To build an enzyme...

JEREMY R. KNOWLES

Departments of Chemistry and Biochemistry, Harvard University, Cambridge, Massachusetts 02138, U.S.A.

SUMMARY

The structural components that lead to enzyme function are discussed for one simple enzyme-catalysed reaction: that mediated by triosephosphate isomerase. First, the recognition and binding of the substrates' phospho group is seen to involve four main-chain —NH— hydrogen bonds, two of which are positioned at the positive end of a short α-helix aimed precisely at the phospho group and interact with the three peripheral phospho group oxygens. Second, the chemical steps (of substrate enolization) are shown to require both base and general acid catalysis. The identity and the positioning of the base, a carboxylate group, nicely fulfils the expectations both of mechanistic economy and of stereoelectronics. The identity of the general acid is shown by Fourier transform infrared and by 15N nuclear magnetic resonance (NMR) to be a neutral imidazole group, lying between the two substrate oxygens. The positioning of the ring is ideal, but its protonation state is unexpected. Thus the pKa of this histidine side-chain is < 4.5, lowered from 6.5 (the value in the denatured protein) by its position at the positive end of another well-aimed α-helix. Third, the need for enzymes to provide kinetic barriers to the loss of reaction intermediates from the active site is emphasized. Triosephosphate isomerase achieves this sequestration of the reaction intermediate by using a flexible loop of the protein, and thus improves the efficiency of the catalysed transformation.

1. INTRODUCTION

There are two views of the catalytic power of enzymes. The first is that the catalytic rates achieved by enzymes are extraordinary, that enzymes represent the rare end products of an extensive search through protein sequence ‘space’ (Maynard-Smith 1970), and, perhaps, that they use mechanistic and kinetic devices that have not yet been recognized by physical-organic chemists. The second view is less awesome, and is that each of today’s enzymes represents one of a large number of more-or-less equally effective possible solutions to the catalytic problem, each of these solutions using functional groups and other catalytic elements in ways that are consonant with the expectations of mechanistic organic chemistry. In this paper, we examine the validity of the second view by dissecting the catalytic apparatus of an enzyme that mediates a chemically simple reaction in the glycolytic pathway: triosephosphate isomerase.

Triosephosphate isomerase catalyses the interconversion of the two triose phosphates, dihydroxyacetone phosphate and D-glyceraldehyde 3-phosphate. The overall reaction simply involves the enolization of substrate to the enediol (or enediolate) intermediate, which then re-ketonizes to form the product (Rieder & Rose 1959; Rose 1962). The enzyme thus mediates two enolizations, back to back, as shown in figure 1. Already in figure 1 we have shown two catalytic groups, a base B to abstract the carbon-bound proton and a general acid H-A to assist in this process. Before any consideration of active site chemistry, however, let us look at the first step of the catalysed process, in which the substrate is recognized and binds to the active site.

2. SUBSTRATE BINDING

In the case of the two triose phosphates, the obvious ‘handle’ for substrate binding is the phosphate ester, which indeed seems to dominate the recognition process. Thus while many phosphate ester analogues of the substrates are competitive inhibitors of the enzyme (these range from molecules that bind relatively weakly such as glycerol 1-phosphate and 3-phosphoglycerate (Collins 1974)) to more impressive inhibitors such as phosphoglycolate and phosphoglycololhydroxamate (Collins 1974)). Any change in the phospho group itself (as in dihydroxyacetone sulphate (Belasco et al. 1978)) results in molecules that are not recognized by the enzyme and do not bind to it. From nuclear magnetic resonance (NMR) experiments and pH-variation studies (Campbell et al. 1978; Belasco et al. 1978), it is known that the phospho group binds as a dianion, and it is instructive to see how the enzyme creates a binding locus for this group. This is shown in figure 2. No cationic amino acid side chains (of lysine, arginine or histidine) are involved, the peripheral oxygens of the phospho group being held by at least four main-chain —NH— hydrogen bonds (Davenport 1986; Lolis & Petsko 1990). As Pauling pointed out many years ago (Pauling 1960), a main-chain —NH— bond is ‘worth’ about half a charge, and we may expect that the arrangement shown in figure 2...
effectively balances the dianionic charge of the phospho
group. Yet if electrostatic neutrality was all that was
important, why did the enzyme not use lysine or
arginine side chains? A persuasive answer to this
question lies in the evident need (which will become
apparent below) for precision in the positioning of
substrate with respect to the catalytic groups. The side
chains of lysine and arginine are both flexible, and
especially all side chains in proteins have greater
freedom and mobility than the atoms of the main
chain. Indeed it seems unlikely that a binding locus of
such constraining exactness as that shown in figure 2
could easily be created by using the floppier side chains
of lysine or arginine. (These views are strongly tinged
with hindsight, of course, for there exist many examples
where phospho group binding sites do contain lysine,
arginine, or histidine side chains (see, for example,
Cotton et al. (1979); Sowadski et al. (1985); Luecke &
Quirocho (1990)). Yet it is true that even in these cases,
the opportunity for multiple hydrogen bonds almost
always ties the side chain cation back to the main-
chain framework.)

There are two other features of figure 2 that deserve
comment. Two of the main-chain hydrogen bonds
derive from glycine residues 232 and 233 that lie at the
positive end of a short α-helix that is beautifully aimed
at the substrates' phospho group (figure 3b). It is hard
to imagine that the precision of this alignment is
accidental, and even if the effect of an α-helix dipole
(Hol 1985) is, as has been suggested, predominantly
because of the alignment of the hydrogen bond donors
and acceptors in the final turn, the arrangement
appears to be a particularly sturdy one. The second
important feature of figure 2 is that glycine-171 lies on
a flexible loop of residues (from 166 to 176) that closes
down over the active site when the substrate binds.
The mechanistic function of this loop movement and the
catalytic consequence of the recruitment of an ad-
ditional group to bind the substrate, are discussed
later. Suffice it here to conclude that the substrates'
phospho group seems to be nicely accommodated by
the constellation of hydrogen bond donors shown in
figure 2.

3. THE FIRST CHEMICAL STEP

The covalency changes that produce the enediol(ate)
intermediate occur within the first-formed Michaelis
complex. Enolizations are both acid- and base-
catalysed, and the kinetic benefit from the concerted

action of both has been well documented (Hegarty &
Jencks 1975). Certainly the organic chemist would be
drawn to propose a base and a general acid arranged
as shown in figure 1. It is, therefore, gratifying that
from the high resolution crystal structures of triose-
phosphate isomerase containing bound substrate or
substrate-like inhibitors (Banner et al. 1975; Alber et al.
1981; Lolis et al. 1990; Lolis & Petsko 1990), a base
(the carboxylate of glutamate-165) and a possible acid
(the imidazole of histidine-95) are nicely placed to
mediate the enolization. The nearer carboxylate
oxygen of glutamate-165 is 2.8 Å and 3.4 Å from C-
1 and C-2 of the substrate, respectively, thus poised for

\[ 1 \text{ Å} = 10^{-10} \text{ m} = 10^{-1} \text{ nm}. \]
the abstraction either of the C-1 pro-R proton of dihydroxyacetone phosphate or of the C-2 proton of glyceraldehyde phosphate (whichever triose phosphate substrate binds). The use by the enzyme of a bidentate base seems particularly appropriate, considering the reaction to be catalysed (figure 1). The base is properly positioned in stereoelectronic terms too, being orthogonal to (what becomes) the plane of the enediol that is, the plane defined by O-1, C-1, C-2 and O-2) as required by Corey & Snee (1956), and using the syn orbitals of the carboxylate, following Gandour (1981). From all chemical points of view, therefore, the identity and position of the catalytic base in triosephosphate isomerase seems ideal. Unsurprisingly, such an attractive arrangement is not unique. One particularly well-studied parallel is the enzyme 3-ketosteroid isomerase (Kuliopulos et al. 1989), where the carboxylate of an aspartate residue is used analogously to effect the abstraction and replacement of carbon-bound protons (in a 1,3 shift, rather than the 1,2 relation in triosephosphate isomerase).

From the crystal structure of the liganded isomerase, it is clear that the imidazole ring of histidine-95 is an excellent candidate for the general acid H-A. The N\(\text{\textdegree}\) of this residue is almost equidistant from the substrate oxygens O-1 and O-2 (the distances are 2.8 \(\AA\) and 2.9 \(\AA\), respectively), well placed in the plane of what will be the enediol and at a distance to make appropriately strong hydrogen bonds to substrate. Once again the positioning seems to be ideal, according nicely with the prejudices of the physical-organic chemist who would like to catalys[e] an enolization. Just as early work by chemical modification had anticipated the existence of glutamate-165 as the isomerase's catalytic base (Hartman 1968; Waley et al. 1970; de la Mare et al. 1972), early chemical and spectroscopic experiments (Webb & Knowles 1975; Belasco & Knowles 1980) had suggested the existence of a general acid that polarizes the carbonyl group of enzyme-bound substrate. Thus the infrared stretching frequency of the carbonyl group of dihydroxyacetone phosphate is moved by 19 cm\(^{-1}\) to lower frequency when this substrate binds to the isomerase (Belasco & Knowles 1980). That this polarizing shift is due to the imidazole ring of histidine-95 has recently been supported by infrared studies on the substrates bound to two mutant isomerases H95N and H95Q, in which histidine-95 has been changed either to asparagine or glutamine. Neither of these mutant enzymes (the catalytic potency of which are 104- and 102-fold less than the wild-type enzyme, respectively, Nickbarg et al. 1988) polarizes the substrate's carbonyl group (Komives et al. 1991), and these data are entirely consistent with the indications from the protein crystal structure that histidine-95 is the catalytic electrophile.

At this point, the chemical reader would naturally presume that histidine-95 would be in its protonated imidazolium form at the pH-values where the enzyme is catalytically active, so as to act as an effective electrophile. The \(pK_a\) of a histidine side chain is normally somewhat less than 7, and the imidazolium cation is (with \(pK_a \approx 7\)) obviously a much stronger electrophile than neutral imidazole (with \(pK_a \approx 14\)). Yet these comforting chemical statements were not consistent with the crystallographic results, which suggested that the imidazole ring of histidine-95 is neutral (Browne et al. 1976). This conclusion was based upon the fact that the distal imidazole nitrogen, \(N^\circ\), lies within hydrogen bond distance of the main-chain –NH– group of residue 97. Clearly if \(N^\circ\) acts as a hydrogen bond acceptor, the imidazole ring cannot be protonated.

To resolve the question, we have investigated the protonation (and hydrogen bonding) state of the two nitrogens of the imidazole ring of histidine-95, by \(^{15}\text{N}\) NMR. To simplify the spectrum and to eliminate the need for resonance assignment, we took the yeast enzyme (which has three histidine residues, at positions 95, 103, and 185) and replaced the two histidines (at positions 103 and 185) in which we had no interest. The resulting double-mutant enzyme H103Q•H185Q is mechanistically and kinetically indistinguishable from the wild type, and the only histidine in the molecule is that at the active site. The mutant enzyme was produced in an auxotrophic his\(^{+}\) Escherichia coli host, grown on singly or doubly \([^{15}\text{N}]\)-labelled histidine. As can be seen from figure 4, the chemical shift of the \(^{15}\text{N}\) resonances of imidazole nitrogens is exquisitely sensitive to protonation state and participation in hydrogen bonds (Bachovchin 1986) and has allowed the nature and environment of the side chain of histidine-95 to be probed (P. Lodi, unpublished data). A typical spectrum from the doubly \([^{15}\text{N}]\)-labelled yeast isomerase is shown in figure 5. From these data, it is evident that, contrary to mechanistic expectation, the imidazole ring of histidine-95 is not protonated at neutral pH. Indeed, as is shown by the insert in figure 5, the ring remains unprotonated down to at least pH 5. The \(pK_a\) of histidine-95 must be less than 4.5! Also evident from

![Figure 4. Typical \(^{15}\text{N}\)NMR chemical shifts for imidazole nitrogens in different environments. All values are upfield (negative) from \(1 \text{m HNO}_3\). The values in bold type are those observed for the double mutant (H103Q•H185Q) yeast triosephosphate isomerase containing ring-labelled \([^{15}\text{N}]\)-imidazole-95, either alone ('unliganded') or in the presence of saturating levels of phosphoglycolohydroxamate ('plus PGH').](http://rstb.royalsocietypublishing.org/)}
the experimental results shown (in bold face) in figure 4, is the fact that when the substrate analogue phosphoglycolohydroxamate binds, the $N^\circ$ of histidine-95 becomes a hydrogen-bond donor. We may conclude, based upon this nmr result (P. Lodi, unpublished data) and the crystal structure of the yeast enzyme with bound inhibitor (Davenport 1986; Lolis & Petsko 1990), that a strong hydrogen bond (stronger, that is, than to any water molecules that occupy the active site in the absence of ligand) is indeed formed between $N^\circ$ and the substrate analogue's carbonyl oxygen.

Two questions are raised by these results. First, how is the $pK_a$ of histidine-95, which we have determined to be 6.5 in the denatured protein, lowered by more than two units in the native enzyme? Inspection of the structure provides an immediate answer, that is shown in figure 3a. The imidazole ring of histidine-95 lies at the positive end of a short $\alpha$-helix, and $N^\circ$ is within hydrogen-bonding distance of two main-chain $-NH-$ groups, of residues 96 and 97 (at 2.7 Å and 2.9 Å, respectively). As has been noted before (Hol 1985; Perutz et al. 1985), and most clearly shown by the work of Fersht and his group on barnase (Sali et al. 1988), such a local environment readily accounts for the lowering of the $pK_a$ of a histidine residue. The second question is more difficult and more teleological: why should the enzyme go to such lengths to use imidazole, a lesser electrophile, in preference to imidazolium? (This question presumes, pending the completion of analogous but technically more difficult experiments with substrate itself, that DHAP and phosphoglycolohydroxamate will have similar effects on the $^{15}$N nmr of the labelled isomerase.) This issue cannot be usefully analysed without more information, though several possibilities can be mentioned. Thus (i) the juxtaposition of two oppositely charged catalytic groups (a carboxylate and an imidazolium) might lead to an ion pairing interaction that would preclude the proper positioning of these groups for catalysis; (ii) the existence of a local imidazolium might lower the basicity of the carboxylate and thus slow the enolization reaction; (iii) since the $pK_a$ of imidazole is known to follow (some seven units away) the $pK_a$ of imidazolium (Bruice & Schmir 1958; Walba & Isensee 1961), it may be that the enzyme lowers the $pK_a$ of imidazole to a value close to that of the enediol intermediate, thus allowing for very rapid proton transfers between O-1 and O-2, and (iv) there may be some unsuspected mechanism by which an imidazolium can catalyse the transfer of protons between the oxygens of a cis-enediolate. Whatever the reasons behind the enzyme’s use of a neutral imidazole as the electrophile, it is clear that this represents a gap in our understanding.

4. THE NEED FOR SEQUESTRATION OF INTERMEDIATES

Any enzyme-catalysed reaction that involves a reaction intermediate en route from substrate to product, faces the problem of avoiding the loss of the intermediate from the active site. As shown in figure 6a, the substrate (or the product) can, when enzyme-bound, partition in two ways: forward, involving the chemical changes of the reaction, or backward, involving the diffusive loss from the active site. These alternatives are necessary and sufficient for all one-substrate enzymic processes. (Extension of these arguments to multi-substrate reactions having either random or ordered binding, is straightforward.) Consider, however, the reaction intermediate, which, as shown in figure 6b, must partition in two directions only, each of which involves the chemical changes of the catalysed reaction: forward to GAP or backward to DHAP. The notational third route, of departure from the active site, must be blocked, so as to ensure a stoichiometric ‘throughput’ of substrate to product. In the case of triosephosphate isomerase, for example, it is known that in free solution the enediol intermediate...
Figure 6. Free energy profiles for the action catalysed by triosephosphate isomerase, showing the need for sequestration of the reaction intermediate. (a), how the complexes of either of the substrates dihydroxyacetone phosphate (DHAP) or R-glyceraldehyde 3-phosphate (GAP), with the enzyme (e), may partition in two ways, either forward (to involve chemistry) or back (to fall off the enzyme); (b), the complex of the reaction intermediate (enediol) with the enzyme (e), must partition only forward or back involving chemistry, and not fall off the enzyme; (c), showing the two ways in which the enzyme can avoid losing its grip on the reaction intermediate, either by the relative thermodynamic stabilization of the enzyme–enediol complex, or by the creation of a kinetic barrier to the loss of the enediol from the active site.

In common with many enzymes, triosephosphate isomerase has a flexible loop of ten amino acid residues (that have recently been more aptly described as a lid: Joseph et al. (1990)) that appears to close over the substrate during the catalytic act. Crystallographic analysis had first indicated that this lid adopts a ‘closed’ position when substrate or a substrate analogue occupies the active site (Banner et al. 1975; Alber et al. 1981), and it was then conjectured that the conformation change (which amounts to induced fit (Koshland 1959), even if new catalytic functionalities are not thus recruited) could prevent loss of the enediol(ate) intermediate from the enzyme. It had been early recognized that the enzyme must (and does) bind its substrates so as to keep the phospho group more or less in the plane of the enediol(ate). Such a conformation steroelectronically disfavours the decomposition of the intermediate to methylglyoxal and P_i, a reaction that, as noted above, is normally much faster than C-protonation (Richard 1984). This facile decomposition makes it especially important that the intermediate not be released. (The consequences of intermediate loss are not, of course, always disastrous, for the intermediate may be transiently stable, too (Pompliano et al. 1990)). The abbreviated loop (which lacks, inter alia, glycine-171, see figure 2) can no longer encapsulate the carbanionic intermediate from amino acid racemases, or the dienolate made by Δ^2-3-ketosteroid isomerases.)

To test these ideas, we deleted four amino acid residues from the lid of triosephosphate isomerase, choosing a segment that minimized any non-local changes in the enzyme’s structure (Pompliano et al. 1990). The abbreviated loop (which lacks, inter alia, glycine-171, see figure 2) can no longer encapsulate the substrate, and model building suggests that the active site of this mutant enzyme is at all times open to solvent. The purified ‘loopless’ mutant enzyme is a much less effective enzyme. Values of $k_{cat}$ are some 10-fold lower than the wild-type, although $K_M$ values rise by less than tenfold. Evidently the transition states for the chemical steps have been most sharply affected, and from the much weaker binding of an analogue of the enediol(ate) intermediate, it seems that the enzyme–intermediate complex is relatively much less stable, too (Pompliano et al. 1990). The first and most
obvious role for the lid residues, then, is preferentially to stabilize the enediol(ate) intermediate (as distinct from the substrate or the product) and the two enolization transition states that flank it. Yet the second function (which is relevant to the question of intermediate sequestration discussed above), to provide a kinetic barrier to the loss of the reactive intermediate, is also evident. Indeed, when the stoichiometry of substrate-product conversion was investigated, we found that for six molecules of glyceraldehyde phosphate consumed by the loopless mutant isomerase, only one molecule of dihydroxyacetone phosphate was formed. The other five molecules ended as methylglyoxal and Pi, the presumed consequence of the loss of the intermediate from the enzyme and its rapid decomposition in free solution.

It thus appears that triosephosphate isomerase has neatly and economically used a flexible loop of the protein, both to speed the reaction by preferential stabilization of the reaction intermediate and of the transition states that lead to it, and to maximize the conversion of substrate to product by creating a kinetic barrier to the loss of the intermediate from the active site.

The contributions of many collaborators, especially Steve Blacklow, Elizabeth Komives, Patricia Lodi, Elias Lolis, Elliott Nickbarg, Greg Petsko, Anusch Peyman, David Pompianio and Ron Raines, and the financial support of the National Institutes of Health, is gratefully acknowledged.

REFERENCES


Phil. Trans. R. Soc. Lond. B (1991)


Introduction

Cyclic sequencings are nucleotide determination and molecular evolution. The DNA sequence has to be determined by a method different from nucleic acid sequence. In this case, a single strand of DNA is used for the determination of nucleotide sequence. To determine the sequence of DNA, the double strand must be cleaved by a restriction enzyme. The cleavage occurs at the specific base pair. The two strands are then separated and each strand is sequenced independently. The sequence of the DNA is then determined by the combination of the two strands.

One of the most fundamental principles is the sequence of the DNA strand. In the basic structure of DNA, there are two complementary strands. The one strand has the sequence of the other strand. The two strands are then combined to form the double strand. The sequence of the DNA is then determined by the combination of the two strands.

This was the basis of molecular evolution. The DNA sequence is determined by the combination of the two strands. The sequence of the DNA is then determined by the combination of the two strands. The sequence of the DNA is then determined by the combination of the two strands. The sequence of the DNA is then determined by the combination of the two strands.