EXPERIENTIA

Vol. 32 – Fasc. 8 Pag. 949–1090 15. 8. 1976

GENERALIA

Theories of Enzyme Specificity and their Application to Proteases and Aminoacyl-Transfer RNA Synthetases*

by H. R. BOSSHARD[1]

Biochemisches Institut der Universität, Zürichbergstrasse 4, CH-8032 Zürich (Switzerland)

Summary. The question of enzyme specificity which is a corollary of the phenomenon of biological recognition is reviewed. The following theories are outlined briefly: non-productive binding, induced fit, transition state binding, the general strain theory and the kinetic proofreading hypothesis. Data on proteolytic enzymes and aminoacyl-tRNA synthetases are discussed in the light of predictions made by the various theories. The specificity of inhibitor and substrate binding to chymotrypsin and subtilisins is revealed at the sub-molecular level as an example of binding specificity. Kinetic specificity is experimentally distinguished from binding specificity. Conformational adaptability of enzyme and substrate, which is crucial in some theories, is documented by data on aminoacyl-tRNA synthetases. Expected and observed specificity of tRNA charging is discussed with regard to a theoretical limit of specificity. Additional means seem necessary beside those contained in the isolated enzyme-substrate system to account for the high specificity of most synthetases. In conclusion, we have arrived at quite good explanations for moderate specificity such as is displayed by many proteases, but there are still ample difficulties in the understanding of highly specific enzyme reactions.

1. Introduction

Many molecules are able to recognize each other specifically. This ability is vitally important for biological systems which have to maintain a highly organized structure vis-à-vis their environment. The phenomenon has been called *biological recognition.* This term can be understood in a very broad sense as covering the whole range of specific interactions between molecules, cells and even species. Since recognizing means detecting specifically one structure among several different ones, specificity is implicit in biological recognition and might be defined as the ability to discriminate between different structures. This discussion will be confined to recognition at the level of biological macromolecules.

A few examples illustrate how widely spread biological recognition, and hence specificity, are throughout the realm of life processes. Of the best studied of examples are enzyme reactions. Enzymes have to select their specific substrates among legion of different molecules with varying but generally astonishing selectivity. In order to provide their power of defence, immunological processes must discriminate between molecules which are often structurally related. Their scope ranges from small haptens to extended molecular complexes on cell surfaces. Most hormones have to select their targets on membranes. Self-assembly of viruses depends on specific interactions between proteins and nucleic acids. And, probably most important, the transcription and translation of genes has to proceed with utmost precision in order to preserve the genetic heritage and to guarantee its faithful expression.

In this article I am going to discuss problems which arise out of this need for specificity in biological processes. The treatment will, however, be confined to the specificity of enzymes which, as mentioned above, excellently represent the range of biologically active macromolecules.

I shall first state and define the questions of enzyme specificity (section 2). Thereafter a selection of theories will be presented which aims to explain, in molecular terms, the remarkable specificity achieved by many enzymes (section 3). Finally, I shall consider experimental results which may corroborate one or the other of the proposed theories (sections 4 and 5). The analysis will be biased towards proteases and aminoacyl-transfer RNA synthetases (amino acid: tRNA ligases). The proteases represent enzymes of low to moderate specificity and the synthetases those of very high specificity. Enzyme specificity is closely related to the

* Dedicated to the memory of the late Professor JOSEF RUDINGER.

[1] I acknowledge a fellowhsip by the Schweizerischer Nationalfonds which I received while doing much of my own work cited in this article.

mechanism of enzyme catalysis, that is to the enzyme's power of accelerating a chemical reaction 'faster than any manmade catalyst'[2] under physiological conditions. However, this aspect of enzyme action will only be dealt with here in connection with its role in enzyme specificity. (The catalytic mechanisms have been treated repeatedly and extensively. For recent reviews, see[3,4] for monographs[5,6]). In other words, the emphasis will be less on how enzymes manage to accelerate reaction rates and more on how they do it so *specifically*.

To emphasize this specificity, consider the following examples. Proteins, nucleic acids and carbohydrates are more or less labile against acid or base. So, in a mixture of the three, it is not possible to hydrolyze, say, only the protein by acid without affecting the integrity of the nucleic acid and the carbohydrate. However, proteases can easily discriminate between these different substances, as can the nucleases and glycosidases. At the next level of specificity, proteases may cleave peptide bonds according to the nature (hydrophobic, acidic, basic, etc.) of the flanking amino acids; nucleases may cleave phosphodiester bonds depending on whether they link purine or pyrimidine nucleosides. On a higher level of discrimination, some proteases can cleave one single peptide bond among hundreds in a protein molecule, as does thrombin in cleaving a single arginylglycine bond in fibrinogen, transforming it to fibrin[7]. Among the nucleases, the highly specific restriction enzymes cleave high molecular weight DNA composed of thousands of nucleotides at only one or a few specific sites[8]. Finally, translation of the genetic information from messenger-RNA into protein in the process of ribosomal protein biosynthesis occurs with an error frequency which was measured in one case as averaging not more than a single error per 3000 peptide bonds formed[9]. One has to remember that a whole series of enzymes work together in the translation process, therefore each enzyme has to perform at an error frequency below the one cited above for the overall process, as long as there is no correction device at a later stage.

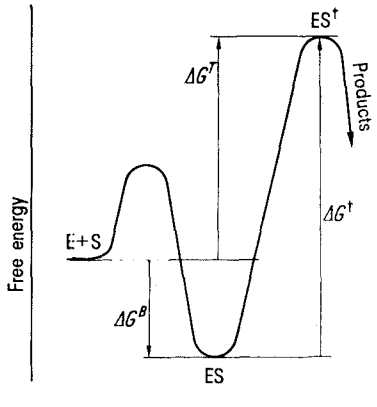


Fig. 1. Relationship between free energy of binding (ΔG^B), free energy of activation ($\Delta G^\dagger$), and free energy of the quasiequilibrium between E + S and the transition state complex ES† (ΔG^T). This simple system may be expanded by additional steps between ES and ES† and between ES† and products, but $\Delta G^\dagger$ is always the largest free energy of activation of the whole reaction.

2. *The levels of specificity*

With several superb crystal structure analyses of enzyme-ligand complexes at hand, one might wonder why specificity still remains one of the foremost puzzles of biological reactions. Indeed, the wealth of intermolecular contacts between a protein macromolecule and its ligands, that have been so splendidly revealed by the crystallographers, substantiates what EMIL FISCHER[10] foresaw in his famous lock-and-key analogy of enzyme action or what PAUL EHRLICH[11] might have had in mind in his dictum 'Corpora non agunt nisi fixata' – substances do not act unless bound.

These historical notions might seem trite to the biochemist of the Seventies but they remain the basis from which all explanatory theories of specificity have to develop: *Specificity stems from the structural complementarity of the partners in a complex.*

Specificity in enzyme catalysis is achieved at two levels. First, specificity can occur in the substrate binding step.

$$\text{Enzyme} + \text{substrate} \underset{k_-}{\overset{k_+}{\rightleftarrows}} \text{enzyme} \cdot \text{substrate} \quad K_B = k_+/k_- \qquad (1)$$

If two substrates are bound with the association constants K_B and K_B', an enzyme may differentiate between them according to K_B/K_B'.

An enzyme seldom discriminates between substrates only according to the single equilibrium of equation (1). In practice it has been repeatedly observed that substrates have the same binding constants and are turned over at different rates, or else that they show large differences in binding and almost none in rate. Therefore there must be an additional level at which specificity is achieved. This will become obvious from a thermodynamic treatment of an enzyme reaction proceeding according to the simple Michaelis-Menten scheme (2).

$$\text{E} + \text{S} \underset{k_-}{\overset{k_+}{\rightleftarrows}} \text{ES} \xrightarrow{k_{cat}} \text{products} \qquad (2)$$

[2] D. R. STORM and D. E. KOSHLAND JR., Proc. natn. Acad. Sci., USA *66*, 445 (1970).
[3] J. F. KIRSCH, A. Rev. Biochem. *42*, 205 (1973).
[4] W. P. JENCKS, Adv. Enzymol. *43*, 219 (1975).
[5] K. J. LAIDLER, *The Chemical Kinetics of Enzyme Action* (Clarendon Press, Oxford 1958).
[6] W. P. JENCKS, *Catalysis in Chemistry and Enzymology* (McGraw-Hill, New York 1969).
[7] H. A. SCHERAGA and M. LASKOWSKI JR., Adv. Protein Chem. *12*, 1 (1957).
[8] D. NATHANS and H. O. SMITH, A. Rev. Biochem. *44*, 273 (1975).
[9] R. B. LOFTFIELD and D. VANDERJAGT, Biochem. J. *128*, 1353 (1972).
[10] E. FISCHER, Chem. Ber. *27*, 2985 (1894).
[11] P. EHRLICH, in *Collected Papers* (Ed. F. HIMMELWEIT; Pergamon, London-New York 1960), vol. 3, p. 561.

The Michaelis-Menten scheme is a good approximation for many practical purposes. The free energy of binding ('freie Enthalpie' in German literature) is

$$\Delta G^B = -RT \cdot lnK_B \quad (3)$$

(R is the gas constant and T the absolute temperature.) The rate constant of product formation is k_{cat}, it relates to the free energy of activation, $\Delta G\dagger$, by

$$\Delta G\dagger = -\mathrm{RT} \cdot lnk_{cat} + RT \cdot lnB \quad (4)$$

B is constant for all reactions at the same temperature. The free energy of activation is that amount of free energy that goes into the chemical process of bond making and bond breaking when the reaction proceeds from the ES complex via the transition state complex ES$\dagger$ to the products. Figure 1 shows the relation between ΔG^B and $\Delta G\dagger$. ΔG^T is the energy that determines the quasiequilibrium between free enzyme and substrate and the transition state complex ES$\dagger$. The free energy of binding, ΔG^B, is concentration dependent. ΔG^B is usually negative (Figure 1) at a 1 M standard state, but at low substrate concentration ΔG^B is often positive, which means ES is less stable than E + S. $\Delta G\dagger$ must always be positive (otherwise there would be no ES-complex).

From Figure 1 and equation (3) and (4) follows:

$$\Delta G^T = \Delta G^B + \Delta G\dagger = -RT \cdot lnK_B - RT \cdot lnk_{cat} + RT \cdot lnB \quad (5)$$

An energy of discrimination between two substrates competing[12] for the same enzyme can thus be defined:

$$\Delta\Delta G^T = \Delta\Delta G^B + \Delta\Delta G\dagger = -RT \cdot ln\,(K_B/K'_B) - RT \cdot ln(k_{cat}/k'_{cat}) \quad (6)$$

K_B, k_{cat} and K'_B, k'_{cat} refer to the two competing substrates. The overall specificity with which the enzyme discriminates between the two substrates is depending on $\Delta\Delta G^T$. From (5) and (6) follows that overall specificity is achieved by a combination of $\Delta\Delta G^B$ and $\Delta\Delta G\dagger$.

BENDER et al.[13] have proposed on these grounds to differentiate between *binding specificity* and *kinetic specificity* (or catalytic specificity). Binding specificity must be taken in the narrow sense of the *specific formation of a complex by two molecules which encounter* (encounter complex) *and happen to be attracted together because of their structural complementarity*. Kinetic specificity is *that fraction of specificity which makes up the differnce between binding specificity and observed overall specificity*. $\Delta\Delta G^B$ and $\Delta\Delta G\dagger$ may be positive or negative, that means specificity in binding can be complemented by specificity in catalysis, or else reduced by 'negative' specificity in the catalytic steps and vice versa.

The term 'specificity' alone will be used henceforth to indicate the overall specificity of a reaction. There are several theories on the physical basis of the two levels of specificity, some of which will be analyzed in the following section. When considering actual data, it is not always possible to distinguish the two levels of specificity.

3. Theories on the origin of specificity

3.1. *Binding specificity.* As briefly mentioned in the foregoing section, the glamorous results of protein X-ray crystallography have furnished the visual representation of binding specificity. The spatial arrangement of a polypeptide chain determines the binding site which holds a ligand molecule by a combination of forces such as van der Waals forces, hydrogen bonds, electrostatic forces, hydrophobic forces, etc. The picture revealed by the X-ray analysis is static, and the comparison of the free enzyme with the enzyme-ligand complex at the usual resolution of around 2–4 Å reveals in many cases only minute structural disturbances of the enzyme upon ligand binding.

Crystal structure analysis gives no quantitative information on the binding specificity of an enzyme. Therefore it has to be combined with kinetic techniques. By relating the binding energies of a series of ligands to their chemical structure, one may map the binding site of a protein and probe its binding specificity. Keeping in mind the definition of binding specifity given in section 2, care has to be taken to fulfill the following conditions: binding must be of the simple type of equation (1) in which the enzyme does not catalyze any reaction. Needless to say, all ligands compared must be shown to bind to the same binding site (or the same set of independent and identical binding sites) in order to make meaningful deductions on binding specificity. The ligands might be substrates, provided the true binding constants of the substrates are compared. This may be difficult if binding specificity has to be deduced from comparing Michaelis-constants (K_m-values) which are not necessarily true binding constants.

Binding specificity of many enzymes, especially proteases, have been compared. Some data will be discussed in section 4.

3.2. *Kinetic specificity.* Ample data indicate that a simple relation between the binding of substrate and the rate of product formation does not hold for many enzyme reactions. Specificity is often more evident from rate constants of product formation rather than from binding constants. Five different theories for kinetic specificity will be presented briefly.

3.2.1. *The non-productive binding theory.* This has also been called the theory of multiple binding modes[14]. It assumes that a good substrate binds in the one mode which is favourable for catalysis (the productive mode), whereas the poor substrate has several

[12] It is important that the 2 substrates compete for the enzyme simultaneously. If, for example, there is no difference in k_{cat} but only in K_B, then each substrate, when tested alone, will be turned over at the same rate at saturating concentration.

[13] M. L. BENDER, F. J. KÉZDY and C. R. GUNTER, J. Am. chem. Soc. *86*, 3714 (1964).

[14] C. L. HAMILTON, C. NIEMANN and G. S. HAMMOND, Proc. natn. Acad. Sci., USA *55*, 664 (1966).

binding modes, only one of which corresponds to the productive mode. Note that 'good' and 'poor' are operationally defined terms. Once a substrate is bound in the productive mode, there is no more discrimination in the rate of product formation.

Good and poor substrates might have the same apparent binding constant, K_{app}, which is composed of the binding constant for productive binding, K_p, and the sum of the constants for non-productive binding, ΣK_{np}. The experimentally determined binding constant averages over a heterogeneity at the molecular level of enzyme-substrate complexes[15]. This heterogeneity is greater for poor substrates and ideally, it disappears for the best substrate for which ΣK_{np} approximates zero. The observed reaction rate is

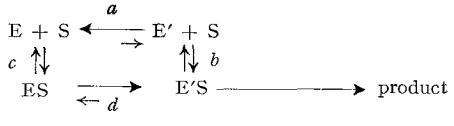
$$v_{app} = v_p K_p/(K_p + \Sigma K_{np})$$

where v_p is the rate for the productive binding mode, the only mode in which the reaction is supposed to take place. According to this equation, the observed rate, v_{app}, equals v_p for the best substrate but is smaller for any substrate that binds also in any other than the productive mode. The apparent lack of correlation between the observed binding constant and the reaction rate is only due to the difficulty of separating experimentally K_p from ΣK_{np}.

In the multiple binding mode theory, one tacitly assumes that enzyme and substrate behave as rigid bodies. The three following theories dwell on the inherent flexibility of enzyme and substrate, or what THEORELL[16] has called the 'fourth dimension' of proteins: '...the proteins cannot be described by one three-dimensional picture. They have, as a fourth dimension, their conformational adaptability.'

A second common theme of the next three theories is that a fraction of the binding energy, ΔG^B, is used to strain the enzyme molecule, or to strain the substrate molecule, or to strain both of them.

3.2.2. *Induced fit.* According to KOSHLAND's[17] theory of induced fit, the functional groups in the active site of an enzyme are not optimally positioned to exert their activity on the substrate. Substrate binding induces a conformational change in the enzyme (or a conformative response as it has also been called[18]) which rearranges the active site into its optimal conformation for catalysis. Binding energy between enzyme and substrate is used to force the enzyme into the catalytically active but energetically less favourable conformation. The following scheme explains the mechanism. E′ indicates the enzyme in the energetically unfavourable but catalytically active conformation.



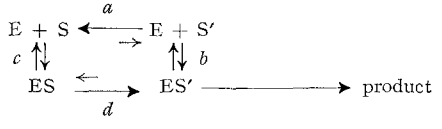
(7)

The overall binding energy ΔG^B is composed of the contributions $\Delta G_a + \Delta G_b$ (or $\Delta G_c + \Delta G_d$). The subscripts refer to the equilibria in (7). ΔG_a is positive and ΔG_b negative. ΔG_a is that fraction of the free energy of binding which is used to bring the enzyme into a catalytically active form. This means that the observed overall binding energy, and hence the association constant of a good substrate, will be lower (weaker binding) compared with the value it would have in absence of a conformative response of the enzyme. As a result, specificity is achieved at a high substrate concentration. The kinetic effect of the induced fit is an increase in the reaction rate.

A poor substrate may well bind to the enzyme but it will lack the structural features necessary to induce the conformative response and will not be turned over into product.

The theory of the induced fit would predict a different conformative response of the enzyme to the binding of good and poor substrates. Identical responses contradict the induced fit. An increase in maximal velocity paralleled by a relative decrease in substrate binding in going from poor to good substrates is also compatible with the induced fit.

3.2.3. *Transition state binding.* The theory of transition state binding assumes that the substrate binds most strongly to the enzyme in its transition state[19–21]. To put it differently, the shape of the substrate binding site of the enzyme is complementary to the substrate in the transition state rather than to the substrate in its state of lowest energy (ground state). The mechanism can be formalized similarly to the induced fit scheme.



(8)

Again the positive ΔG_a reduces the overall binding energy of the substrate. But this time binding energy equivalent in amount to ΔG_a is used to strain the substrate molecule S towards the energetically unfavourable transition state S′. Hence the free energy of activation, $\Delta G^{\dagger}$, is reduced by strain in the substrate. Specificity is also achieved at high substrate concentrations; that means the kinetic effect is an increase in the reaction rate with little change or even a decrease in substrate binding.

[15] It has been pointed out by a referee that multiple-binding can occur in 2 ways: a) The substrate binds to the enzyme in one mode, comes off again and diffuses back to be bound in a second mode. b) The substrate binds to the enzyme and oscillates between different modes of binding without leaving the enzyme. Since K_p and K_{np} are true binding constants case a) is formally treated here. But I am not aware of any experiments that would rigorously differentiate between a) and b).

[16] H. THEORELL, Harvey Lectures, Series *61*, 17 (1967).
[17] D. E. KOSHLAND JR., Proc. natn. Acad. Sci. USA *44*, 98 (1958).
[18] N. CITRI, Adv. Enzymol. *37*, 397 (1973).
[19] L. PAULING, Chem. Eng. News *24*, 1375 (1946).
[20] R. WOLFENDEN, Acc. chem. Res. *5*, 10 (1972).
[21] G. E. LIENHARD, Science *180*, 149 (1973).

There is a pitfall in this theory, since a transition state is not an intermediate. An intermediate is stable in the sense that it may be isolated under suitable conditions. In contrast, a transition state is an unstable species which can never be isolated and which is interposed between reactants and products. It has a structure somewhere in between that of substrate and product. For this reason scheme (8) could be misleading unless one keeps in mind that ES′ and ES are of very different nature. LUMRY[22] stressed this point by proposing that there is not one enzyme conformation complementary to the transition state conformation of the substrate, but that the enzyme follows the process of bond rearrangement by consecutive conformational changes; complementarity between enzyme and substrate is not static but dynamic.

3.2.4. *The general strain theory.* According to this theory (4, 6), binding forces are used directly to strain the substrate *and* the enzyme *towards* the transition state. The observed binding energy for the substrate is lower than it would be in the absence of strain. The theory can be explained with the help of Figure 2. E*S* stands for the enzyme-substrate complex that has a conformation more similar to the transition state ES† than the non-strained enzyme-substrate complex ES. Free energy correponding to ΔG_{strain} has been used to strain the hypothetical ES-complex to E*S*. This free energy has been deducted from the intrinsic[4] binding energy ΔG^B and the effect is a reduced binding energy ΔG^B_{obs}. ΔG^B_{obs} is the quantity measurable by the usual binding studies. The reaction rate constant is enhanced because the free energy of activation, $\Delta G\dagger$, is reduced by ΔG_{strain} to $\Delta G\dagger$.

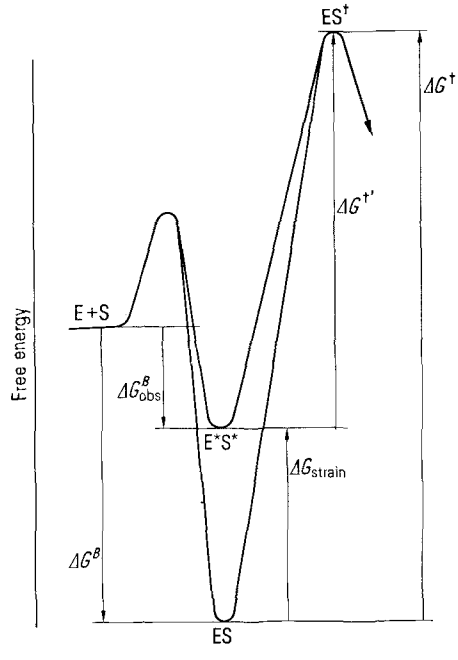


Fig. 2. Energy profile to explain the general strain theory. ΔG^B is the intrinsic binding energy corresponding to the hypothetical enzyme-substrate complex ES. Experimentally only ΔG^B_{obs} can be measured and the actual enzyme-substrate complex E*S* has a conformation more similar to the transition state ES†. Part of the intrinsic binding energy is used to strain ES to E*S*, correspondingly the free energy of activation is lowered from $\Delta G\dagger$ to $\Delta G\dagger'$.

Kinetic specificity in the general strain theory is achieved by a substrate that can provide a large portion of its intrinsic binding energy to strain the enzyme-substrate complex towards the transition state. A poor substrate might well bind, it may even show a larger ΔG^B_{obs} than a good substrate, but little of the intrinsic binding energy ΔG^B goes into ΔG_{strain}, and $\Delta G\dagger'$ is therefore larger.

The difference between the induced-fit and the general strain theory is that in the induced-fit binding energy is used to change a catalytically inactive enzyme conformation to an active conformation. The free energy of activation is not lowered by the induced-fit mechanism. However, according to the general strain theory, much of the free energy of binding is used to reduce $\Delta G\dagger$ and therefore to increase k_{cat} of the reaction. The induced-fit and the non-productive binding theory can explain enhanced specificity but they do not account for a lowered free energy of activation. The general strain theory is superior in the sense that it provides an explanation for both enhanced specificity and a larger k_{cat}.

It should be added here that the induction of strain at the cost of the binding energy is only one of several possible mechanisms to reduce the free energy of activation. Part of the binding energy may also be used to overcome the loss of translational and overall rotational entropy of the substrate when it binds to the enzyme, or to pay for desolvation of a polar substrate molecule that binds to a less polar binding site[4].

Experimental indications predicted by the general strain theory are the same as those for induced fit and transition state binding. Absence of a conformative response of the enzyme would be incompatible with the general strain theory.

Any combination of the four theories presented so far is possible. In particular, non-productive binding modes of poor substrates can always be superimposed on conformative responses of enzyme and substrate.

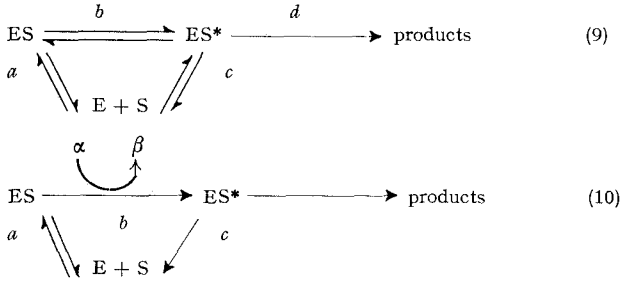
Parenthetically, I come back to the conformative response of the enzyme. There are two ways relating substrate binding and conformative response. Either the substrate is given a guiding or instructive role such that the conformative response is a consequence of substrate binding (path $c \rightarrow d$ in scheme (7)). Alternatively, the enzyme could pre-exist in the two conformations E and E′ and the substrate could bind better to either of the two conformations (path $a \rightarrow b$ in scheme (7)). In this case, the substrate acts selec-

[22] R. LUMRY, Ann. N.Y. Acad. Sci. 227, 46 (1974).

tively rather than guiding. Incidentally, this question of pre-existing versus induced conformational changes is at the centre of the two leading models of allosteric control, the concerted model[23] and the sequential model[24].

3.2.5. *The kinetic proofreading hypothesis.* There must be a limit to the overall specificity if for a given enzyme-substrate system binding specificity and kinetic specificity ultimately rely on the structural complementarity of enzyme and substrate. In thermodynamic notation, the frequency by which errors are made in a simple enzyme-substrate system is given by $\Delta\Delta G^T$, where $\Delta\Delta G^T$ is the largest difference between the free energy of two substrates moving along the reaction coordinate (from E + S to ES† (equation (6) and Figure 1).

HOPFIELD[25] has proposed that this theoretically limited specificity will be enhanced when the enzymatic reaction is coupled to an energy providing system. Consider the following two reactions:



Equation (9) is the Michaelis-Menten pathway of scheme (2) extended by the intermediate ES* (the asterisk indicates only that the intermediate differs from ES, nothing is implied about the kind of difference). To simplify the discussion, let step *d* be irreversible and rate limiting, i.e. much slower than all the rates of steps *a* to *c*. Further, let step *b* be unspecific with respect to competing substrates. Discrimination between good and poor substrates will then be achieved only in the binding steps *a* and *c*. But since steps *a* through *c* are reversible, it can be shown[25] that the minimal error to be achieved in (9) is the same as that in the simpler system (2); this error is given by $\Delta\Delta G^T$. Equation (10) shows the same reaction coupled to the energy providing reaction $\alpha \rightarrow \beta$ which may be, for example, ATP → AMP + PP*i*. In (10) step *b* is irreversible because of the coupled reaction $\alpha \rightarrow \beta$. ES* may be a high energy intermediate and step *c* becomes irreversible too. The first discrimination between competing substrates happens in the binding equilibrium *a*. Once the substrate molecule reaches ES*, a second discrimination via step *c* follows. This second discrimination has been nicknamed 'proofreading' since errors from step *a* might now be corrected in step *c*, for example by preferentially expelling the wrong substrate. Figure 3 compares the energy profiles for equations (9) and (10), assuming ATP-hydrolysis as the energy providing reaction. The free energy of ATP-hydrolysis is used partly to make the reverse reaction ES ← ES* less likely by ΔG_1, and partly to make the reversal of step *c* more 'expensive' by ΔG_2. The sum of ΔG_1 and ΔG_2 amounts to the free energy of hydrolysis of ATP which is large enough (about —7 kcal · mol^{-1} at *M* reactants, pH 7.0 and 25 °C) to render steps *b* and *c* near to irreversible. The dashed profiles in Figure 3 represent a poor substrate which is discriminated against in steps *a* and *c*. For the sake of simplicity, it is assumed that discrimination against this poor substrate is in the off-rates only (k_--values).

One way to test such a proofreading mechanism is to measure the ratio of energy consumed in step *b* to the amount of product formed. For a good substrate this ratio is low and it increases for a poor substrate which escapes product formation in the proofreading step *c*. Careful analysis of the presteady state kinetics would

[23] J. MONOD, J. WYMAN and J.-P. CHANGEUX, J. molec. Biol. *6*, 306 (1963).
[24] D. E. KOSHLAND JR., G. NEMETHY and D. FILMER, Biochemistry *5*, 365 (1966).
[25] J. J. HOPFIELD, Proc. natn. Acad. Sci. USA *71*, 4135 (1974).

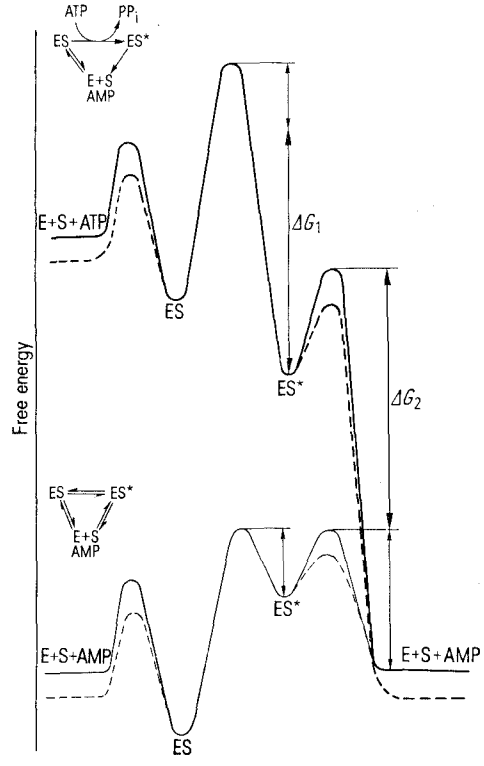


Fig. 3. The proofreading hypothesis. Energy profiles for the equilibrium reaction (9) (thin line) and for the energy driven 'one-way-reaction' (10) (thick line). The relative magnitudes of the 'valleys' and 'mountains' are chosen arbitrarily. The dashed lines indicate the profile for a poor substrate. AMP has been added to equation (9) only to facilitate the comparison of (9) and (10) with respect to the free energy scale. $\Delta G_1 + \Delta G_2$ is the free energy of hydrolysis of ATP. Further explanations are given in the text.

reveal a lag period between mixing of enzyme and substrate and appearance of product. The lag will be due to the coupled build-up of intermediate ES*.

The above covers the exposition of a few theories on enzyme specificity. I shall now turn to some practical considerations.

4. Specificity of proteases

Proteases, or proteolytic enzymes, catalyze the cleavage of peptide bonds. Some of them also cleave ester bonds. Proteases are specific with respect to the position of the peptide bond within the peptide chain, viz. exopeptidases and endopeptidases, and also with respect to the amino acids forming the bond to be cleaved. The latter are the amino acids donating the carbonyl group (on the left of the bond in the standard notation) and the amino group (on the right) to the scissible peptide bond. The amino acid contributing the carbonyl group usually dominates specificity. There is a wide spectrum of specificity from a very broad preference for, say, hydrophobic side chains (e.g. elastase, thermolysin, subtilisin, papain) to a distinct preference for phenylalanine, tyrosine and tryptophan by chymotrypsin, and for lysine and arginine by trypsin. The proteases of highest specificity are those responsible for restricted metabolic tasks such as the already mentioned thrombin or enzymes of the complement system[26]. These highly specific proteases indicate that it cannot always be one of the flanking amino acids of the scissible bond which controls specificity. Some proteases must be able to recognize larger stretches of the amino acid sequence around the cleavage site, or even different parts of the primary structure brought together by folding up the peptide chain in the three-dimensional structure of the protein molecule.

In the following, binding and kinetic specificity of mainly chymotrypsin and two subtilisins will be analyzed in more detail.

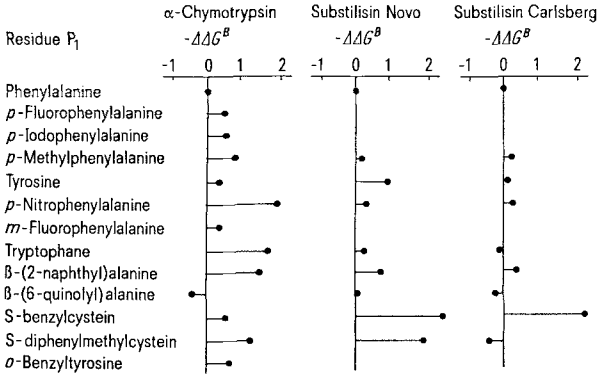


Fig. 4. Binding specificity of α-chymotrypsin, subtilisin Novo and subtilisin Carlsberg. Differences in standard free energies of binding ($\Delta\Delta G^B$ in kcal/mol) between Z-Ala-Ala-P_1 and Z-Ala-Ala-Phe were obtained kinetically (see ref.[30] for experimental data). All amino acids are of L-configuration.

4.1. *Binding specificity of proteases.* Binding specificity of proteases has been convincingly demonstrated in a series of experiments, notably by the late ARIEH BERGER and his coworkers[27–30]. Data related uniquely to binding specificity were obtained by comparing the interaction of inhibitors with an enzyme's binding site. Inhibitors were chosen so as to resemble real substrates as closely as possible. Such pseudosubstrates, or virtual substrates, are, e.g., Z-Ala-Ala-Phe[31] for chymotrypsin[30] or Ac-Phe-Leu[31] for papain[29]. In these two peptides, the carboxy-terminal amino acids, phenylalanine and leucine, occupy a position at the enzyme's binding site which in a real substrate would correspond to the amino acid next to the cleavable bond towards the amino end of the peptide chain.

In order to make meaningful comparisons among sets of inhibitors, care had to be taken to guarantee unique and uniform binding modes. Since this condition could be fulfilled by using substrate analogues, the free energy of binding of inhibitors could be divided into contributions by each amino acid.

In essence, it was found that proteases exhibit binding specificity towards several residues on both sides of the scissible peptide bond[27–29]. Proteases have therefore quite extended binding sites which can be divided formally into subsites S_1, S_2, etc. from the site of the bond cleaved toward the amino end of the substrate and S'_1, S'_2, etc. toward the carboxyl end. Accordingly, amino acid residues of the inhibitor interacting at subsites S_1, S_2, ... are named P_1, P_2, ...[27]. Many proteases are able to differentiate delicately among amino acid side chains interacting at the same subsite of the enzyme. Specificity is usually greatest at S_1, which is regarded as the main specificity determining site of a protease, and which, by definition, must be nearest to or overlapping with the catalytically important groups of the enzyme (active site).

The specificity pattern at S_1 of chymotrypsin and subtilisins type Novo and Carlsberg is illustrated in Figure 4. S_1 of chymotrypsin and subtilisins corresponds roughly to the crystallographically defined hydrophobic pockets[32,33]. Peptide inhibitors of the type Z-Ala-Ala-P_1, where P_1 stands for various amino acids, were compared[30]. Binding of phenylalanine to

[26] H. J. MUELLER-EBERHARD, A. Rev. Biochem. *44*, 697 (1975).
[27] I. SCHECHTER and A. BERGER, Biochem. biophys. Res. Commun. *27*, 157 (1967).
[28] A. BERGER and I. SCHECHTER, Phil. Trans. R. Soc. London Ser. B *257*, 249 (1970).
[29] A. BERGER, I. SCHECHTER, H. BENDERLY and N. KURN, in *Peptides 1969* (North Holland Publ. Co., Amsterdam 1971), p. 290.
[30] H. R. BOSSHARD and A. BERGER, Biochemistry *13*, 266 (1974).
[31] Z, benzyloxycarbonyl; Ac, acetyl; abbreviations of amino acids and peptides according to IUPAC-IUB rules, J. biol. Chem. *247*, 977 (1972).
[32] T. A. STEITZ, R. HENDERSON and D. M. BLOW, J. molec. Biol. *46*, 337 (1969).
[33] C. S. WRIGHT, R. A. ALDEN and J. KRAUT, Nature, Lond. *221*, 235 (1969).

subsite S_1 has been taken as a reference point and arbitrarily set as zero.

Many characteristics of the hydrophobic pockets of chymotrypsin and the subtilisins can be deduced from Figure 4. This has been done in detail elsewhere[30]. A few salient features are noteworthy. There are two factors which increase binding: substitution in the *para-* and, probably, *meta*-position of the phenyl ring and enlargement of the side chain from phenyl to naphthalene or indole. An increase in binding also occurs when the phenyl ring is moved away from C_α of P_1, i.e. probably pushed deeper into the S_1 pocket, by insertion of $-SCH_2-$ in S-benzylcystein. The largest increase in binding energy is found with *p*-nitrophenyl-alanine. This is tentatively explained by the pronounced hydrophobicity of the *p*-nitrophenyl group and by a possible hydrogen bond between the nitro group and serine residue 189 of the enzyme which lies deep in the hydrophobic pocket[32]. The predilection for hydrophobic side chains in P_1 explains the decreased binding of β-(6-quinolyl) alanine, the aromatic nitrogen of which is hydrated in aqueous solution but probably not so when bound to the hydrophobic pocket.

Binding specificity may be achieved either in the rate of complex formation (k_+ in equations (1) and (2)) or in the dissociation rate k_-. It was found that, at least for the few pseudosubstrates tested, specificity resides in dissociation rates alone, the 'on-rates' being near to diffusion controlled and independent of the structure of residue P_1[34].

A comparison of the binding specificity of chymotrypsin and subtilisins type Novo and Carlsberg points to subtle but distinct differences between these functionally similar but structurally and phylogenetically unrelated[35] proteases (Figure 4). First of all, chymotrypsin discriminates delicately among variously substituted phenylalanines and structural isomers of aryl side chains, whereas subtilisins accomodate almost any aromatic residue. This correlates well with earlier observations about the subtilisins showing lower substrate specificity than chymotrypsin and possessing a more extended and more open binding site ('pocket' versus 'crevice'[33]).

Comparison of the data with the two subtilisins exposes some remarkable difference between these two structurally related proteases (30% sequence difference)[36]. Whereas both bind the extended S-benzylcystein residue, the more bulky side chain of S-diphenylmethylcystein is rejected by the Carlsberg enzyme but not by subtilisin Novo. Tentatively one may conclude that the S_1 crevice of subtilisin Carlsberg is narrower than that of the Novo enzyme. The crystal structure of the Novo but not of the Carlsberg enzyme is known[37]. Topographical differences in the active site region of the two types of subtilisins have been deduced also from other, independent experiments[38].

Table I. Kinetic specificity of chymotrypsin towards hydrolysis of N-acetyl-L-amino acid methyl esters compared to the binding specificity towards corresponding virtual substrates[a, b]

Amino acid in P_1	K_B[b] (M^{-1})	k_{cat}[c] (sec^{-1})	$\frac{k_{cat}}{K_m}$[c] ($M^{-1}\,sec^{-1}$)
p-Nitrophenylalanine	12 000	inhibitor[g]	
p-Iodophenylalanine	12 000	inhibitor[d]	
S-benzylcystein	9 200	3.0[e]	6 280[e]
Leucine	3 100	2.9[f]	1 590[f]
Phenylalanine	4 500	52.5[f]	42 000[f]
Tyrosine	8 400	117[f]	365 000[f]
Tryptophane	36 000	49[f]	420 000[f]

[a] Binding of inhibitors of the type Z-L-Ala-L-Ala-L-P_1 is compared to the hydrolysis of the corresponding methyl esters of the type Ac-L-P_1-OMe. P_1 stands for the variable amino acid. [b] K_B is proportional to the free energy of binding, ΔG^B, according to equation (3). Values from ref.[30]. [c] k_{cat} relates to the free energy of activation, $\Delta G^\dagger$, according to equation (4). k_{cat}/K_m is a measure for the overall specificity since it is proportional to ΔG^T (equation 5). The proportionality holds even if K_m is not a true dissociation constant. [d] Ref.[44]. [e] Ref.[45]. [f] Ref.[46]. [g] Author's unpublished experiments. [h] Extensive comparison of substrate binding constants and individual rate constants for chymotrypsin catalyzed hydrolysis of esters has been reported recently by DUTLER et al.[47].

What is the physical nature of all this variability in binding specificity? The type of non-covalent interactions between some of the inhibitors of Figure 4 and the substrate binding sites of chymotrypsin and subtilisin Novo have been revealed by crystal structure analysis[39,40]. Besides the binding of residue P_1 to the hydrophobic pocket, interactions occur between subsites S_1, S_2 and S_3 and residues P_1, P_2 and P_3. The predominant interactions are with an extended segment of the main chain of the enzyme which forms an antiparallel β-structure. There are several hydrophobic bonds and also a number of possible van der Waals contacts. Thus, the inhibitors exhibit various topographically fixed interactions, they show multipoint attachment to the extended binding site. This is not only important for strong binding but also for uniqueness of the binding mode and therefore for specificity.

It has to be remembered that thermodynamic evidence is against hydrogen bonds making a major contribution to the free energy of protein-substrate interaction[41,42]. The same applies to the less energetic but

[34] H. R. BOSSHARD, I. PECHT, N. KURN and A. BERGER, 9th Int. Congr. Biochem., Stockholm (1973), Abstract Book, p. 46.
[35] B. S. HARTLEY, Phil. Trans. R. Soc. London, Ser. B *257*, 77 (1970).
[36] E. L. SMITH, R. J. DE LANGE, W. H. EVANS, M. LANDON and F. S. MARKLAND, J. biol. Chem. *243*, 2184 (1968).
[37] J. KRAUT, in *The Enzymes* (Ed. P. D. BOYER; Academic Press, New York 1971), vol. 3, p. 547.
[38] H. R. BOSSHARD, FEBS Lett. *30*, 105 (1973).
[39] J. D. ROBERTUS, J. KRAUT, R. A. ALDEN and J. J. BIRKTOFT, Biochemistry *11*, 4293 (1972).
[40] D. M. SEGAL, J. C. POWERS, G. H. COHEN, D. R. DAVIES and P. E. WILCOX, Biochemistry *10*, 3728 (1971).
[41] J. A. SHELLMAN, C. r. Trav. Lab. Carlsberg, Serv. Chim. *29*, 223 (1955).
[42] I. M. KLOTZ and J. S. FRANZEN, J. Am. chem. Soc. *84*, 3461 (1962).

more numerous van der Waals contacts[43]. Hydrogen bonds and van der Waals contacts demand precise juxtaposition of atoms and thus high complementarity between substrate and binding site. The main force to stabilize protein complexes has been described as the hydrophobic interactions[43]. Hydrophobicity is much less specific since much less dependent on the spatial arrangement of atoms contributing to hydrophobic binding. On these grounds, it has been postulated that hydrophobic forces contribute the stability to the complex and hydrogen bonds, salt bridges and van der Waals contacts determine specificity[43]. Although the point-to-point contacts contribute little to stability, lack of them entails large unfavourable enthalpies due to loss of hydrogen bonds and loose packing compared to the free, solvated enzyme and substrate molecules.

4.2. *Kinetic specificity of proteases.* The concept of kinetic specificity was formulated to interpret the apparent lack of a relationship between rate constants and binding specificity. A meaningful comparison between binding specificity and catalytic rate constants is difficult to obtain from K_m- and k_{cat}-values unless K_m is a true binding constant. Table I shows an attempt to overcome this difficulty. Rate constants for the hydrolysis of N-acetyl-L-amino acid esters are placed side by side with the binding constants of the corresponding pseudosubstrates of the structure Z-Ala-Ala-P_1. It has been reasoned elsewhere that such a comparison is feasible, thanks to the substrate-like positioning of virtual substrates on chymotrypsin's binding site[30]. Comparison with the corresponding N-acetyl amino acids would be unreliable because such small and weak inhibitors tend to show multiple binding modes. Obviously there is no such correlation as better binding: better catalysis for chymotrypsin. Whereas binding increases in the order phenylalanie < tyrosine < tryptophane, kinetic specificity expressed in k_{cat} is highest for the tyrosine derivative. Hydrolysis of the S-benzylcysteine and leucine substrates is much slower than expected from the tightness of binding of the corresponding virtual substrates. The esters of N-acetyl-*p*-nitrophenylalanine and N-acetyl-*p*-iodophenyl-alanine are not hydrolyzed at all but are good inhibitors. Several possible explanations for kinetic specificity have been outlined in section 3. Are there any data to sustain these theories? As to the induced fit and the general strain theory, there is evidence for a conformative response of chymotrypsin, and carboxypeptidase A based on physicochemical data and, most convincing, on crystal structure analysis (reviewed in[18]). The conformative response of chymotrypsin as monitored by difference spectroscopy is strongly dependent on the structure of the binding ligands[22]. As to kinetic evidence, an increase in rate, with little change in overall binding on going from poor to good substrates, is predicted by the three theories in sections 3.2.2. to 3.2.4. Such a relationship

Table II. Binding constants and differences in free energies of binding, $\Delta\Delta G^B$, for pairs of diastereomeric tripeptides Z-L-Ala-L-Ala-L(or D)-P_1 towards chymotrypsin (from ref.[55])

Amino acid P_1	K_B (M^{-1})	$\Delta\Delta G^B$ (kcal/mol)
β-(6-quinolyl)-L-alanine	2 600	0.6
β-(6-quinolyl)-D-alanine	1 000	
L-Phenylalanine	4 500	0.7
D-Phenylalanine	1 400	
m-Fluoro-L-phenylalanine	9 000	0.5
m-Fluoro-D-phenylalanine	3 800	
p-Methyl-L-phenylalanine	13 000	0.7
p-Methyl-D-phenylalanine	4 100	
β-(1-naphthyl)-L-alanine	16 000	0.8
β-(1-naphthyl)-D-alanine	4 400	0.8
L-tryptophane	36 000	0.6
D-tryptophane	13 000	0.6

exists between some substrates for pepsin[48,49], papain[29] and others[50,51]. However, it is worth mentioning that for chymotrypsin it is possible purposely to choose a series of substrates for which better binding parallels better catalysis[46], As to the transition state theory, preferential binding of transition state analogues to chymotrypsin has been noticed[52].

The theory of the multiple binding modes is upheld by various experimental data. Different binding modes of pseudosubstrates to subtilisin Novo were observed by crystal structure analysis[39]. Stringent evidence for non-productive binding of various substrates was obtained from a comparison of rate constants and K_m-values with the hydrophobicity of substrates[53]. On the basis of semi-empirical energy calculations, even different binding modes for the substrate molecule N-acetyl-L-tryptophane amide to chymotrypsin have been predicted[54]. When approach of this substrate to the binding site was simulated, four different modes of good binding were calculated[54]. Only one of these is identical with the mode deduced crystallographically.

There is also an interesting indication for different *productive* binding modes of substrate molecules on subtilisins. By kinetic techniques, it was found that the

[43] C. Chothia and J. Janin, Nature, Lond. *256*, 705 (1975).
[44] C. J. Garratt and D. M. Harrison, FEBS Lett. *11*, 17 (1970).
[45] A. P. Damoglou, H. Lindley and I. W. Stapleton, Biochem. J. *118*, 553 (1970).
[46] J. R. Knowles, J. theor. Biol. *9*, 213 (1965).
[47] S. A. Bizzozero, W. K. Baumann and H. Dutler, Eur. J. Biochem. *58*, 167 (1975).
[48] K. Inouye and J. S. Fruton, Biochemistry *6*, 1765 (1967).
[49] J. S. Fruton, Adv. Enzymol. *33*, 401 (1970).
[50] D. Atlas, J. molec. Biol. *93*, 39 (1975).
[51] K. Morihara, T. Oka and H. Tsuzuki, Arch. Biochem. Biophys. *138*, 515 (1970).
[52] A. Ito, K. Tokawa and B. Shimizu, Biochem. biophys. Res. Commun. *49*, 343 (1972).
[53] J. Fastrez and A. R. Fersht, Biochemistry *12*, 1067 (1973).
[54] K. E. B. Platzer, F. A. Momany and H. A. Scheraga, Int. J. Pept. Protein Res. *4*, 201 (1972).

substrate N-benzoyl-L-arginine ethyl ester binds in a mode different to that of the usual peptide substrates, e.g. Ala-Ala-Ala-Phe-Ala[38].

Further examples of multiple binding modes are encountered with D-amino acids which interact differently from L-amino acids at the binding sites of proteases. One sort of evidence came from binding studies with a series of diastereomeric pseudosubstrates and chymotrypsin[55]. Table II shows the results. Binding specificity of chymotrypsin towards configurational isomers is similar ($\Delta\Delta G^B$ = —0.5 to —0.8 kcal/mol) regardless of the structure of the side chain in P_1. Binding of both isomers is also governed by the same pK of 7.3, presumable histidine-57 of the active site[55]. These fidings are consistent with the following interpretation (Figure 5): The strong interaction of the aromatic side chains of residue P_1 with the hydrophobic pocket at S_1 is preserved as is the multipoint attachment of the 'Z-Ala-Ala' portion at subsites S_2 to S_4. But the carboxyl groups and the C_α protons of amino acid P_1 switch places at the active site. In the case of substrates of D-amino acids, this must lead to a positioning where the active site residues of chymotrypsin are not optimally oriented to cleave the peptide or ester bond. Binding of N-acetyl-L (or D)- amino acids differs similarly[55]. In conclusion, stereospecificity of chymotrypsin is based on non-productive binding of D-enantiomers and not on exclusion from binding of the configurational isomers. A study with subtilisins did reveal several different binding modes of D-amino acids as compared to L-amino acids[56].

In summary, there are many observations on proteases to support the four theories discussed in sections 3.2.1. to 3.2.4. Depending on what sort of substrates or inhibitors are compared, one or the other theory might seem to prevail. None of them can explain all the experimental data. The general strain theory is corroborated best, since there is no refuting observation at present.

The proteases discussed so far are moderately specific; discrimination between two very similar substrates occurs by factors of the order of 1 to 10.

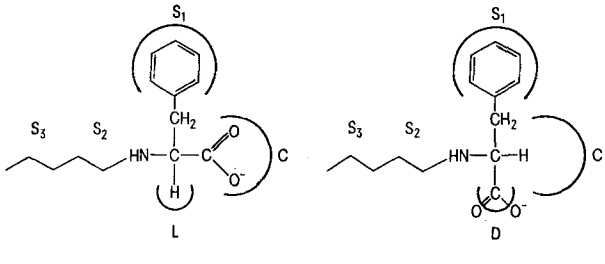


Fig. 5. Schematic representation of the binding of Z-L-Ala-L-Ala-L-Phe and Z-L-Ala-L-Ala-D-Phe to the extended binding site of α-chymotrypsin. The wavy line indicates the Z-L-Ala-L-Ala portion of the pseudosubstrate. S_1, S_2 etc. are the subsites as defined in the text, c indicates the site on the enzyme for the carboxyl group of an L-amino acid residue interacting at S_1.

Specificity depends on several residues around the cleavage point. But what about thrombin and other highly specific proteases? They cleave peptide bonds with a selectivity of the order of one per several hundred or event thousand. At present we do not know any detailed reason for this remarkable specificity. One hint could be the extended binding sites of the moderately specific proteases. It seems reasonable to hypothesize that the larger the number of point-to-point contacts necessary for productive substrate binding the smaller will be the number of good substrates and the higher the specificity. But there are no quantitative data to sustain this notion.

Some of the open questions with highly specific proteases will be encountered also with aminoacyl-tRNA synthetases, the family of enzyme I am going to discuss next.

5. Specificity of aminoacyl-tRNA synthetases

Aminoacyl-tRNA synthetases are indirectly responsible for the correct translation of genetic information into proteins (for a recent review, see ref[57]). The enzymes catalyze the esterification of amino acids to the 3'-end of tRNA in a highly specific manner. The reaction can be formally divided into two steps, activation of the amino acid to the 5'-aminoacyladenylate by means of ATP and then transfer of the activated amino acid to the 3'-terminal adenosine of a particular tRNA.

$$E + aa + ATP \rightleftharpoons E.aa{\sim}AMP + PPi \quad (11)$$

$$E.aa{\sim}AMP + tRNA \rightleftharpoons aa\text{-}tRNA + E + AMP \quad (12)$$

aa = amino acid, E = enzyme, PP_i = inorganic pyrophosphate

Whether the aminoacyladenylate (aa~AMP) is an obligatory intermediate in the sequence of reaction, or whether a concerted mechanism is likely at least for some synthetases is not yet known with certainty[58]. The enzymes show bispecificity towards amino acid and tRNA.

I have argued in the introduction that the error frequency of each step of the ribosomal protein biosynthesis must be below the observed overall error frequency, which is one error in 3 000 peptide bonds formed for the incorporation by rabbit reticulocytes of valine versus isoleucine in the amino-terminal peptide of globin[9]. Assuming that such an error frequency is representative, which is by no means certain, one out of 12 protein molecules of a length of 250 residues will contain a single wrong amino acid somewhere in its sequence. On further assuming that only 10% or so of all errors seriously interfere with the biological function of the protein, we arrive at 1 protein in about 100 which might interfere with metabolism. It would seem that

[55] H. R. BOSSHARD, FEBS Lett. *38*, 139 (1974).
[56] H. R. BOSSHARD, Israel J. Chem. *12*, 495 (1974).
[57] L. L. KISSELEV and O. O. FAVOROVA, Adv. Enzymol. *40*, 141 (1974).
[58] R. B. LOFTFIELD, Progr. nucl. Acid. Res. molec. Biol. *12*, 87 (1972).

any organism might cope with such a burden and that therefore the error of 1 in 3000 might at least not be seriously underestimated. If this is so, aminoacyl-tRNA synthetases would have to charge tRNA's at an error frequency well below 1 in 3000. Yet there are numerous reports about misacylations of tRNA's in vitro.

Mixed systems composed of synthetases and tRNA's from different organisms[59] are especially prone to errors in acylation. A striking example is charging of some tRNA species from *Bacillus stearothermophilus* by valyl-tRNA synthetase from yeast. Mischarging to the extent of 92% for $tRNA^{Met}$, 81% for $tRNA^{Ala}$, 53% for $tRNA^{Pro}$ and so forth has been reported[60,61]. Admittedly this system has been selected for mischarging and does not represent the situation in vivo. In homologous systems, non-cognate binding of tRNA (K_m-values) has been estimated to be diminished by one to two orders of magnitude with respect to cognate binding and non-cognate charging (k_{cat}) by two to four orders of magnitude[62,63]. Specificity of the order of 10^7 in favour of correct acylation has been calculated from individually determined maximal velocities and Michaelis constants for cognate and non-cognate acylation in vitro[60].

5.1. *Binding specificity of synthetases.* There are few data on the binding specificity of synthetases. The affinity of individual substrates to the synthetase molecule in absence of other substrates necessary for the full reaction to occur might not be relevant to the real affinities. Conformative responses do often occur upon substrate binding (section 5.2.1.).

Generally, affinities of amino acid and ATP are weaker than that of tRNA. Mainly the side chain and the amino group of the amino acid are specificity determining and the carboxyl group plays little part in the association[57]. Studies on the binding of leucine and histidine and their analogues to the corresponding enzymes have been reported[64,65], as have experiments on the binding of ATP analogues to the methionine specific enzyme[66].

Several approaches have been tried to get insight into the binding of tRNA to the synthetase. For the sake of simplicity they may be divided into those focussing on the tRNA and those dealing with the synthetase. In both approaches either tRNA or synthetase or the tRNA-synthetase complex are studied. For obvious reasons the complex should yield the least ambiguous results. The approach from the tRNA side has been favoured so far, for the simple reason that much more tRNA's than synthetases have been isolated and sequenced. Some of the techniques applied were chemical modification of tRNA and its implication on complex formation and aminoacylation, nuclease digestion of free and complexed RNA, nucleotide sequence comparison with the aim of finding common structures indicating synthetase binding sites, and binding studies with mutant tRNA's (reviewed in[57]). Photo-irradiation of enzyme-tRNA complexes yields covalent linkages between enzyme and nucleic acid[67]. Cross-linking of several cognate and non-cognate complexes was achieved[67,68]. Apart from some differences in the contact area between cognate and non-cognate complexes, there emerged a surprizingly uniform picture of the protein-nucleic acid interactions in the different complexes analyzed to date[69]. The main contacts are centered on that area of the tRNA molecule where the two helical branches of the L-shaped molecule are fused[70,71].

Studies approaching the problem from the side of the synthetase are scarce. There have been attempts to introduce reactive groups into the tRNA molecule and to crosslink the resulting affinity label with the synthetase[72,73]. I have pursued a novel approach in localizing some lysine residues of the synthetase which are in contact with the tRNA molecule[74]. The method consists in comparing chemical reactivities of ε-amino groups of lysine residues in the free synthetase and the synthetase-tRNA complex. The assumption made is that, in the synthetase-tRNA complex, some of the ε-amino groups have a decreased reactivity toward acylating agents. This occurs because they might be shielded, for example by forming salt bridges to the phosphate backbone of the tRNA, or more generally, because access of the acylating agent to the lysine side chains is sterically hindered in the complex. Alternatively, an altered conformation imposed on the

[59] Reactions between tRNA's and synthetases from different organisms are called 'heterologous' as compared to 'homologous' for the reaction between enzyme and substrate of the same origin.

[60] R. Giegé, D. Kern, J. P. Ebel, H. Grosjean, S. de Henau and H. Chantrenne, Eur. J. Biochem. *45*, 351 (1974).

[61] The superscript in $tRNA^{Met}$ etc. indicates the amino acid which the tRNA is mainly specific for. Charging according to this main specificity by the corresponding synthetase is called 'cognate', thus valyl-tRNA synthetase and $tRNA^{Val}$ is a cognate pair. The term 'non-cognate' is self-explanatory. Cognate and non-cognate charging can of course be homologous as well as heterologous.

[62] J. P. Ebel, R. Giegé, J. Bonnet, D. Kern, N. Befort, C. Bollack, F. Fasiolo, J. Gangloff and G. Dirheimer, Biochimie *55*, 547 (1973).

[63] J. Bonnet and J. P. Ebel, FEBS Lett. *39*, 259 (1974).

[64] J. Flossdorf and M. G. Kula, Eur. J. Biochem. *36*, 534 (1973).

[65] G. C. Lepore, P. di Natale, L. Guarini and F. de Lorenzo, Eur. J. Biochem. *56*, 369 (1975).

[66] F. Lawrence, D. J. Shire and J. P. Waller, Eur. J. Biochem. *41*, 73 (1974).

[67] H. J. P. Schoemaker and P. R. Schimmel, J. molec. Biol. *84*, 503 (1974).

[68] G. P. Budzik, S. M. M. Lam, H. J. P. Schoemaker and P. R. Schimmel, J. biol. Chem. *250*, 4433 (1975).

[69] H. J. P. Schoemaker, G. P. Budzik, R. Giegé and P. R. Schimmel, J. biol. Chem. *250*, 4440 (1975).

[70] J. D. Robertus, J. E. Ladner, J. T. Finck, D. Rhodes, R. S. Brown, B. F. C. Clark and A. Klug, Nature, Lond. *250*, 546 (1974).

[71] S. H. Kim, F. L. Suddath, G. J. Quigley, A. McPherson, J. L. Sussman, A. H. J. Wang, N. C. Seeman and A. Rich, Science *185*, 435 (1974).

[72] C. J. Bruton and B. S. Hartley, J. molec. Biol. *52*, 165 (1970).

[73] D. V. Santi and S. O. Cunnion, Biochemistry *13*, 481 (1974).

[74] H. R. Bosshard, to be published.

Table III. Asymmetry of substrate binding to aminoacyl-tRNA synthetases from *B. stearothermophilus*

Synthetase specific for	Subunit[a] composition	Number of binding sites for			
		Amino acid from equil. dialysis)		Aminoacyladenylate	
		No ATP	With ATP	From gel filtration	Kinetically determined
Tyrosine	2× 45 000	1[b]	2[d]	1[b]	2[d]
Methionine	2× 66 000	1[c]	–	2[c]	—
Leucine	1×110 000	1[c]	–	1[c]	2[e]
Valine	1×110 000	1[c]	2[d]	1[c]	2[d]
Tryptophane	2× 35 000	1[c]	–	2[c]	–

[a] Ref.[80]; [b] ref.[76]; [c] ref.[81]; [d] ref.[77]; [e] A. R. FERSHT and R. S. MULVEY, personal communication.

synthetase by complex formation with the tRNA might also reduce reactivity of ε-amino groups outside the tRNA binding site. The method on its own is not able to distinguish strictly between these two interpretations of altered reactivities.

The method was applied to the complex of $tRNA^{Tyr}$ with tyrosyl-tRNA synthetase from *B. stearothermophilus*[74]. It was found that, while the majority of lysine residues are equally accessible in the free enzyme and the enzyme-tRNA complex, a few show decreased reactivity (up to 7-fold) in the complex. The decreases are specifically due to complex formation with the cognate tRNA only. It will be most interesting to localize these contact-making lysine residues in the primary sequence of the tyrosyl-tRNA synthetase and finally to demarcate the tRNA binding site on the three-dimensional model of the enzyme[75]. This work is in progress and should be instrumental in the elucidation at the 'near-to-atomic' level of nucleic acid-protein interactions.

Evidence for binding of non-cognate tRNAs to synthetase comes from inhibition experiments in which non-cognate tRNAs inhibit acylation of $tRNA^{Val}$ by valyl-tRNA synthetase[62].

5.2. *Kinetic specificity of synthetases*

5.2.1. *Non-equivalence of substrate binding sites.* Asymmetry in the binding of amino acid and tRNA to synthetases has been observed[76–79]. Table III summarizes results obtained with 5 enzymes from *B. stearothermophilus*. The dimeric enzymes are composed of identical subunits. In the tyrosine-specific enzyme, the monomer is the asymmetric unit of the crystal[75]. The leucine and valine specific enzymes are likely to be composed of two similar or identical domains on the three-dimensional level which are formed by sequence repeats within the primary structure[78]. The situation is reminiscent of the one with immunoglobulins where domains have been predicted and crystallographically confirmed[82]. The domains in the leucine and valine specific enzymes might be the structural basis for the two binding sites in the monomeric enzymes.

The 5 synthetases from *B. stearothermophilus* tested (Table III) have two binding sites for amino acid and aminoacyladenylate, and most probably also two active sites irrespective of their subunit composition[76,77,81]. On the tyrosine and valine enzyme, one amino acid binding site is detectable by equilibrium dialysis in absence of ATP and two in its presence. Only those aminoacyladenylate-enzyme complexes stable over the period of the experiment are detected by the gel filtration method. In fact, the complexes have half-lives of several hours at 4 °C[76]. The second adenylate binding sites are much less stable and only deduced indirectly from kinetic analysis[77].

The sites are either non-equivalent to begin with, or the second site becomes partially blocked upon binding of the first substrate molecule. There is evidence discussed elsewhere[76] that asymmetry is more likely to be induced in the tyrosyl-tRNA synthetase by the binding of the tyrosine molecule, i.e. the ligand guides rather than selects (see section 3.2.4.).

All these data demand conformational adaptability of the synthetase molecule, a feature compatible with the induced fit and the general strain theory. It is not known whether non-cognate amino acids also show such asymmetric binding, since under the experimental conditions prevailing only binding of cognate amino acids was strong enough to be detected[76,81].

Binding of tRNAs to several synthetases is asymmetric too. There was found a strong and a weak site for the cognate tRNAs on tyrosyl-tRNA synthetase from *B. stearothermophilus*[76], and *E. coli*[79], and on phenylalanyl- and seryl-tRNA synthetases from *E.*

[75] B. R. REID, G. L. E. KOCH, Y. BOULANGER, B. S. HARTLEY and D. M. BLOW, J. molec. Biol. *80*, 199 (1973).
[76] H. R. BOSSHARD, G. L. E. KOCH and B. S. HARTLEY, Eur. J. Biochem. *53*, 493 (1975).
[77] A. R. FERSHT, Biochemistry *14*, 5 (1975).
[78] G. KRAUSS, A. PINGOUD, D. BOEHME, D. RIESNER, F. PETERS and G. MAASS, Eur. J. Biochem. *55*, 517 (1975).
[79] A. PINGOUD, D. BOEHME, D. RIESNER, R. KOWNATZKI and G. MAASS, Eur. J. Biochem. *56*, 617 (1975).
[80] G. L. E. KOCH, Y. BOULANGER and B. S. HARTLEY, Nature, Lond. *249*, 316 (1974).
[81] Y. BOULANGER, H. R. BOSSHARD and G. L. E. KOCH, Eur. J. Biochem., in press.

coli[78]. It has been shown further that $tRNA^{Tyr}$ acts by guiding in the complex with the *E. coli*, i.e. there are not two pre-existing conformations of the enzyme[79].

Many synthetases do not show asymmetry in substrate binding, e.g. the tryptophanyl-tRNA synthetase from *B. stearothermophilus*[81] (Table III). But there is still another indication for conformational adaptability in the fact that substrate binding often occurs synergistically, that means binding of one substrate is increased in presence of a second substrate. Such is the case with phenylalanine and ATP binding to the corresponding enzyme from *E. coli*[83,84].

5.2.2. *Kinetic parameters for cognate and non-cognate aminoacylation.* The general strain theory demands that binding energy is used partially to strain the substrate towards the transition state. A better substrate would use a larger fraction of the intrinsic binding energy to induce strain in the enzyme-substrate complex. On these grounds it could be predicted that the better the substrate the weaker will be the overall binding. Such a relationship can be observed with phenylalanyl-tRNA synthetase from yeast. Binding of various homologous and heterologous and cognate and non-cognate tRNAs has been compared to the overall specificity with which phenylalanine is charged to the various tRNAs (Figure 6). In Figure 6 $\log K_m$ is plotted versus $\log(V_{max}/K_m)$ since the two logarithms are linearly related to the free energy of binding, ΔG^B, and the free energy of equilibrium, ΔG^T. The latter determines the overall specificity (see equations 5 and 6). The increase in specificity roughly coincides with a decrease in affinity of the tRNAs from *E. coli*. There is a break in the relationship for the cognate $tRNA^{Val}$ from yeast and wheat. These two tRNAs are charged about two orders of magnitude more specifically than the tRNAs from *E. coli*. The jump in specificity is brought about partly by better binding but mainly by higher rates.

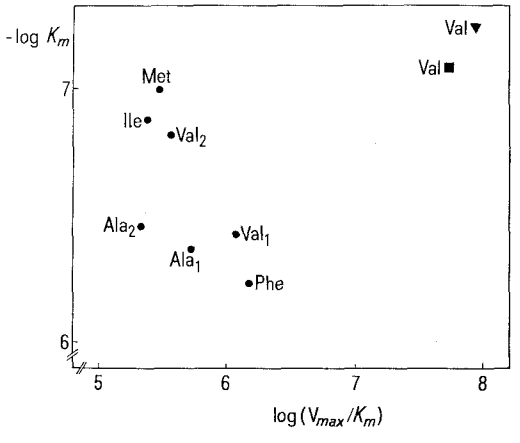


Fig. 6. Binding specificity and overall specificity of phenylalanyl-tRNA synthetase from yeast. tRNAs from *E. coli* ●, wheat ■, and yeast ▼. Ala_1, Ala_2 etc. are iso-accepting tRNA species. K_m- and V_{max}-values, obtained at 30° and pH 8.2, are taken from ref.[85].

It is not known whether the compared K_m-values are true dissociation constants, but this does not affect the result as long as the mechanism of charging is the same for all tRNAs. The latter assumption is necessary, though still unproven.

In essence it seems that the general strain theory holds for the interaction of tRNA with phenylalanyl-tRNA synthetase from yeast. It has been concluded[62] from maximal velocities and Michaelis-constants of the charging of tRNAs by a single synthetase that changes in K_m are generally one to two orders of magnitude smaller than changes in V_{max}, in other words, specificity increase is expressed mainly in $\Delta\Delta G^{\dagger}$ and little in $\Delta\Delta G^B$.

5.2.3. *'Superspecificity' in synthetases.* I have argued above that the allowed error frequency of synthetases must be below 1:3000. Calculated error frequencies have been put as low as 1:10[62,63]. Can an enzyme achieve such a low error frequency if it has to make its choice solely on the basis of structural differencies between amino acids and tRNAs? The well known example is that of isoleucyl-tRNA synthetase discriminating between isoleucine and valine. These two amino acids differ only by one methylene group. In a much cited article PAULING[86] has emphasized many years ago that a methylene group can contribute at most 1.8 kcal/mol to the difference in free energy of binding given optimal fit between substrate and binding site. In terms of binding constants, this is a factor of 20. In the case of isoleucyl-tRNA synthetase catalyzing the formation of isoleucyl- and valyladenylate, respectively, this factor of 20 may be achieved either solely in a difference in binding energies (pure binding specificity, $\Delta\Delta G^B$) or in a difference in rate of adenylate formation (pure kinetic specificity, $\Delta\Delta G^{\dagger}$) or in both. But the overall specificity of aminoacyladenylate formation should never differ by more than a factor of 20 or so. Another factor of 20 or more may be achieved in the specificity towards tRNA. Together this is quite insufficient to account for the predicted (see above) and observed specificity of isoleucyl-tRNA synthetase[87]. Even if the above numbers are grossly under-

[82] R. J. POLJAK, L. M. AMZEL, B. L. CHEN, R. P. PHIZACHERLEY and F. SAUL, Proc. natn. Acad. Sci. USA *71*, 3440 (1974).
[83] H. M. KOSAKOWSKI and E. HOLLER, Eur. J. Biochem. *38*, 274 (1973).
[84] E. HOLLER, B. HAMMER-RABER, T. HANKE and P. BARTMANN, Biochemistry *14*, 2496 (1975).
[85] B. ROE, M. SIROVER and B. DUDOCK, Biochemistry *12*, 4146 (1973).
[86] L. PAULING, *Festschrift Arthur Stoll* (Birkhäuser Verlag Basel 1958), p. 597.
[87] F. H. BERGMANN, P. BERG and M. DIECKMANN, J. biol. Chem. *236*, 1735 (1961).

estimated, in other words even if discrimination between valine and isoleucine on the basis of their structure is possible by a much larger factor than has been calculated[86], one might search for other means than the differences in free energies between substrates which would empower synthetases to achieve 'superspecificity'. I am going to discuss three of them.

5.2.4. *Competition among synthetases in vivo.* A possibility not obvious from in vitro systems dealing with single enzymes is that in vivo the whole collection of different tRNAs and synthetases is available simultaneously. Thus each synthetase may be complexed with its cognate tRNA and amino acid. This reduces the concentration of free enzyme and free tRNA available for mischarging as opposed to the in vitro system where a single synthetase is being offered a collection of tRNAs[62,88]. A simple calculation illustrates the point. The intracellular concentration of $tRNA^{Ile}$ and isoleucyl-tRNA synthetase in *E. coli* has been estimated[89] to 3×10^{-6} *M* and 5×10^{-7} *M*, respectively. A binding constant of 1.4×10^{-6} *M* for $tRNA^{Ile}$ to the isoleucyl-tRNA synthetase has been found[89]. Two systems will be compared: a) $tRNA^{Val}$ + Ile-tRNA synthetase; b) $tRNA^{Val}$ + $tRNA^{Ile}$ + Ile-tRNA synthetase. In system a) only misacylation occurs. Its rate shall be arbitrarily set to one. In system (b), however, 99.44% of the isoleucyl-tRNA synthetase are calculated to be in complex with the cognate $tRNA^{Ile}$. Therefore misacylation which is proportional to the concentration of free enzyme, is reduced by a factor of about 200 in system (b). If valyl-tRNA synthetase is added to system (b), the concentration of free $tRNA^{Val}$ will also be reduced which diminishes the rate of misacylation even further.

It is quite possible that in vivo the actual concentrations and binding constants are occasionally more or less in favour of such a concentration-dependent reduction in misacylation[62]. But in principle this effect of the multitude of enzymes and substrates is certainly a device to reduce the frequency of misacylation. The overall specificity of the individual synthetase has, of course, not been increased.

5.2.5. *Verification mechanisms.* A verification mechanism has been proposed in which tRNAs cognate to a given synthetase but carrying the wrong amino acid, are rapidly deacylated by hydrolysis in the absence of other substrates[90].

$$\text{Ile} + \text{ATP} + \text{tRNA}^{\text{Phe}} \xrightarrow{\text{PheRS}} \text{Ile-tRNA}^{\text{Phe}} \qquad (13)$$

$$\text{Ile-tRNA}^{\text{Phe}} \xrightarrow{\text{PheRS}} \text{Ile} + \text{tRNA}^{\text{Phe}} \qquad (14)$$

Thus, if phenylalanyl-tRNA synthetase (PheRS) has erroneously charged the cognate $tRNA^{Phe}$ with isoleucine (reaction 13), the enzyme is able to correct this error by catalyzing the hydrolysis of Ile-$tRNA^{Phe}$ via reaction (14). It has to be stressed that reaction (14) cannot be simply the reversal of reaction (13), otherwise there would be no increase in specificity on top of that achieved already in reaction (13).

Several synthetases possess the ability to catalyze hydrolysis of mischarged cognate tRNAs. It was even postulated that hitherto undetected enzymes genetically related to synthetases might account for verification in vivo[91]. Deacylation is not very specific, either with respect to tRNA or to the amino acid, and deacylation of a mischarged tRNA is not always faster than that of a correctly charged tRNA. Consequently the verification hypothesis has been questioned as an important contribution to 'superspecificity'[63].

5.2.6. *Do synthetases proofread?* The mechanism of aminoacylation by synthetases apparently has all the requirements required by the proofreading hypothesis[25]. The following scheme was devised which is analogous to scheme (10):

$$\begin{array}{l} \text{E} + \text{aa} + \text{ATP} \overset{a}{\rightleftharpoons} \text{aa(E}\cdot\text{ATP)} \overset{b}{\rightarrow} \text{(aa}\sim\text{AMP) E} \overset{d}{\rightarrow} \text{aa-tRNA} \\ \qquad\qquad\qquad\qquad\quad \downarrow \text{PP}_i \qquad\quad \downarrow c \\ \qquad\qquad\qquad\qquad\quad \text{E} + \text{aa} + \text{AMP} \end{array} \qquad (15)$$

Reaction *a* is specific and reaction *b*, as found experimentally[57], is unspecific with respect to the amino acid (aa). As demanded by the theory, reaction *b* is strongly driven by the hydrolysis of ATP. In addition, low intracellular concentration of inorganic pyrophosphate (PP_i) favours equilibrium to the side of the enzyme-aminoacyladenylate complex, (aa~AMP)E. Step *c* is also known to occur and would have a functional purpose according to the theory. Probably step *c* is irreversible, due to the instability of the aminoacyladenylate, (aa~AMP), in aqueous medium. Specificity is achieved in steps *a* and *c*. In particular it is assumed that incorrectly formed aminoacyladenylate is preferentially removed via step *c*. Alternatively a mechanism is feasible where the second specificity determining step, the proofreading step, is in reaction *d*.

[88] M. YARUS, Nature New Biol. *239*, 106 (1972).
[89] M. YARUS and P. BERG, J. molec. Biol. *42*, 171 (1969).
[90] M. YARUS, Proc. natn. Acad. Sci. USA *69*, 1915 (1972).
[91] M. YARUS, Nature New Biol. *245*, 5 (1973).

If proofreading occurs in step *c*, the theory demands further that reaction *d*, which is a composite and could be expanded, is slow with respect to steps *a* through *c*, in other words the enzyme-aminoacyladenylate complex should build up in the pre-steady state of the reaction. Several studies indicate that the rate limiting step in the overall reaction is release of aminoacyl-tRNA from complex with the synthetase[89,92], a result compatible with kinetic proofreading.

A concerted mechanism, in which no intermediate is ever formed between substrate binding and product formation, i.e. in which the overall reaction catalyzed by a synthetase cannot be divided into reactions (11) and (12), is incompatible with the proofreading hypothesis. Such a mechanism might apply for some synthetases[60,93]. It is, however, difficult to rule out pre-steady state formation of aminoacyladenylate convincingly[94]. There are indications that tyrosyl-tRNA synthetase from *B. stearothermophilus* operates by both mechanisms, step-wise and concerted, depending on the prevailing substrate concentrations[94].

One crucial experiment to test kinetic proofreading has to my knowledge not yet been done, and that is to determine the amount of ATP hydrolyzed per molecule of incorrectly formed aminoacyl-tRNA. Not suprizingly, there is one to one stoichiometry in case of the cognate charging, that is one molecule of ATP is consumed per molecule of correct aminoacyl-tRNA produced[94]. Therefore reaction *c* in (1.5) is insignificant for the correct aminoacylation as forescasted by the theory. But whether this is so for misacylation remains to be shown.

5.2.7. *Negative cooperativity and specificity increase.* Two non-equivalent binding sites for tRNA and amino acid have been found on several synthetases (section 5.2.1.). Could this help to advance specificity above the limiting value inherent in the structural differences between competing substrates, i.e. to achieve 'super-specificity'? The answer is *no* on the grounds given above (sections 3.2.5. and 5.2.3.), namely that in a single-site and in a two-site enzyme overall specificity at each site ($\Delta\Delta G^T$ in equation 6) is limited by the structural differences between enzyme and the competing substrates. As an example, tyrosyl-tRNA synthetase discriminating between tyrosine and phenylalanine might be considered. At saturating amino acid concentrations, the complexes E.Tyr and E.Phe apply to the one-site enzyme and E.Tyr.Tyr, E.Phe.Phe and E.Tyr.Phe to the two-site enzyme. In both cases correct tyrosyladenylate is formed at half of the sites and wrong phenylalanyladenylate at the other half. Specificity is governed only by the different rates of formation of tyrosyl- and phenylalanyl-adenylate at the different sites, which is in turn given by $\Delta\Delta G^{\dagger}$. Nothing alters if subsaturating amino acid concentrations prevail or if there is cooperation between the sites of the two-site enzyme.

Extra binding energy from the second binding site could conceivably be used to drive adenylate formation at the first site, e.g. to strain the amino acid at the first site toward the transition state on the path to the adenylate, as proposed by the general strain theory. The two-site enzyme would therefore have a catalytic advantage over the one-site enzyme. But in this case enhanced catalysis does not entail enhanced overall specificity. There is no asset for specificity in two sites.

6. *Conclusion*

There is at present no theory which would embrace all the various facets of specificity that have been observed experimentally. I have discussed those hypotheses only which seem most promizing to me. The future task will be to conceive experiments which are still better suited to test the various theories. As for the synthetases it will be most important to determine more accurately the intracellular concentrations of enzymes and substrates. Then, a careful comparison of correct and incorrect acylation by a specific enzyme with regard to enzyme mechanism must be conducted. Lastly, it will be crucial to assess the limit of error bearable by in vivo protein biosynthesis.

To sum up we may say that the selectivity of less specific enzymes no longer poses insurmountable conceptual difficulties. However, understanding the stupendous accuracy of some biological processes, for example the transcription of a gene or the ribosomal protein synthesis, is still a matter of speculation, and we have but few experimental data that could help to corroborate and improve existing theories[95].

[92] E. W. Eldred and P. R. Schimmel, Biochemistry *11*, 17 (1972)

[93] T. N. E. Loevgren, J. Heinonen and R. B. Loftfield, J. biol. Chem. *250*, 3884 (1975).

[94] A. R. Fersht and R. Jakes, Biochemistry *14*, 3350 (1975).

[95] I am grateful to many of my colleagues for critical advice and constructive comments to this article. My particular thanks go to Gordon L. E. Koch for his linguistic help.