chemical labelling methods when the structural data is available. All rational mutation implies knowledge of the structure. According to the position of the most likely implicated amino acids, the use of molecular modelling and energy minimisation methods will help to conceive the most sensible replacement. An important point to consider is that of internal dynamics of the protein. Indeed, the substitution of an amino acid can have important structural consequences in a region of the protein far from the mutation site. The use of genetic methods for determining catalytic groups of enzymes calls however some reservations. They do not permit like chemical labelling methods achieving a kinetic analysis of the results. In addition, the loss of activity engendered by the mutation can result in a secondary effect and does not necessarily indicate that the modified group is directly implicated in the catalysis.

9.3.3. SOME EXAMPLES

Since the introduction of the method, the work using directed mutagenesis to determine the groups essential to the activity of enzymes has rapidly increased. Also there is no question in this chapter of making an exhaustive review. Only some examples are given here in order to illustrate the potentials of the method. Other applications will be given relevant to particular enzymatic systems in Part IV.

β-lactamase possesses a serine in position 70 and a threonine in position 71 that are conserved in all species. The mutation Ser70 ⟶ Thr brings about a total loss of enzymatic activity showing the essential role of this residue for catalysis. The mutant Thr71 ⟶ Ser, on the other hand conserves still 15% of the activity of the wild type enzyme. The replacement of His170 by tyrosine in anthranilate synthase, the enzyme that catalyses the first reaction in the biosynthesis of tryptophan, brings about the loss of activity of the enzyme with the substrate glutamine, but not with ammoniac. This reaction proceeds normally *via* the formation of a covalent intermediate, glutaminyl-Cys84. The results obtained after directed mutagenesis indicate the essential role of histidine that seems to interfere like a general base catalyst in the glutaminylation of cysteine 84.

In the case of glycyl-tRNA synthetase, chemical modification experiments with N-ethylmaleimide had suggested that Cys395 is essential to the catalysis. A mutant Cys395 ⟶ Glu was constructed. It presented a weak but significant activity, representing about 10% of that of the wild type enzyme. The mutation was not affect-ing the rate of formation of aminoacyl adenylate, but merely the step depending on tRNA. These experiments showed that the cysteine residue is not a catalytic residue. Actually the mutation, like the labelling by NEM, limits the reaction by a steric effect.

The α subunit of tryptophan synthase catalyses the following reaction:

indole-3-glycerol phosphate ⇌ indole + glyceraldehyde-3-phosphate

The β subunit catalyses the formation of tryptophan starting with indole and serine. It was shown that the residue glutamate 49 of the α subunit is a group essential to the activity. Indeed its replacement by whatever one of the 19 other amino acids by directed mutagenesis brings about a totally inactive enzyme.

9.4. STRUCTURAL STUDIES BY RADIOCRYSTALLOGRAPHY AND NUCLEAR MAGNETIC RESONANCE OF THE ACTIVE CENTRE OF ENZYMES

The precise knowledge of protein structures resulting from development of crystallographic and nuclear magnetic resonance studies (NMR) constitutes an important contribution to the determination of active sites of enzymes and enzyme-substrate interactions. It permits specifying the position of enzyme groups in relation to diverse substrate atoms, therefore the knowledge of groups localised to the proximity

of the disrupted substrate bond and groups susceptible to interfering in the catalysis. Moreover it permits determination of structural changes that interfere when the substrate is bound to the enzyme. If crystallisation stays a limiting stage in the determination of the structure of enzymes, the crystallogenesis techniques have progressed considerably over the course of recent years going to the automation of methods, including weak volumes on the order of a nanoliter (KNIL et al., 2002). On the other hand, the use of genetic engineering permits the overexpression of genes and the production of large quantities of the protein studied. Nevertheless this overexpression frequently leads to the formation of aggregates called inclusion bodies. In favorable cases, the native protein can be obtained starting from these aggregates by dissolving in a denaturant and then refolding (YON, 2004).

However radiocrystallographic studies are difficult carried out with real substrates; the enzyme-substrate complex decomposes into reaction products during the acquisition of X-ray data. One must resort to substrate analogs or specific inhibitors. To obtain the complexes between an enzyme and its pseudo-substrates, there are several possibilities. The protein crystals contain around 50% of solvent and some substrates diffuse in the crystal without perturbing the enzyme structure. In other cases, it is possible to co-crystallise the enzyme in the presence of pseudo-substrates. In these conditions, if there are no profound modifications in the enzyme structure in the presence of its substrate analog, it is possible to solve the structure of the complex by the FOURIER difference technique. The method permits measuring differences between diffraction diagrams of crystals of the free enzyme and crystals of the enzyme having co-crystallised with its ligand. Therefore, the changes can be observed without necessarily solving the entire structure of the complex. The first attempts to determine the structure of enzyme-substrate complexes were based on the extrapolation of results obtained starting with enzyme-inhibitor complexes. In the case of lysozyme, a portion of the substrate playing the role of the inhibitor was bound to the enzyme, and the rest was determined starting from molecular models. Today molecular modelling methods constitute an effect-ive tool in the determination of enzyme-substrate interactions.

Starting from coordinates deposited in the Protein Data Bank (PDB), in using modelling software like for example SwissPDB Viewer accessible on the internet, it is possible to view the three-dimensional structure of a protein and to turn it in space in order to visualise the regions of interest.

Despite their power and their precision, radiocrystallography methods exclude neither the chemical methods nor the genetic methods for determination of active sites of enzymes; these approaches are complementary. The use of chemical labelling and the analysis which is derived from it stays very useful for determining the role of amino acids in contact with substrate atoms. In addition, the chemical approach was used to respond to an important question: is the structure of the active site of an enzyme the same in the crystal where constraints exist and in solution? Use of bifunctional reagents has permitted in certain cases to evaluate the distance between

functional groups of an enzyme in solution and to compare it to those derived from the structural analysis. In other cases the comparison was done starting with the nuclear magnetic resonance (NMR) analysis. In fact, in most systems, the structure of an enzyme is analogous in the crystal and in solution. However, when a protein oscillates between two or several conformations, crystallography can trap one of these conformations. This is observed for example in the case of allosteric enzymes, and in precise cases, the contribution of these methods has permitted the determination of steps of this transition (see Part V).

For certain enzymes, like lysozyme, the nature of catalytic groups was not known before structural studies. For others, structural studies confirmed and stated precisely the results of chemical studies; in some cases, they were refuted. In order to illustrate the complementarity between the two approaches, some examples are given in Chap. 12 where are treated in detail some enzymatic systems of which the three-dimensional structure is known. The chosen examples show the import-ance of structural studies in the determination of groups essential to the enzymatic activity and in the determination of catalytic mechanisms; they equally underline the limits. They also indicate the necessity of uniting the structural studies, the chemical and/or genetic approaches and the kinetic studies. It is only by the conjunction of these different methods that a catalytic mechanism can be elucidated.

High field NMR for the study of proteins was developed more recently. Beginning from the 1970s, NMR of the proton began to be used for the study of proteins. Then, more recently the study of nuclei ^{13}C and ^{15}N was considerably improved permitting the application of multidimensional heteronuclear spectroscopy methods. NMR has the advantage of applying to proteins in solution and no longer in the crystal; contrary to X-ray diffraction, it identifies the proton. However it requires important protein concentrations on the order of 0.5–1 mM. Taking into account this limitation, it is important to have specific probes in the study of protein-ligand interactions and the determination of active sites of enzymes. The most judicious is to use probes carried by the ligand after isotopic enrichment in ^{13}C or ^{15}N or even ^{19}F or ^{2}H after chemical modification. With phosphorylated compounds, the resonance spectroscopy of ^{31}P is very utilised. For example, the spectrum of ^{31}P of free NADPH and that of the coenzyme bound to a dehydrogenase give information on the conformation of the bound coenzyme and on its interactions with the enzyme. The results showed that the conformation of the bound coenzyme differs from that of the free coenzyme and revealed an electrostatic interaction with the positively charged histidine of the enzyme. Another study showed that the binding of 3'CMP to ribonuclease affects the position of protons C2 and C4 of histidine 12 and of histidine 119 that are catalytic residues.

For the methodological aspect of radiocrystallography and of NMR of proteins, we redirect the reader to two fundamental books: *Biophysical chemistry. Part II for the study of biological structure and function*, by C. CANTOR & P.R. SCHIMMEL (FREEMAN ed., 1980) and *Biologie structurale: principes et méthodes biophysiques*, by J. JANIN & M. DELEPIERRE (HERMANN ed., 1994).

In summary, the study of the topology of the active centre of enzymes requires an ensemble of complementary methods. The kinetic approach as a function of pH permits determining the role of groups implicated in the activity, but does not authorise the conclusion of the nature of these groups. The research on the nature of amino acid residues implicated in the catalysis necessitates the utilisation of chemical and genetic methods bringing about in parallel enzymatic studies.

Different strategies are employed for the chemical modification of residues of the active centre of an enzyme. The use of pseudo-substrates and of affinity or photoaffinity reagents when it is possible to prepare the adequate reagent brings about a specific labelling. However, when this strategy cannot be used, there exists an ensemble of more or less selective reagents for the group that one wants to label. In choosing sensibly the conditions of pH, ionic force, it is possible to selectively modify a precise residue. It is important to underline that chemical methods require numerous controls and a rigorous analysis in order to establish if the labelled residue is essential to the catalysis.

Directed mutagenesis methods that are highly developed constitute another way to approach very efficiently the determination of residues essential to the catalysis; they are more and more utilised. In this case equally, the interpretation of the results must be based on a rigorous analysis, and numerous controls of structural properties of the mutated protein must be carried out.

The combination of all these methods largely benefits from precise structural knowledge in order to rationally orient experimentation. The number of structures of proteins currently known is on the order of several tens of thousands and still increasing, offering therefore great possibilities to study enzymatic function. Despite these diverse approaches and the power of tools at our disposal, many mechanisms are not yet totally elucidated.

BIBLIOGRAPHY

BOOKS

CANTOR C.R. & SCHIMMEL P.R. –1980– *Biophysical chemistry. Part II: Techniques for the study of biological structure and function*, Freeman, San Francisco.

DIXON M. & WEBB E.C. –1964– *Enzymes*, Longmans, London.

GLAZER A.N., DELANGE R.J. & SIGMAN D.S. –1975– Chemical modifications of proteins. Selected methods and analytical procedures, in *Laboratory techniques in biochemistry*, T.S. WORK & E. WORK eds, North Holland American, Elsevier, New York.

JANIN J. & DELEPIERRE M. –1994– *Biologie structurale: principes et méthodes biophysiques*, Hermann, Paris.

LUNDBLAD R.L. & NOYES C.M. –1985– *Chemical reagents for protein modifications*, Vol. I and II, CRC Press, Boca Raton.

MC PHERSON M.J. –1990– *Directed mutagenesis: a practical approach*, Oxford University Press.

MEANS G.E. & FEENEY R.E. –1971– *Chemical modifications of proteins*, Holden Day Inc., San Francisco.

REVIEWS

BAYLEY H. & KNOWLES J.R. –1977– Photoaffinity labeling. *Methods Enzymol.* **46**, 69–114.

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

BODLAENDER, P. FEINSTEIN G. & SHAW E. –1969– *Biochemistry* **8**, 4941.

BORNMANN L. & HESS, H. –1974– *Hoppe Seylers Z. Physiok. Chem.* **355**, 1073.

CHOWDHRY V. & WESTHEIMER L.H. –1979– Photoaffinity labelling of biological systems. *Annu. Rev. Biochem.* **48**, 293–325.

COHEN L.A. –1968– Group specific reagents in protein chemistry. *Annu. Rev. Biochem.* **37**, 695–726.

GLAZER A.N. –1970– Specific chemical modifications of proteins. *Annu. Rev. Biochem.* **39**, 101–130.

KOSHLAND D.E. –1960– The active site and enzyme action. *Adv. Enzymol.* **22**, 45–96.

KOSHLAND M.E., Englberger F. & Koshland D.E. –1959– *Proc. Natl. Acad. Sci. USA* **45**, 785.

LALLEGERIE P., LEGLER G. & YON J.M. –1982– The use of inhibitors in the study of glycosidases. *Biochimie* **64**, 977–1000.

MCMURRAY, CH. & TRENTHAM D.R. –1968– *Biochem. J.* **115**, 913.

O'NEIL K.T. & HOESS R.H. –1995– Phage display: protein engineering by directed evolution. *Curr. Opin. Struct. Biol.* **5**, 443–449.

PHILIPPS A.T. –1977– Differential labeling: a general technique for selective modification of binding sites. *Methods Enzymol.* **46**, 59–68.

SCHROEDER W.A., SHELTON J.R. & ROBERTSON B. –1967– *Biochim. Biophys. Acta* **147**, 590.

SINGER S.J. –1967– Covalent labeling of active sites. *Adv. Prot. Chem.* **22**, 1–54.

TENU J.P. –1978– Thèse Université de Paris-Sud Orsay.

TOBIN M.B., GUSTAFSSON G. & HUISMAN G.N. –2000– Directed evolution: the rational basis for irrational design. *Curr. Opin. Struct. Biol.* **10**, 421–427.

VALLEE B.L. & RIORDAN J.F. –1969– Chemical approaches to the properties of active sites of enzymes. *Annu. Rev. Biochem.* **38**, 733–794.

WALSH C.T. –1984– Suicide substrates, mechanism-based enzyme inactivators: recent developments. *Annu. Rev. Biochem.* **53**, 493–535.

WESTHEIMER F.H. –1962– *Adv. Enzym.* **24**, 441.

WILCOX P.E. –1967– Esterification. *Methods Enzymol.* **11**, 605–617.

WILSON O.S. & SZOSTAK J.W. –1998– In vitro selection of functional nucleic acids. *Annu. Rev. Biochem.* **68**, 611–645.

WOLD F. –1967– Bifunctional reagents. *Methods Enzymol.* **11**, 617–640.

WOLFENDEN R. –1977– Transition state analogs as potential affinity labeling reagents. *Methods Enzymol.* **46**, 15–28.

YON J.M. –2004– Protein aggregation, in *Encyclopedia of molecular biology and molecular medicine*, Vol. V, R.A. MEYERS ed., J. Wiley VCH, New York.

SPECIALISED ARTICLES

ALBER T., PETSKO G.A. & TSERNOGLOU D. –1976– *Nature* **263**, 297, London.

ALBERTY R.A. & MASSEY V. –1954– *Biochim. Biophys. Acta* **13**, 347.

D'ALBIS A. & BECHET J.J. –1967– *Biochim. Biophys. Acta* **140**, 435.

BECHET J.J., DUPAIX A., YON J.M., WAKSELMAN M., ROBERT J.C. & VILKAS M. –1973– *Eur. J. Biochem.* **35**, 527.

BECHET J.J., GARDIENNET M.C. & YON J.M. –1966– *Biochim. Biophys. Acta* **122**, 101.

BLOW D.M., BIRKTOFT J.J. & HARTLEY B.S. –1969– *Nature* **221**, 337.

BODE W., CHEN Z., BARTELS K., KUNTZBACH C., SCHMIDT KASTEN G. & BARTURIK H. – 1983– *J. Mol. Biol.* **164**, 237.

BODLAENDER P., FEINSTEIN G. & SHAW E. –1969– *Biochemistry* **8**, 4941.

CHEVALLIER J. & YON J.M. –1966– *Biochim. Biophys. Acta* **112**, 116.

COHEN J.A. & WARRINGA M.G. –1953– *Biochim. Biophys. Acta* **11**, 52.

CRESTFIELD A.M., MOORE S. & STEIN W.H. –1963– *J. Biol. Chem.* **238**, 622.

CRESTFIELD., STEIN W.H. & MOORE S. –1963– *J. Biol. Chem.* **238**, 2413 and 2421.

DEGANI Y., NEUMANN H. & PATCHORNIK A. –1970– *J. Am. Chem. Soc.* **92**, 6969.

DEGANI Y. & PATCHORNIK A. –1974– *Biochemistry* **13**, 1.

ELLMAN G.L. –1959– *Arch. Biochim. Biophys.* **82**, 70.

ENGELBERGER F. & KOSHLAND D.E. –1959– *Proc. Natl Acad. Sci. USA* **45**, 1470.

FEINSTEIN G., BODLAENDER P. & SHAW E. –1969– *Biochemistry* **8**, 4949.

GOLD A.A. –1965– *Biochemistry* **4**, 897.

HANES J. & PLÜCKTHUN A. –1997– *Proc. Natl Acad. Sci. USA* **94**, 4957.

HARSHMAN S. & NAJJAR V.A. –1965– *Biochemistry* **4**, 2526.

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

HORTON H.R. & KOSHLAND D.E. –1965– *J. Am. Chem. Soc.* **87**, 1126.

HUGHES P.L., SIEKES L.C., BIETH J. & DIMICOLI J.L. –1982– *J. Mol. Biol.* **162**, 645.

JACOBSON G.R., SCHAFFER M.H., STARK G.R. & VANAMAN T.C. –1973– *J. Biol. Chem.* **248**, 6583.

KARTHA G., BELLO J. & HARKER D. –1967– *Nature* **213**, 862, London.

KEMPE D. & STARK G.R. –1975– *J. Biol. Chem.* **250**, 6861.

KNIL M.E., BODENSTAFF E.R., HOENMAEKER F.J. & ABHAHAMS J.P. –2002– *Enzyme Microb. Tech.* **30**, 262.

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

LIPSCOMB W.N. –1973– *Proc. Natl Acad. Sci. USA* **70**, 37.

MASSEY V. & ALBERTY R.A. –1954– *Biochim. Biophys. Acta* **13**, 354.

MATTHEWS B.W., SIGLER P.B., HENDERSON R. & BLOW D.M. –1967– *Nature* **214**, 652.

MILHAUSEN M. & LEVY H.R. –1975– *Eur. J. Biochem.* **50**, 453.

OPPENHEIMER L., HESS G.P. & LABOUESSE B. –1966– *J. Biol. Chem.* **241**, 2727.

PARSONS S.M., JAO L., DALQUIST F.W., BORDES C.L., GROFF T., RACS J. & RAFTERY M.A. –1969– *Biochemistry* **8**, 700.

RAFTERY M.A. & COLE R.D. –1963– *Biochem. Biophys. Res. Commun.* **10**, 467.

RAY W.J. & KOSHLAND D.E. –1961– *J. Biol. Chem.* **236**, 1073.

SANGER F. –1945– *Biochem. J.* **39**, 507.

SCHOELLMANN G. & SHAW A.E. –1963– *Biochemistry* **2**, 252.

SHAW A.E., MARES-GUIA A. & COHEN W. –1965– *Biochemistry* **4**, 2219.

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

SINGH A., THORNTON E.R. & WESTHEIMER F.H. –1962– *J. Biol. Chem.* **237**, 3006.

STEITZ T.A., HENDERSON R. & BLOW D.M. –1969– *J. Mol. Biol.* **46**, 337.

TAKAHASHI K. –1968– *J. Biol. Chem.* **243**, 6171.

TSOU C.L. –1962– *Scientia Sinica* **11**, 1535.

WANG S.S. & CARPENTER F.H. –1968– *J. Biol. Chem.* **243**, 3702.

WITTER A. & TUPPY H. –1960– *Biochim. Biophys. Acta* **45**, 4.