

Structural Origins of the
Catalytic Power of Triose Phosphate Isomerase

by

Thomas Clifford Alber

A. B., University of California at Santa Cruz

(1976)

SUBMITTED TO THE DEPARTMENT OF
BIOLOGY
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE
DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 1981

© Thomas Clifford Alber, 1981
The author hereby grants to M.I.T. permission
to reproduce and distribute copies
of this document in whole or in part.

Signature of Author _____
Department of Biology
September 1981

Certified by _____
Gregory A. Petsko
Thesis Supervisor

Accepted by _____
Archives
Chairperson, Departmental Graduate Committee
Mary Lou Pardue

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

DEC 22 1981

1

LIBRARIES

Structural Origins of the
Catalytic Power of Triose Phosphate Isomerase

by

THOMAS CLIFFORD ALBER

Submitted to the Department of Biology
on September 9, 1981 in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in
Biology

ABSTRACT

The goal of this project is to determine the structures of all the kinetically significant intermediates in an enzymatic reaction. Triose phosphate isomerase (E.C. 5.3.1.1 (TIM)) is a dimeric protein with a molecular weight of about 53,000. It catalyzes the interconversion of dihydroxyacetone phosphate (DHAP) and glyceraldehyde-3-phosphate (GAP) in glycolysis. The rate of the catalytic reaction is almost diffusion controlled, making TIM one of the most efficient enzymes in living things.

In collaboration with Dr. G. Kawasaki, the sequence of the gene coding for yeast TIM was determined using the methods of Maxam and Gilbert. The amino acid sequence deduced from the order of nucleotides is approximately 50% homologous with TIM's from rabbit, chicken, and coelacanth and 37% homologous with the Bacillus stearothermophilus enzyme. The residues which are thought to be catalytically important are conserved. The coding region contains no introns. The nucleotide sequences which specify three helical regions of the protein are approximately homologous. However, since these sequences are translated in the same reading frame, their presence does not provide strong evidence for genetic duplication of domains during the evolution of the protein. A deletion of 3000 base pairs in the 5' flanking sequence reduced the expression of the gene 10-fold in yeast and enhanced its expression in *E. coli*. A comparison of the mutant and wild type nucleotide sequences provides clues about the basis of the efficient expression of the gene in yeast.

The x-ray crystal structure of yeast TIM was determined at 3.0 Å resolution using the method of multiple isomorphous replacement. Each monomer contains a core of eight strands of parallel β-sheet arching around the surface of a cylinder with a right handed twist. This β-barrel is surrounded by twelve α-helices which form a secondary cylinder on the surface of each subunit. Irregular loops connect the secondary structural features. The loop containing residues 70 to 80 makes many of the intersubunit contacts, and residues 168 to 177 are disordered in the free enzyme. The structure is similar to that of chicken TIM, despite only 50% homology of their amino acid sequences. X-ray diffraction data were collected to 1.9 Å resolution, and

crystallographic refinement of the structure is in progress. These studies will facilitate a detailed comparison of the yeast and chicken enzyme structures.

Using a flow cell, crystals of the enzyme were exposed to the substrate, DHAP, at -5° C, and x-ray diffraction data were collected to 3.5 Å resolution. At this temperature an equilibrium mixture of DHAP and GAP formed in the crystal. Difference maps revealed the substrate in both active sites of the dimer. The loop which was disordered in the free enzyme bound the substrate and became ordered in the E-S complex. Several residues in the loop moved over 10 Å on substrate binding. These results support model building studies on the interactions of the enzyme with its substrates.

These studies are consistent with more than one hypothesis about the enzyme mechanism. Glu 165 probably acts as a catalytic base which shuttles protons between C-1 and C-2 of the substrates. Catalysis due to electrophilic strain of the ground state of the substrates and stabilization of the transition states is provided by His 95 at the positive terminus of a helix dipole and possibly by Lys 13. These groups may act electrostatically to stabilize the negatively charged transition states or as general acids which aid in the formation and breakdown of the enzyme-bound intermediate. The phosphate moiety of the substrates is bound at the positive end of a helix dipole by Gly's 232 and 233, by residues 209 to 211 and by interactions with residues 171 to 173 in the disordered loop. Lys 13 may also form part of the phosphate binding site. The interaction between the substrate and the disordered loop accounts for some of the specificity of the reaction, inhibits the non-enzymatic decomposition of the enzyme-bound intermediate and raises the energy of the enzyme-substrate complex to speed the release of products. Higher resolution crystallographic studies and the construction of site specific mutants have been undertaken to test these ideas.

Yeast TIM was also crystallized in the presence of the putative transition state analogue, phosphoglycolate. These crystals have unit cell dimensions which are very different from those of the native enzyme, and they diffract to at least 1.8 Å resolution. Thus, another biologically interesting conformational state of the enzyme may be studied crystallographically.

Appendices to the thesis outline theoretical studies on unfolded regions of protein binding sites as sources of negative entropy and the mechanism of protein secretion in eukaryotes.

TABLE OF CONTENTS

Title Page	1
Abstract	2
Table of Contents	4
Chapter I. Introduction	5
Chapter II. Nucleotide Sequence of the Yeast Triose Phosphate Isomerase Gene	56
Chapter III. Crystal Structure of Yeast Triose Phosphate Isomerase at 3.0 Å Resolution	92
Chapter IV. Crystal Structure of a Michaelis Complex: Substrate Interactions with Triose Phosphate Isomerase	140
Chapter V. Conclusions	169
Appendix I. Energetic Consequences of Unfolded Regions in the Binding Sites of Proteins	191
Appendix II. Recognition of Eukaryotic Signal Sequences	220
Acknowledgements and Testimonials	240
Biographical Note	242

CHAPTER I.

INTRODUCTION

I. The Eyeteeth of Molecular Enzymology

The elucidation of the crystal structure of hen egg white lysozyme in 1964 marked the advent of molecular enzymology (1). This structure, and the many that have been solved since, have revealed the details of the conformations of a wide variety of enzymes, and, in concert with amino acid sequence information, this has inaugurated the study of enzymatic reactions in terms of the geometry and chemical properties of the participating groups. By the late 1960's, crystals of ribonuclease and γ -chymotrypsin were shown to be catalytically active (2,3). In their paper on chymotrypsin, Rossi and Bernhard concluded that "molecular inferences regarding the . . . mechanism of enzyme action, derived from crystallographic studies of acyl-chymotrypsins, are validly extendable to aqueous solutions" (3). This work has been extended to the reactions of a number of crystalline enzymes, and has led to a growing confidence that the packing of enzyme molecules in a crystal lattice does not necessarily obscure their fundamental catalytic properties (4).

So rich was the view through the crystallographic looking glass that Koshland and Neet, writing in the Annual Review of Biochemistry said, "Enzyme chemistry in the year 1968 has arrived at a particularly significant watershed. The determination of the three-dimensional structures of some enzymes has produced a 'moment of truth' for many of the past predictions of protein structure and function. It has also provided a glimpse onto a future in which structure determination, protein modification, and other new and sophisticated techniques may allow enzymologists to explain those twin miracles of enzyme catalytic power and enzyme specificity" (5). The truth about enzymes, however, has remained elusive, and crystallographic studies have allowed only more detailed questions to be framed. A large part of the problem turns on the absence of direct information about the events of the catalytic cycle, and the experiments described in this thesis are directed toward obtaining such information. Early crystallographic studies were necessarily indirect, because transient intermediates in enzymatic reactions could not be stabilized for the one to two week period required to collect x-ray data.

The standard methods which are used to obtain structural information about the interactions of enzymes and substrates were first applied in the study of hen egg white lysozyme. Phillips, Blake and coworkers solved the structures of complexes of hen egg white lysozyme with inhibitors and products, and they were able to observe the interactions of these ligands with the active site residues (6-8). Model building studies indicated that the hexasaccharide substrate could

not be accommodated in the active site unless the sugar adjacent to the labile glycosidic bond adopted a strained sofa conformation (8). On the basis of these studies, they proposed that destabilization of the ground state of the substrate brought about by steric strain may be central to the catalytic mechanism (9). Because this hypothesis was based on indirect inferences about the structure of the kinetically significant enzyme-substrate complexes, it has been susceptible to challenges by several groups who contend that electrostatic stabilization of the transient carbonium ion intermediate may make a more important contribution to the enzymatic enhancement of the reaction rate (10-12).

Nonetheless, structural studies of complexes of enzymes with poor substrates, nonproductive enzyme-inhibitor complexes and complexes of multiple-substrate enzymes with each each of their substrates in turn has led to a much better appreciation of the catalytic process. In particular, the conformational changes which can occur during catalysis, whose existence had been inferred from a large number of physical measurements were observed in detail, at least in shadow form. One surprising example was a change of 12 Å in the average position of the hydroxyl group of Tyr 248 which was seen when glycy-L-tyrosine was bound to carboxypeptidase A (13).

These results have been reviewed by Blow and Steitz (14) and Matthews (15) and will not be described in detail here. In general, the results are epistemologically suspect, for they try to deduce the progression of productive conformations through the study of non-productive states. It's like trying to learn how to hit home runs

by watching films of pop-ups, or, worse yet, in the cases where only the native enzyme structure has been determined, by merely observing that baseball is a game for 18 people, played with a ball and a bat on a field of a certain size, etc.

The roots of this problem were described in 1970 by David Eisenberg (16), "A limitation of crystallography, inherent in the phenomenon of diffraction, is that the resulting electron density function is an averaging over all the unit cells of the crystal and over the time of the experiment. This means, of course, that crystallography can yield very little direct information on the dynamics of an enzyme reaction ..."

II. Biting the Apple

Taking this point of view as a challenge rather than a truth, a number of workers in the ensuing decade found ways of directly observing productive enzyme-substrate complexes using x-ray crystallography. Petsko, Douzou, Fink and their coworkers demonstrated that it was possible to kinetically trap productive complexes of crystalline enzymes at low temperatures for a period of time long enough to collect x-ray data (4,17-20). The application of this methodology has led, in the last year, to the determination of the structure at atomic resolution of every kinetically significant enzyme-bound intermediate in the reaction catalyzed by ribonuclease A (W. Gilbert and G.A. Petsko, unpublished results). These data are currently the subject of intense scrutiny and analysis.

Another approach was taken by D.C. Phillips and coworkers. It was realized over ten years ago that because isomerase reactions are readily reversible, the equilibrium mixture of productive enzyme-substrate complexes could be observed crystallographically at ambient temperatures. This idea is illustrated in Figure 1. For the reaction of a hydrolytic enzyme, substrate which is soaked into an enzyme crystal will be rapidly converted to product and diffuse away from the active site. The lifetimes of transient enzyme-substrate complexes will be too short to contribute significantly to the x-ray diffraction pattern and will not be observed.

The situation is quite different in a crystal of an isomerase. Substrate which diffuses into the crystal is rapidly converted to product, but the product is the substrate for the back reaction. The substrates will be continuously interconverted, and, if the substrate concentration is higher than the dissociation constants, crystallographic analysis should, in theory, reveal the weighted average structure of the enzyme-substrate complexes present at equilibrium. This was the rationale behind the studies on chicken muscle triose phosphate isomerase (TIM) initiated over a decade ago in the laboratory of Molecular Biophysics in Oxford, England. For a number of technical reasons described in detail below, the crystals of the chicken muscle enzyme have proven especially difficult to study. In the hope of skirting these technical problems, Drs. Hartman, Petsko and Tsernoglou began crystallographic experiments on the triose phosphate isomerase

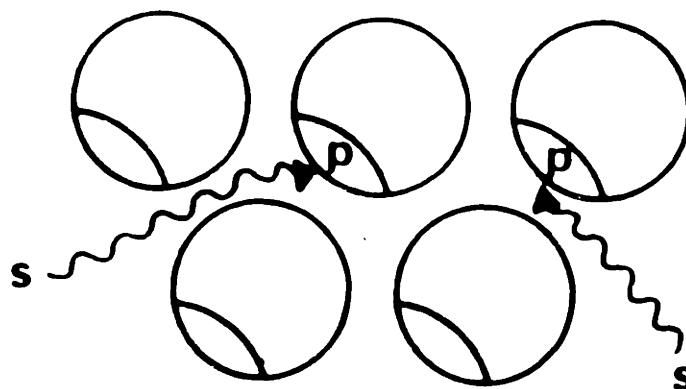
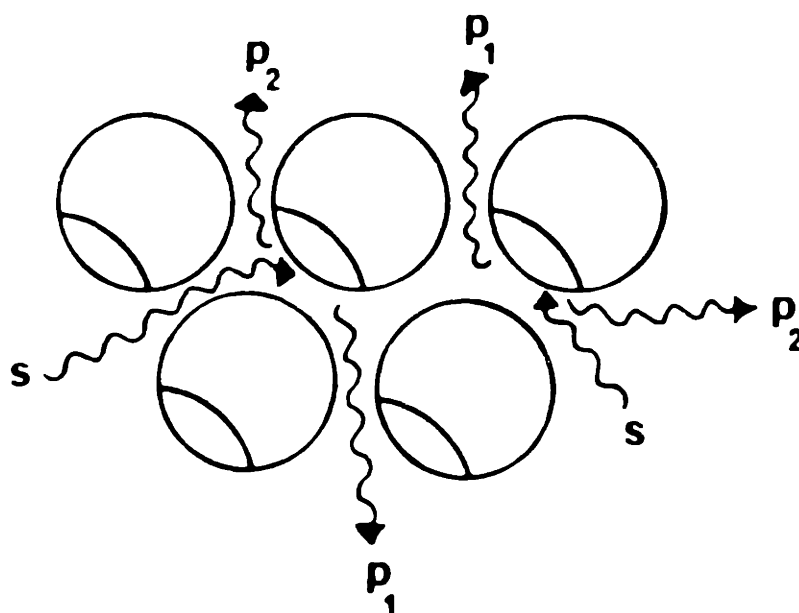
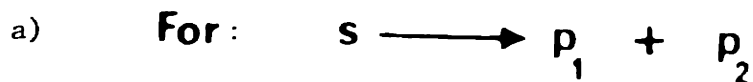


Figure 1. Reactions of Crystalline Enzymes. a) In a crystal of a hydrolytic enzyme, substrate is rapidly converted to product at ambient temperatures. The enzyme-substrate complex is short lived and cannot be observed by x-ray diffraction experiments. b) In a crystal of an isomerase an equilibrium mixture of enzyme-bound species will accumulate if the substrate concentration is above K_D . Though each molecule is rapidly interconverting the substrates, the average structure of all the enzyme-substrate complexes in the crystal will remain constant and can be observed directly by x-ray crystallographic methods.

(TIM) from yeast about three years ago. Structural studies on this enzyme form the bulk of the work described in this thesis.

III. Triose Phosphate Isomerase - The System (The Tree of Knowledge).

This work is directed toward understanding the structural basis of the specificity and rate enhancing ability of an enzyme. In addition to the fact that the productive enzyme-substrate complex should be directly observable, there are a number of features of the TIM reaction that make it an excellent model system. First, the reaction is extremely fast and simple. A 10^9 fold rate enhancement over the uncatalyzed reaction is brought about by the enzyme, making it one of the most efficient in living things (21,22) (Table 1). The reaction is almost diffusion-controlled, and the K_m 's are matched to the intracellular concentrations of substrates. Knowles has described it as an almost perfect catalyst, all but insensitive to evolutionary pressure (22). The insights gained into the mechanism of this fast enzyme may be more generally applicable to the reactions of those which are less efficient.

From a structural perspective, as well, TIM is a good candidate for study. The triose phosphate isomerases from four organisms have been sequenced (23-26), and the x-ray structure of the chicken muscle enzyme has been solved and recently refined at 2.5 Å resolution (27). Low resolution studies of its interactions with substrates and inhibitors have also been carried out (28). These investigations will permit a detailed comparison of the structures of the chicken and yeast enzymes

TABLE 1. Bimolecular Rate Constants for Product Formation*

Enzyme	Substrate	k_{cat}/K_m (sec ⁻¹ M ⁻¹)
Adenylosuccinate lyase	Adenylosuccinate	7×10^9
Δ^5 -3-Ketosteroid Isomerase	Pregnenedione	4.4×10^9
Superoxide dismutase	Superoxide	2.3×10^9
Fumarase	Malate	1.4×10^9
Cytochrome C Peroxidase	Cytochrome C	6×10^8
Triose Phosphate Isomerase	Glyceraldehyde-3-Phosphate	3.6×10^8
Triose Phosphate Isomerase	Dihydroxyacetone Phosphate	1.0×10^7

* from references 21 and 78.

and may facilitate an analysis of the motions of the atoms in the protein (29).

As described below, the crystals of the free yeast triose phosphate isomerase diffract to at least 1.5 Å resolution, and non-isomorphous crystals of the enzyme in the presence of a transition state analogue have been obtained (30). This raises the possibility of looking at the protein in several biologically interesting conformational states at atomic resolution.

The possibility of genetic experiments, realized through the work described in Chapter II, also makes the yeast enzyme a potentially powerful subject of investigation. The combination of the methods of modern protein crystallography and molecular genetics may yield qualitatively new information about the catalytic mechanism.

The most attractive feature of the triose phosphate isomerase reaction, however, is that it has been extensively studied using a wide range of physical and biochemical methods. Consequently, there is a large body of biochemical phenomenology to be explained in structural terms. The intellectual fusion of structure and function, then, is limited by our ability to design structural experiments to test specific hypotheses rather than by a paucity of information about the properties of the protein.

IV. Triose Phosphate Isomerase -- The Reaction.

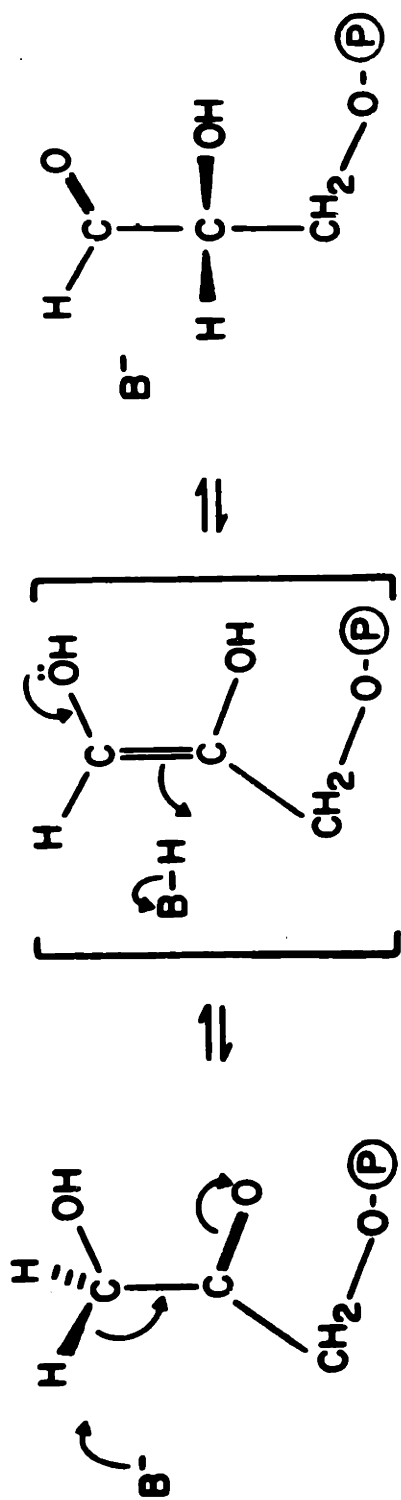
Triose phosphate isomerase catalyzes the interconversion of dihydroxyacetone phosphate (DHAP) and D-glyceraldehyde-3-phosphate (GAP), and, in doing so, it plays a critical role in glycolysis, gluconeogenesis, the pentose pathway and fatty acid metabolism (31,32). The reaction is shown in Figure 2.

A. The Flight from the Garden.

In the 1950's, two groups demonstrated that the enzyme introduces a single tritium into DHAP when the reaction is carried out in tritiated water (33,34). When this ^3H -labelled DHAP was exposed to the enzyme in H_2O , none of the label was found in the resulting GAP (trapped as 3-phosphoglycerate by including excess glyceraldehyde-3-phosphate dehydrogenase, NAD^+ and arsenate in the reaction mixture). On the basis of these results, a direct hydride transfer between C-1 and C-2, which would yield stoichiometric amounts of ^3H -GAP, was ruled out. By analogy to phosphoglucose isomerase, it was proposed that the reaction occurs by a proton transfer mechanism and that the substrates are interconverted through an enzyme-bound enediol intermediate (34,35).

The proton transfer was found to be completely stereospecific, occurring only between C-2 of D-GAP and the pro-R position of C-1 of DHAP (36). On the basis of the economical assumption that proton abstraction and donation occur on the same face of the planar intermediate, Reider and Rose argued that the stereospecificity of the reaction implies that

Triosephosphate isomerase



**dihydroxyacetone
phosphate**

**D-glyceraldehyde
phosphate**

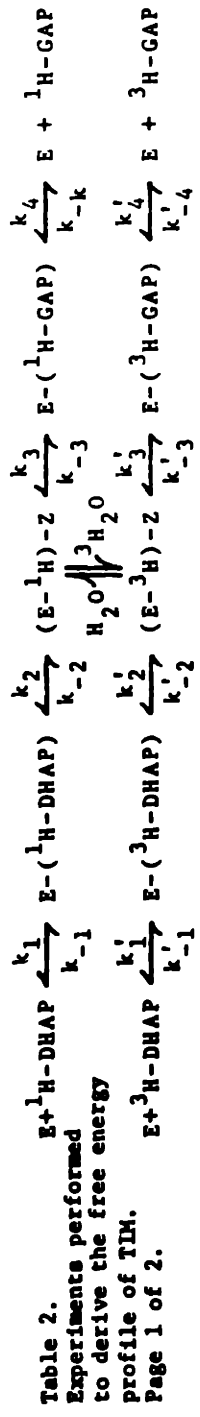
Figure 2. The triose phosphate isomerase reaction, after Reider and Rose, 1956 (34).

the intermediate is cis rather than trans. This species is shown in Figure 2, though formally the data are also consistent with a symmetrical enediolate as the enzyme-bound intermediate species.

Subsequent work demonstrated that 3 to 6% of the label in [1(R)-³H]-DHAP was transferred to GAP in the TIM reaction (37). This had two important consequences. 1) It supported the idea that proton transfer between C-1 and C-2 is catalyzed by a single base on the enzyme which accepts and donates the labile proton. Studies directed toward the identification of this base will be described in detail below. 2) It demonstrated that, even with the intermediate bound to the enzyme, the base is in almost complete equilibrium with the solvent.

Knowles, Albery and coworkers were able to exploit this situation through the judicious use of isotopic label to follow the fates of specific deuterium and tritium atoms in the substrates and the solvent during the TIM reaction. Starting with the scheme shown in Table 2, the theory which relates the results of sixteen different isotope exchange experiments to the microscopic rate constants was derived (38). Eleven of these experiments were performed and analyzed to give the Gibbs free energies of all of the intermediates and transition states in the reaction (39-41). These experiments are summarized in Table 2.

In general, three kinds of experiments were carried out. 1) The extent of transfer of isotope from labelled DHAP to GAP and the change in the specific activity of the substrate were measured as a function of the extent of the reaction (37,42). 2) Starting with unlabelled

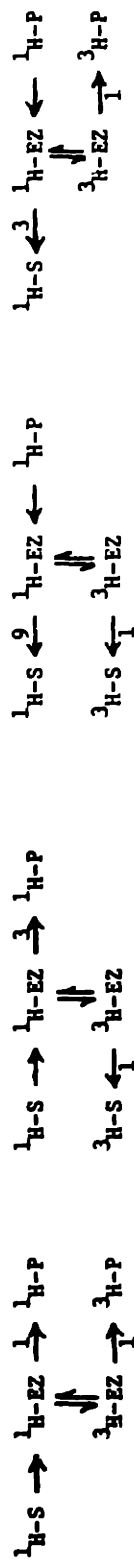


Isotope exchange experiments performed by Knowles, Albery and Coworkers.

Substrate	Solvent	Measure	Qualitative Results and Conclusions
1. ${}^3\text{H-DHAP} + {}^1\text{H-DHAP}$	${}^1\text{H}_2\text{O}$	Trap GAP with GAPDH* and measure its specific activity as a function of the extent of the reaction.	GAP has 3-6% of the specific activity of the starting material. This is consistent with a single enzyme base which catalyzes proton transfer. The E-Z complex is in almost complete isotopic equilibrium with the solvent. Proton exchange from the conjugate acid of the catalytic base is fast. (37)
2. ${}^3\text{H-DHAP} + {}^1\text{H-DHAP}$	${}^1\text{H}_2\text{O}$	Trap GAP with GAPDH and measure the specific activity of DHAP as a function of the extent of the reaction.	The specific activity of DHAP is about 3 times higher at the end of the reaction than at the beginning. This is a reflection of two competing processes: 1) a primary kinetic isotope effect at k_2 which causes ${}^1\text{H-DHAP}$ to react faster than ${}^3\text{H-DHAP}$ and 2) the equilibration of $(E-{}^3\text{H})-Z$ with solvent which washes the label out of substrate and product. (37)
3. ${}^1\text{H-DHAP}$	${}^3\text{H}_2\text{O} + {}^1\text{H}_2\text{O}$	Trap GAP with GAPDH and measure its specific activity as a function of the extent of the reaction.	Product GAP has a specific activity which is 0.77 that of the solvent. This corresponds to very little isotopic desamination (1.3) in the product forming step. Therefore, $E-({}^1\text{H-GAP})$, $(E-{}^1\text{H})-Z$, $(E-{}^3\text{H})-Z$ and $E-({}^3\text{H-GAP})$ are in almost complete isotopic equilibrium. Release of product involves passing through the rate determining transition state. $k_4' = k_4$, k_{-3} , k_3 , proton exchange, k_3' , & k_{-3}' . (42)
4. ${}^1\text{H-DHAP}$	${}^3\text{H}_2\text{O} + {}^1\text{H}_2\text{O}$	Trap GAP with GAPDH and measure the specific activity of DHAP as a function of the extent of the reaction.	${}^3\text{H-DHAP}$ accumulates as the reaction proceeds. This is a function of the partition ratio of E-Z and the kinetic isotope effect on k_2 . (42)
5. ${}^1\text{H-GAP}$	${}^3\text{H}_2\text{O} + {}^1\text{H}_2\text{O}$	Trap DHAP with COPDH** and measure the specific activity as a function of the extent of the reaction.	The specific activity of the product DHAP is 11% that of the solvent. Therefore, k_{-2} is kinetically significant; k_2 , k_{-1} and $k_{-2}' = 9 \times k_{-2}'$. Product release is faster than product formation. (43)
6. ${}^1\text{H-GAP}$	${}^3\text{H}_2\text{O} + {}^1\text{H}_2\text{O}$	Trap DHAP with COPDH and measure the specific activity of GAP as a function of the extent of the reaction.	${}^3\text{H-GAP}$ accumulates during the course of the reaction. ${}^1\text{H-DHAP}$ is formed 3 times faster than ${}^3\text{H-GAP}$. (43)

Table 2. Continued . . .

Qualitative Summary of Experiments 3 - 6.



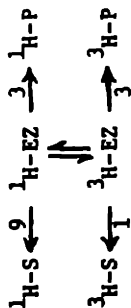
Experiment 3

4

5

6

Combining these relationships



The numbers above the arrows

yields:

represent the relative rate of that step.

Additional information was obtained by studying kinetic isotope effects.

Substrate	Solvent	Measure	Qualitative Results and Conclusions
7. $^2\text{H-DHAP}$	$^1\text{H}_2\text{O}$	k_{cat} and K_m	Combining the results of 7 & 8, 1) K_m is not altered by isotopic substitution, and 2) $(k_{\text{cat}}^{\text{H}}/k_{\text{cat}}^{\text{D}}) = 2.9$. Since E-GAP is in isotopic equilibrium with the solvent, k_{cat} its rate of breakdown is not affected by the H-isotope which is present on C-1 of the starting DHAP. The rate of the breakdown of E-GAP is dependent on its rate of formation, so the kinetic isotope effect arises from a difference in the rate of proton (deuterium) abstraction from the E-DHAP complex. (44)
8. $^1\text{H-DHAP}$	$^1\text{H}_2\text{O}$	k_{cat} and K_m	No kinetic isotope effect is observed, because E-GAP is in isotopic equilibrium with the solvent. The $^2\text{H-GAP}$ loses its deuterium before the rate limiting proton transfer in the conversion to DHAP. (44)
9. $^2\text{H-GAP}$	$^1\text{H}_2\text{O}$	k_{cat} and K_m	No isotope effect was observed in the extent of transfer of ^2H vs. ^3H from C-1 to C-2. Consequently, the isotopic equilibrium between the conjugate acid of the catalytic base and the solvent is not rate limiting in the formation of GAP. (45)
10. $^1\text{H-GAP}$	$^1\text{H}_2\text{O}$	k_{cat} and K_m	No isotope effect was observed in the extent of transfer of ^2H vs. ^3H from C-1 to C-2. Consequently, the isotopic equilibrium between the conjugate acid of the catalytic base and the solvent is not rate limiting in the formation of GAP. (45)
11. $^2\text{H-DHAP}$	$^1\text{H}_2\text{O}$	As in experiment #1, trap GAP with GAPDH and measure ^2H at C-2 as a function of the extent of the reaction	

* -- glyceraldehyde-3-phosphate dehydrogenase
 ** -- -glycerolphosphate dehydrogenase

" $^1\text{H-DHAP}$ " = $[\text{}^1\text{H}(\text{R})-\text{}^1\text{H}]\text{-dihydroxyacetone phosphate}$
 " $^2\text{H-GAP}$ " = $2\text{-}^2\text{H-glyceraldehyde-3-phosphate}$

substrates in tritiated water, the specific activity of the substrate and product were measured as a function of the extent of the reaction (43,44). DHAP and GAP could not be labelled directly, but only in the course of the breakdown of the enzyme-bound intermediate. Consequently, these experiments gave information about the partitioning of the enzyme-intermediate complex and the rate at which it exchanges protons with the solvent. 3) Deuterium kinetic isotope effects in the overall reaction were measured to probe the kinetic significance of proton transfer steps (45).

In all cases, the product of the isomerase reaction was trapped using the appropriate dehydrogenase to prevent the back reaction of the released product from contributing to the observed distribution of isotope. GAP-dehydrogenase was used to scavenge free GAP, and DHAP was trapped using α -glycerolphosphate dehydrogenase. The change in absorbance of the NAD^+ and NADH cofactors of these dehydrogenases also provided a measure of the rate of both enzymatic reactions.

The free energy profile for the TIM reaction determined by Knowles, Albery and coworkers is shown in Figure 3 (40). The standard state was defined in relation to the concentration of triose phosphates in muscle (40 mM). In the conversion of DHAP to GAP, the slowest step is the enolization of the substrate to form the enzyme-bound intermediate, but the rate limiting transition state is not encountered until the release of product. The back reaction from enzyme-bound intermediate is three times faster than the release of GAP (Table 2).

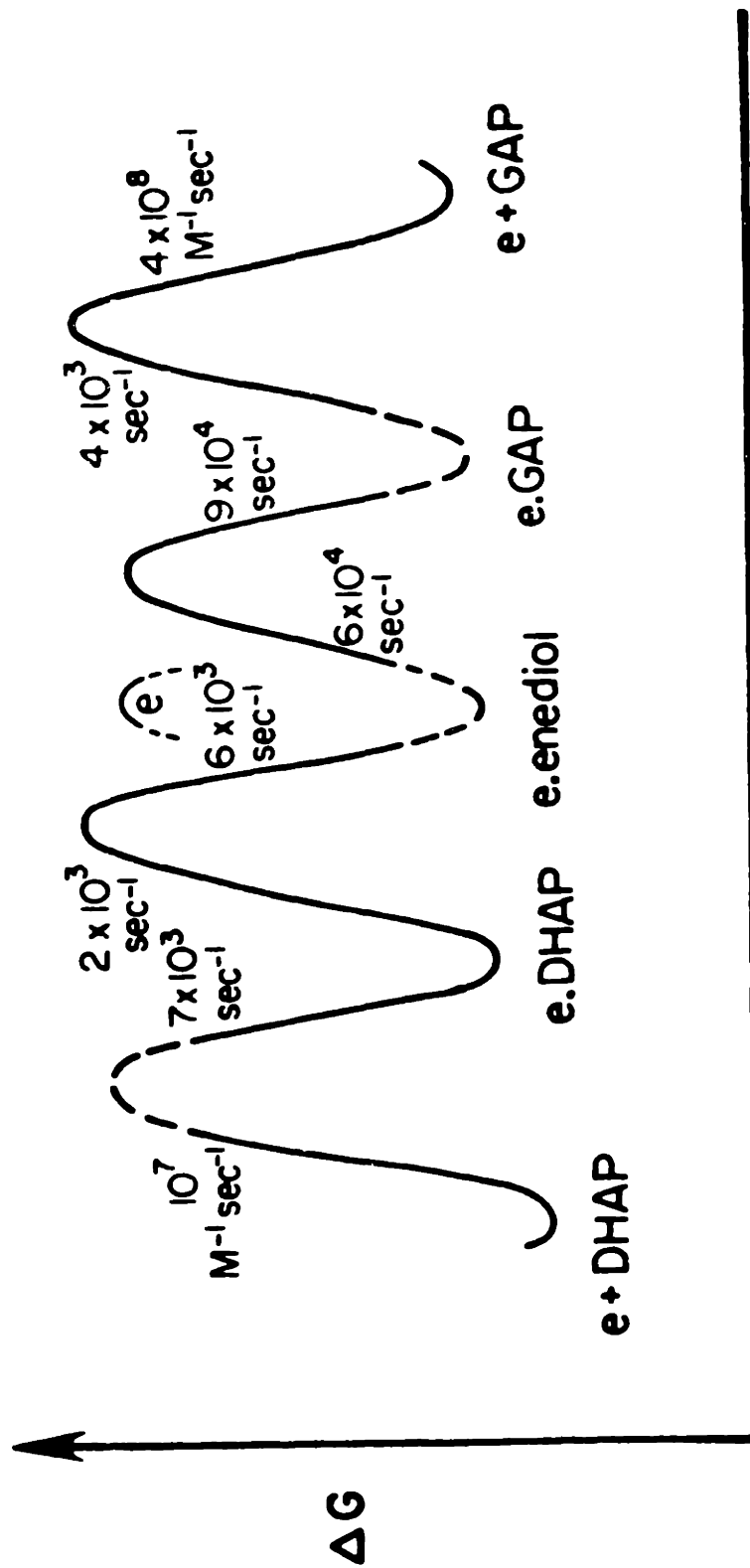


Figure 3. Free energy profile for the triose phosphate isomerase catalyzed reaction. The dashed parts of the profile are less precisely established, and the rate constants for the elementary steps involving these states are lower limits. These states are probably kinetically insignificant. The barrier marked "e" is that for the exchange of protons between the enzyme-intermediate and the solvent. The intermediate may be an enediol or enediolate. From Albery and Knowles, 1977 (41).

In the conversion of GAP to DHAP, the reaction rate may be close to the diffusion controlled limit. The binding of GAP involves the highest free energy barrier, corresponding to a rate of $4 \times 10^8 \text{ s}^{-1}\text{M}^{-1}$. The slowest chemical step (which may be partially rate limiting) is the formation of the enzyme-bound product. The activation free energy barrier for the binding and release of DHAP is not kinetically significant.

The exchange of protons between to enzyme-intermediate complex and the solvent involves the second lowest free energy barrier of the reaction. The enzyme-intermediate complex forms E-GAP about 10 times faster than E-DHAP ($k_3 = 10 \times k_{-2}$, Table 2). This partition ratio is similar to that for the non-enzymatic collapse of the enediol intermediate in aqueous solution (46). If the enzyme-bound intermediate is, indeed, an enediol, this suggests that both transition states for enolization have been stabilized to the same extent by the enzyme (46). The free energy of stabilization is about 7 kcal/mole, corresponding to a rate enhancement of 2×10^9 at 37° C .

The enzyme-intermediate and enzyme-GAP complexes, marked in dashed lines in Figure 3, are probably not kinetically significant. All the enzyme-bound species are higher in free energy than enzyme + free DHAP.

One important feature of the free energy profile is that the E-DHAP complex is the most stable enzyme-bound species. This implies that the crystal structure of the TIM-substrate complex will be dominated by contributions from this species (41). Assuming that the free energy

T °C	$k_{-1} \text{ s}^{-1}$	$k_2 \text{ s}^{-1}$	$k_{-2} \text{ s}^{-1}$	$k_3 \text{ s}^{-1}$	$k_{-3} \text{ s}^{-1}$	$k_4 \text{ s}^{-1}$
30	7.7×10^3	1.7×10^3	5.7×10^3	6.2×10^4	8.6×10^4	4.1×10^3
25	5.4×10^3	1.2×10^3	3.9×10^3	4.5×10^4	6.2×10^4	2.8×10^3
0	7.3×10^2	1.4×10^2	5.2×10^2	7.3×10^3	1.1×10^4	3.6×10^2
-20	1.1×10^2	18.2	77.2	1.3×10^3	2.0×10^3	51.9
-40	12.4	1.7	8.4	1.9×10^2	2.8×10^2	5.4
-60	0.9	0.1	0.6	17.9	28.4	0.4
-80	4.1×10^{-2}	3.8×10^{-3}	2.5×10^{-2}	1.1	1.8	1.5×10^{-2}
-90	6.6×10^{-3}	5.5×10^{-4}	4.0×10^{-3}	0.2	0.4	2.3×10^{-3}
-100	8.8×10^{-4}	6.3×10^{-5}	5.2×10^{-4}	3.4×10^{-2}	6.0×10^{-2}	2.9×10^{-4}
-110	9.1×10^{-5}	5.5×10^{-6}	5.2×10^{-5}	4.4×10^{-3}	8.0×10^{-3}	2.8×10^{-5}

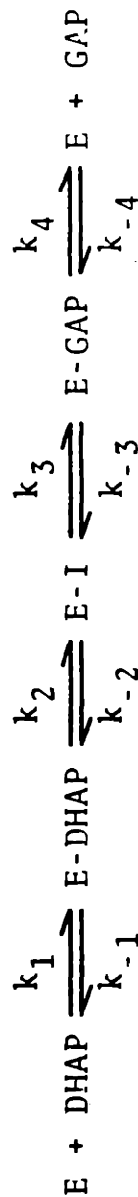


Table 3. Microscopic rate constants for the triose phosphate isomerase reaction extrapolated to low temperatures. The rates and activation energies reported by Albery and Knowles have been assumed to be exact and constant as a function of temperature. The expression, $\Delta G^\ddagger = -RT \ln(k_f/\kappa kT)$, where k_f is the rate constant, h is the Boltzmann constant, κ is the transmission coefficient and k is Planck's constant, was evaluated. (40)

profile does not change as a function of temperature, the microscopic rate constants can be extrapolated to low temperatures. The results of such a calculation are shown in Table 3. For the upper limit case in which the enzyme is as fast in the crystal as it is in solution, it is clear that the enzyme-GAP complex can be trapped long enough to collect high resolution x-ray data (10^5 - 10^6 seconds) only at temperatures well below -100° C. Though enzymatic reactions are often slowed by restraints to substrate diffusion and protein motion imposed by the crystal environment (20,47), the prospect of having to resort to such low temperatures makes such experiments technically challenging.

The relative free energies of the enzyme-bound species calculated by Knowles and Albery have recently been challenged (48,49). Iyengar and Rose mixed ^{32}P -DHAP with excess enzyme and analyzed the labelled species liberated by acid denaturation of the equilibrium mixture (49). DHAP, inorganic phosphate and GAP were found in the ratio 42 : 5 : 52. Since the substrates are relatively stable in acid, this result implies that the ratio of E-GAP to E-DHAP at equilibrium is about 1.2, a clear contradiction to the conclusion from the free energy profile that E-DHAP should predominate (41). This is also a large change from the equivalent ratio of unhydrated substrates in solution (about 0.003 (51)) and implies that the enzyme binds GAP more tightly than DHAP.

Belasco and Knowles have attempted to probe directly the nature and distribution of enzyme-bound triose phosphates using Fourier transform infrared spectroscopy (52). The infrared spectrum of DHAP bound to the enzyme showed two bands, with a 3:1 intensity ratio, which were not

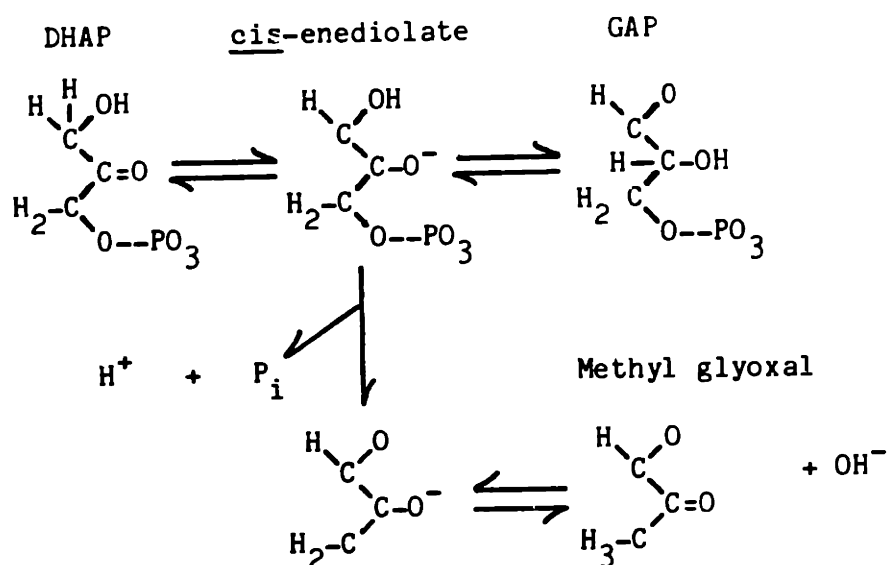
observed when the enzyme active sites were blocked prior to addition of substrate. Neither band was shifted more than 2 cm^{-1} in the spectrum of bound 1(S)- ^2H -DHAP, and both bands were shifted under the amide I absorption of the protein by isotopic labelling of the substrate carbonyls with ^{18}O . Thus, both bands were attributed to the carbonyl absorption of DHAP bound at the active site. This assignment has been confirmed by measuring the spectrum of bound 1- ^{13}C -DHAP (J.G. Belasco and J.R. Knowles, personal communication). It is curious that two bands corresponding to bound DHAP were found, and it was hypothesized that the bands represent two different conformations of this substrate in the active site (52).

The failure of the infrared experiment to detect bound GAP is enigmatic in light of the findings of Iyengar and Rose that this complex represents 52% of the enzyme-bound species at equilibrium (49). The spectra indicate that either the E-GAP complex is not present (as predicted by the free energy profile of Knowles and Albery) or that this complex is not observable by infrared spectroscopy because the C-1 carbonyl resonance is shifted under the amide I band of the protein (52). Though this latter possibility is unlikely (J.R. Knowles, personal communication), it is, nonetheless, disquieting. If the E-GAP complex could be stabilized long enough to collect the Fourier transform infrared spectrum (about 10 minutes), then the question of whether is possible to detect it with this technique could be resolved, and the current spectra could be interpreted with more confidence. At least two strategies could be adopted here. One is to mix the enzyme and GAP and record the spectrum at a sufficiently low temperature to prevent the

breakdown of the E-GAP complex (Table 3). Control experiments would have to be done to show that the aqueous-organic solvent used in such an experiment does not alter the structure of the protein or the energetics of the catalytic reaction (53,54). A second approach involves using a mutant triose phosphate isomerase which binds GAP and polarizes the C-1 carbonyl group but doesn't form or break down the enzyme-bound intermediate. An enzyme which lacks the active site base may satisfy these criteria. The possibility of constructing such mutants is described in detail in Chapters II and V. If such experiments show that bound GAP cannot be observed by infrared spectroscopy, other approaches will be needed to resolve the conflict concerning the ratio of enzyme-bound species in the steady state.

Alternatively, the root of the conflict may lie in the failure of the results of Iyengar and Rose to reflect the true equilibrium ratio of enzyme-triosephosphate complexes. Since the denaturation of the E-S complex in trichloroacetic acid is likely to be slower than the interconversion of substrates, the pathway of acid denaturation may force a peculiar partitioning of the enzyme-bound species. If this were the case, different denaturation pathways may release different ratios of substrates and intermediates. This possibility could be tested by quenching the E-S complex using a variety of conditions -- including rapid changes in temperature or pressure and/or exposure of the E-S complex to detergents or polar aprotic organic solvents -- and comparing the ratios of substrate species which are released. Unfortunately, a negative result in such experiments would not be conclusive.

Nonetheless, the acid quenching experiments of Iyengar and Rose have provided clues about the nature of the enzyme-bound intermediate and the basis of the specificity of the enzymatic reaction (49,50). Several workers have reported that triose phosphate isomerase causes a slow ($k = 0.5 \text{ s}^{-1}$ at pH 6) and irreversible breakdown of its substrates into methyl glyoxal and inorganic phosphate (49,55). This side reaction is thought to arise from the non-enzymatic decomposition of the enediol/enediolate intermediate:



The reaction depends on the presence of active enzyme, and it is not inhibited when enzyme is in excess. Consequently, the reaction occurs while the intermediate is bound (50). Iyengar and Rose argued that the inorganic phosphate produced by quenching the TIM reaction in acid arises from the analogous dephosphorylation of the free enediol (48,49). Their results suggest that the ratio of bound intermediate to bound substrates is 1 : 20.

The kinetics of the formation of inorganic phosphate in the acid quenching experiments were studied in order to determine the nature of

the labile species (50). If the quenched reaction mixture was neutralized and mixed with excess native triose phosphate isomerase within 5 ms of denaturation, the formation of P_i could be inhibited by 80%. By varying the time between quenching the equilibrium mixture and trapping the released triose phosphates with excess native enzyme, the half-time for the appearance of P_i was found to be about 15 ms ($k = 60 \text{ s}^{-1}$) in 0.2 N trichloroacetic acid (pH 1). Extrapolation to zero time between quenching and trapping showed that 100% of the labile triose phosphate was in a form that could be stabilized by the enzyme. This suggested that only one acid-labile species is released by denaturation, that its break down occurs in solution and that, once released into the solvent, this material is easily rebound by the enzyme. These data support the enediol mechanism shown in Figure 2.

By varying the pH of the neutralizing solution between 4.5 and 8.5, the pH dependence of the rate of P_i formation was measured (50). The rate varied between 60 s^{-1} and 160 s^{-1} with a minimum at pH 6.5. The change of pH did not alter the the partitioning of the free enediol in the absence of trapping enzyme. A comparison of the enzymatic and non-enzymatic rates of enediol fragmentation shows that the enzyme slows down the dephosphorylation reaction by a factor of 10^3 at pH 5.5 and 3×10^5 at pH 7.5 (50). This helps give the enzyme kinetic specificity. Since ester hydrolysis is favored by an out-of-plane conformation of the bridge oxygen (56), the enzyme may inhibit dephosphorylation by binding the phosphate group in an in-plane conformation.

The acid quenching experiments also provide evidence which suggests that the enzyme-bound intermediate is an enediol rather than an enediolate (50). This point is critical to an understanding of the catalytic events, because the obligate formation of an enediol would favor a general-acid/general-base mechanism. At the same time it would tend to rule out mechanisms involving only electrostatic stabilization of the enediolate anion by the enzyme and/or by the proton of the neighboring hydroxyl group of the substrate.

Since the pK_a of an enediolate is high, the intermediate which is formed by acid denaturation and subsequently trapped by native TIM will be an enediol phosphate. The dissociation constant of the species is very low ($<10^{-10}$ M), and its rate of association with TIM is apparently diffusion controlled (50). The enzyme concentration at which the trapping is half-maximal is independent of pH between 4.5 and 8.5, indicating that the enediol is turned over by TIM. Therefore, this species is on the reaction pathway and can join the reaction pathway rapidly. Arguing against the presence of a fortuitous base that could accept a proton to form an enediolate, Iyengar and Rose have suggested that the enediol phosphate is the catalytic intermediate (50).

However, this point has not been definitively settled. Preliminary experiments indicate that the enediol phosphate is not catalytically competent; it is not turned over as rapidly as GAP (I.A. Rose, personal communication). This suggests that a structural rearrangement of the enzyme or the loss of a proton from the enediol to form the enediolate limits the reaction rate. Structural studies described below suggest

that Glu 165, the catalytic base, may be able to accept a proton from the enediol and exchange it with the solvent.

B. The Nature of the Catalytic Groups (The Wilderness).

As mentioned above, the transfer of tritium from C-1 of DHAP to C-2 of GAP by triose phosphate isomerase suggests the presence of a single enzymatic base which mediates the shuttling of protons (57,58). Covalent modification of the protein using substrate analogues has led to the identification of this basic group. The compounds which have been used include glycidol phosphate, glycidol plus inorganic phosphate (!), haloacetol phosphates and chloroacetol sulfate, and their reactivity toward the triose phosphate isomerases from yeast, chicken muscle, rabbit muscle and human erythrocytes has been investigated (59-65). In all cases, the reaction is rapid, stoichiometric, leads to complete inactivation of the enzyme and is inhibited by substrates and noncovalent inhibitors. D-glycidol phosphate reacts ten times faster than the L- isomer, mimicking the substrate specificity of the enzyme. The inactivation leads to the formation of a 1-acyl ester of Glu 165, a residue which is part of a conserved region of the protein sequence.

The pKa of Glu 165 was probed by measuring the pH dependence of the inactivation rate. Using glycidol phosphate as the active-site-directed alkylating agent, Waley (66) concluded that the pK of the carboxylate was close to 6.0, in apparent conflict with the results of Schray et al who found no pH dependence of the inactivation rate between

pH 5.5 and 10 (62). These results are not easily interpreted because the phosphate moiety of the reagent changes its ionization state in the pH range tested. To circumvent this problem, Hartman et al measured the pH dependence of the rate of inactivation of the yeast enzyme with chloroacetol sulfate (67). This reagent selectively alkylated Glu 165, and their measurements were consistent with a pKa of 3.9 for that residue. This pK_a was low enough to account for the extremely rapid exchange of protons during catalysis.

Unfortunately, chloroacetol sulfate exists as a mono-anion over the pH range tested, and substrates and inhibitors have been found to bind predominantly as di-anionic species (see below). In addition, structural studies suggest that the conformation of the protein does not respond identically to the binding of sulfate and phosphate (28). As a result, the electrostatic potential in the region of Glu 165 in the presence of bound substrates may not be adequately modeled by the binding of chloroacetol sulfate, and the pK_a of 3.9 for Glu 165 is subject to this caution.

The fast alkylation of the enzyme by D-glycidol phosphate also provided indirect evidence for the presence of an acidic or electrophilic group in the active site of the enzyme which stabilizes developing negative charge on the ring oxygen (67). This active site electrophile could also stabilize the developing enolate species during the course of the catalytic reaction. The absence of a pH effect on the inactivation rate between pH 5 and 10 suggests that the electrophilic group, if it acts as an acid, has a pK higher than 10 (67) and/or that

proton donation is not rate-limiting, even at high pH. It is also consistent with the electrophilic catalysis being electrostatic rather than acidic in nature.

Additional evidence for electrophilic catalysis in the reaction of triose phosphate isomerase comes from the work of Knowles and coworkers. Webb and Knowles reported that the reduction of enzyme-bound DHAP by sodium borohydride is completely stereospecific and is eight times faster than the reduction of free substrate (68). This is consistent with the presence of an electrophilic group in the active site that polarizes the carbonyl group of DHAP and destabilizes its ground state conformation. Theoretically, this ground state destabilization could accelerate both the rate of sodium borohydride reduction at C-2 and facilitate the abstraction of the pro-R proton of C-1 by Glu 165 during the normal catalytic reaction. No reduction of the carbonyl group of enzyme-bound GAP was observed, presumably because C-1 is not accessible to sodium borohydride (68).

Though Webb and Knowles did not distinguish between the neighboring hydroxide of the substrate and a group on the enzyme as the source of the electrophilic center, at least two lines of evidence implicated the enzyme in this process. 1) The rate of alkylation of Glu 165 by D-glycidol phosphate is accelerated by the enzyme even though this substrate analogue contains no neighboring hydroxide. 2) Were the neighboring hydroxide a sufficiently potent electrophile, a 10^9 -fold difference in the enzyme-catalyzed and uncatalyzed rates of interconversion of GAP and DHAP would not be expected.

The experiments of Webb and Knowles were subject to the caveat that specific binding of NaBH_4 to the active site prior to its reaction with DHAP may have caused the observed increase in the reduction rate. In addition, they could not formally distinguish between ground state destabilization and transition state stabilization (or both) as the root cause of the observed catalysis. Direct observation of substrate distortion by triose phosphate isomerase was accomplished by Belasco and Knowles using Fourier transform infrared spectroscopy (52). These experiments have been described in some detail above. Of the two bands in the infrared spectra which correspond to bound DHAP, the more intense band was shifted 19 cm^{-1} relative to the carbonyl absorption of DHAP in solution. This indicates that the ground state conformation of the substrate is strained on binding to the enzyme. The same two bands were observed using the yeast and chicken enzymes suggesting that the environment of the active site and, presumably, the origin of the catalytic power are similar in both proteins.

Kinetic experiments have also been carried out to help define the nature and behavior of the catalytic groups. As described above, the enzymatic reaction is extraordinarily fast. For triose phosphate isomerases from a number of organisms, the overall rate constant for the conversion of GAP to DHAP is about 2800 sec^{-1} , and k_{cat} for the reverse reaction is 250 sec^{-1} (69).

The K_m values for substrate interactions with TIMs from several organisms are listed in Table 4 (51). A number of methods -- including

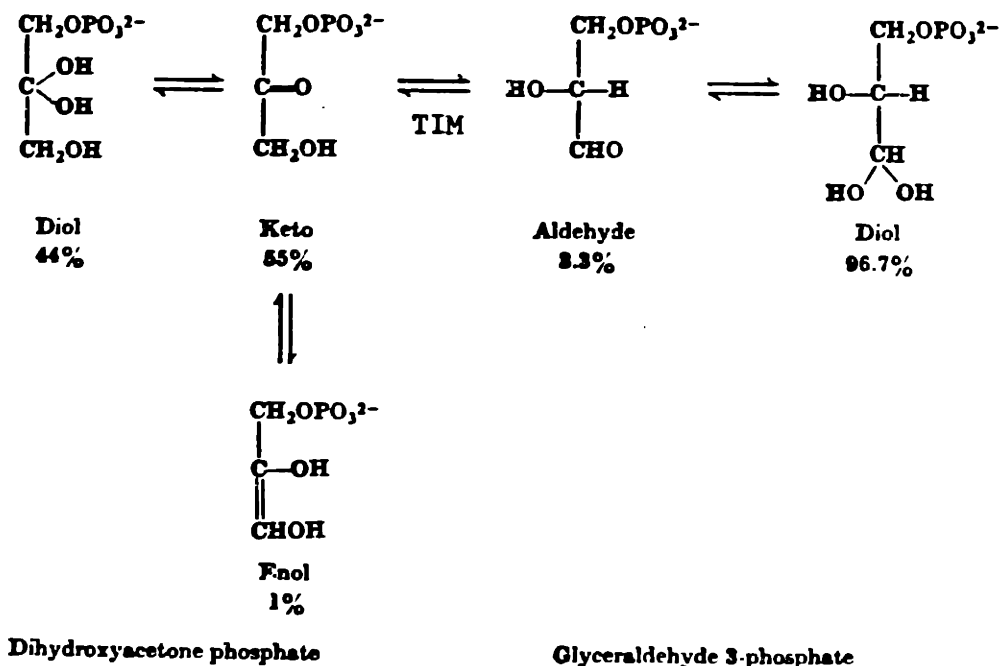


Figure 4. In vitro concentrations of triose phosphates at equilibrium at 20° C. Only the unhydrated forms are TIM substrates (51).

Table 4. Kinetic and thermodynamic constants for the TIM reaction. The discovery that only a fraction of the triose phosphates in solution are TIM substrates required a reassessment of the kinetic and thermodynamic constants of the enzyme (51).

(a) Equilibrium data	20°C	37°C
Glyceraldehyde-3- P_{tot} (%)	3.3	4.3
Dihydroxyacetone- P_{tot} (%)	55	32
K_{eq} (dihydroxyacetone- P /glyceraldehyde-3- P)		
Original value	22	22
Recalculated	367	420
ΔG° (kcal/mol)		
Original value	-1.82	-1.90
Recalculated	-3.49	-3.17
(b) Kinetic parameters		
K_m (glyceraldehyde-3- P)	Reported (μ)	Recalculated (μ)
Rabbit muscle (Burton & Waley, 1968)	4.6×10^{-4}	15×10^{-4}
Rabbit muscle (Krietsch <i>et al.</i> 1970)	3.2×10^{-4}	11×10^{-4}
Rabbit liver (Krietsch <i>et al.</i> 1970)	3.2×10^{-4}	11×10^{-4}
Human erythrocyte (Schneider, Valentine, Hattori & Heina, 1965)	3.5×10^{-4}	12×10^{-4}
Yeast (Krietsch <i>et al.</i> 1970)	12.7×10^{-4}	42×10^{-4}
K_m (dihydroxyacetone- P)		
Rabbit muscle (Burton & Waley, 1968)	8.7×10^{-4}	4.8×10^{-4}
Rabbit muscle (Krietsch <i>et al.</i> 1970)	6.2×10^{-4}	3.4×10^{-4}
Rabbit liver (Krietsch <i>et al.</i> 1970)	6.0×10^{-4}	3.3×10^{-4}
Yeast (Krietsch <i>et al.</i> 1970)	12.3×10^{-4}	6.8×10^{-4}

stopped flow coupled enzyme assays (51,70), infrared and ultraviolet spectroscopy (70,71), nuclear magnetic resonance (51,55,72), and isotope partition experiments (73) -- have been used to demonstrate that the active forms of the substrates are unhydrated di-anions. At 20° C, only 3% of the GAP and 55% of the DHAP are present in their unhydrated forms at equilibrium (Figure 4) (51,70). These ratios are temperature dependent. Consequently, the equilibrium constant for total substrate of about 22 in favor of DHAP (74-77) implies that the ratio of keto-DHAP to aldehydic-GAP is 420 in solution (51).

In vivo, however, the gem-diols are probably not present in appreciable concentrations, because the enzymes that utilize triose phosphates are fast and abundant, and the hydration reaction is slow (51). TIM, aldolase, α -glycerolphosphate dehydrogenase and glyceraldehyde-3-phosphate dehydrogenase all act on unhydrated substrates, which consequently predominate in the steady state, and the triose phosphate pool in vivo does not reach equilibrium.

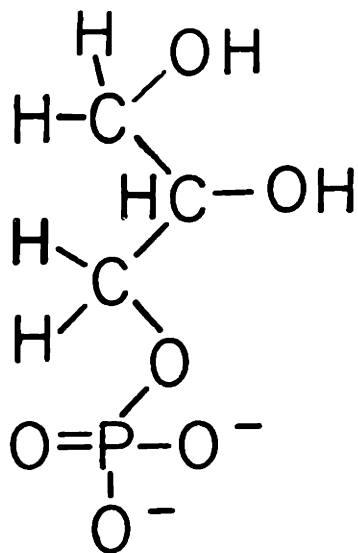
The pH dependences of the kinetic parameters k_{cat} and K_m for the chicken muscle enzyme were determined by Plaut and Knowles (78). The value of k_{cat}/K_m is highest between pH 7 and 8 and has apparent pK_a values of 6.0 and 9.0. The value of k_{cat} is sigmoid with respect to pH, with pK_a 's of 5.9 for the conversion of GAP to DHAP and 6.05 for the reverse reaction. For yeast TIM, Hartman and Ratrie found that k_{cat} and k_{cat}/K_m exhibit pK_a 's of 4.6 and 5.9, respectively (79). The phosphonate analogues of DHAP and GAP are poor substrates for TIM and titrate near pH 7.5, over a pH unit higher than the corresponding triose

phosphates (73). Under second order conditions ($[S_o] \ll K_m$), k_{cat}/K_m for the isomerizations of the GAP analogue, 2-hydroxy-4-phosphonobutyralsdehyde, has a pK_a of 7.5. These results implicate the ionization state of the substrate in the pH dependence of the reaction rate and are inconsistent with early suggestions that the pK_a of k_{cat}/K_m near 6.0 arises exclusively from the ionization of a group on the enzyme (73). Instead, the pH dependence of the reaction is as complex as it is controversial, involving a change in the pK_a of the substrate and/or enzyme on formation of the E-S complex.

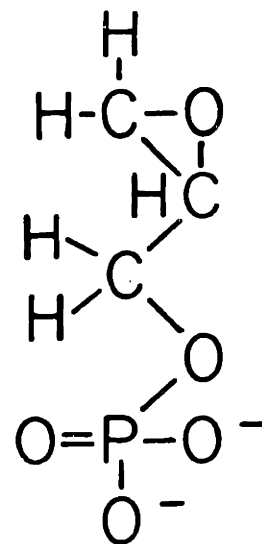
Many substances have been found which competitively inhibit the TIM reaction. Inhibition constants for some of these compounds are listed in Table 5 (69,80). 2-Phosphoglycolate (PGA) binds slightly more tightly than GAP, and it has been put forward as a potential transition state analogue (81,82). It binds at the active site as a tri-anion, and complex formation is accompanied by the uptake of a proton by the enzyme (71,72). The pH dependence of its binding is different from that of substrates, and it induces larger changes in the unit cell dimensions of the chicken enzyme crystals (82). The spectral changes resulting from phosphoglycolate binding are similar to those caused by the binding of D- α -glycerol phosphate, a substrate analogue (71). Nonetheless, because its inhibition constant is not markedly lower than K_m for GAP, its role as a transition state analogue remains controversial (69). Instead, it may mimic the enediol/enediolate intermediate or the substrates. Waley et al have proposed that the carboxyl group of PGA shares a proton with Glu 165 when it is bound in the active site (72). This interaction may be a model for nonproductive binding of the substrates and could explain

Table 5. Inhibition constants for the triose phosphate isomerase reaction (69,80). Data for the enzymes from several different organisms and tissues is given.

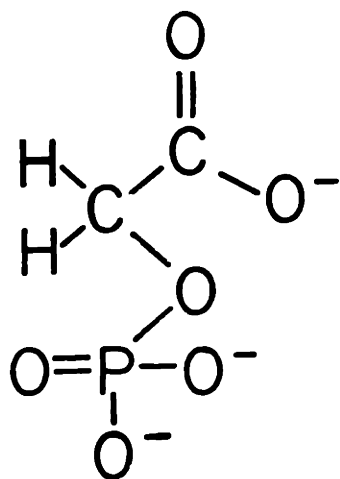
Compound	Organism			
	Yeast	Rabbit Muscle	Rabbit Liver	Chicken
D- α -glycerol phosphate		0.1 mM		
DL- α -glycerol phosphate		0.2 mM		0.15 mM
Phosphoglycolate		6 μ M		6.8 μ M
Phosphoglycolo- hydroxamate		4 μ M		
Phosphate		4.8 mM		5.5 mM
Sulfate		5.0 mM		5.7 mM
Arsenite	12.0 mM	6.2 mM	6.3 mM	
Phosphoenol- pyruvate	9.0 mM	3.0 mM	2.5 mM	
Allyl-phosphate		1.3 mM		1.0 mM
2-Hydroxyethyl- phosphate		2.0 mM		1.4 mM
D-ribose-1-phosphate		4.8 mM		3.1 mM
Glycolate		5.4 mM		10.0 mM
Succinate		1.4 mM		1.0 mM
Maleate		1.7 mM		3.5 mM
Malonate		0.18 mM		1.5 mM
Oxalate		9.0 mM		6.2 mM
Acetate		7.7 mM		10.0 mM



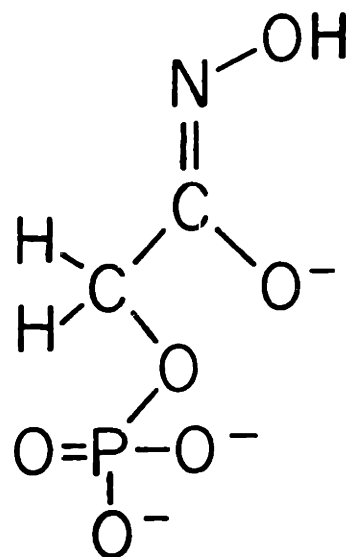
α -Glycerol phosphate



Glycidol phosphate



2-Phosphoglycolate



Phosphoglycolohydroxamate

Figure 5. Inhibitors of triose phosphate isomerase. α -Glycerol phosphate is a substrate-analogue, and glycidol phosphate is a suicide inhibitor. 2-Phosphoglycolate and phosphoglycolohydroxamate may be transition-state or intermediate-analogues. The binding of 2-phosphoglycolate may also mimic nonproductive binding of the substrate.

the presence of two bands in the Fourier transform infrared spectrum of enzyme-bound DHAP.

The specificity of the enzyme is reflected in the inhibition constants for related compounds. For example, D- α -glycerol phosphate binds about as tightly as DHAP, while binding of the L- enantiomer cannot be detected (55). In addition, an extremely surprising result is that dihydroxyactone sulfate, a close cousin of DHAP, does not bind to the enzyme (K_i is greater than 100 mM), let alone act as a substrate (73)! The lack of binding has been ascribed to its inability to form a sulfate di-anion. This cannot be the only prerequisite, however, because the mono-anionic forms of the substrates and D- α -glycerolphosphate do bind to the enzyme, albeit weakly (72,73). Structural studies indicate that inorganic phosphate and sulfate (both di-anionic) bind at the same place in the active site and yet stabilize quite different conformations of the enzyme (28). Thus, productive binding of substrates seems to require a di-anionic species with a particular van der Waals radius and particular electronic properties.

In summary, the kinetic and thermodynamic studies of the triose phosphate isomerase reaction are consistent with the presence of three enzyme-bound species on the reaction coordinate. These are complexes of di-anionic DHAP, GAP and the enediol or endiolate intermediate. There is disagreement over the ratio of these species in the steady state. Their rapid formation and breakdown is facilitated by active site groups which provide specificity of binding, transfer protons, polarize the carbonyl groups, stabilize the developing transition states and inhibit

the non-enzymatic breakdown of the unstable intermediate. The form of the intermediate (diol or diolate) and, as a result, the nature of the electrophilic catalysis (acidic or electrostatic) is still controversial. Glu 165 is the most promising candidate for the active site base. Structural studies using x-ray crystallography and protein sequencing methods were undertaken to determine what features of the enzyme lie at the root of these functions.

V. Triose Phosphate Isomerase - The Protein (The Hand of Chance).

Triose phosphate isomerases from a number of organisms and tissues have been purified and characterized (69). In all cases, the protein is a dimer with a molecular weight of about 53,000, and contains no cofactors or metal ions. Protein sequencing experiments suggest that the subunits are chemically identical, and there is no evidence for kinetic cooperativity in binding or catalysis.

Waley has found that the appearance of enzyme activity is second order with respect to subunit concentration when the denatured subunits are diluted from 5 M into 17 mM guanidine hydrochloride. This suggests that the dimer is the active species (83). Formally, the apparent inactivity of monomers could be an artifact of a folding pathway which is peculiar to the conditions of the experiment.

In contrast, White and Fell have reported that monomers crosslinked to an insoluble matrix have enzymatic activity in the absence of denaturants (84,85). Unfortunately, control experiments describing the

relative yield of crosslinked species, the microenvironment of the protein on the beads, the specific activity of crosslinked enzyme, and the nature of the refolding reaction and the activity resulting from crosslinking after denaturation were not done, making it impossible to conclusively interpret these results.

Triose phosphate isomerases from four organisms -- rabbit, chicken, coelacanth and Bacillus stearothermophilus -- have been sequenced (23-26). The amino acid sequences of the vertebrate enzymes are 85% homologous, while the sequence of the thermophilic protein is only 30-40% homologous with the other three. The vertebrate enzymes have been evolving at a snail's pace -- one change per 50 residues per 100 million years. This rate of evolution is comparable to that of glyceraldehyde-3-phosphate dehydrogenase but slower than cytochrome C (25,26).

The regions of the enzyme which are thought to be catalytically important are conserved in all four proteins. Active heterodimers of the chicken and B. stearothermophilus enzymes have been made, suggesting that the monomers adopt similar structures despite the sequence heterogeneity (86).

Crystals good enough for x-ray diffraction studies have been produced using the chicken and rabbit muscle enzymes (82,87). Rabbit triose phosphate isomerase was crystallized as monoclinic needles and hexagonal bipyramids. The hexagonal crystals have the symmetry of space group $P6_122$ and diffract to 3.0\AA resolution. The unit cell dimensions

are $a = b = 99.4 \text{ \AA}$, $c = 89.6 \text{ \AA}$. The asymmetric unit is the monomer, indicating that the subunits are identical. The needles have the symmetry of space group $P2_1$ and unit cell dimensions of $a = 36.6 \text{ \AA}$, $b = 76.4 \text{ \AA}$, $c = 89.6 \text{ \AA}$ and $\beta = 100.5^\circ$. The chicken enzyme crystals are orthorhombic (space group $P2_12_12_1$, with $a = 106.0 \text{ \AA}$, $b = 74.8 \text{ \AA}$ and $c = 61.7 \text{ \AA}$) and diffract to 2.5 \AA resolution. The dimer forms the asymmetric unit in these crystals. Microcrystals of the enzymes from brewer's yeast, rabbit liver, calf muscle and human erythrocytes have also been reported (63,88-90).

The structure of chicken muscle triose phosphate isomerase was solved at 2.5 \AA resolution (27) and recently refined to an R factor of 21% by Phillips and coworkers (D.C. Phillips, personal communication).

The structure of the protein is extraordinarily beautiful (Figure 6). The monomer is largely globular, with a radius of about 35 \AA . Its core is composed of eight strands of parallel β -sheet arching around the surface of a cylinder with a right handed twist. This β -barrel is surrounded by twelve α -helices which help connect the strands of β -sheet and, like a corona, form a secondary cylinder on the outer surface of the protein. 55% of the residues are in α -helices and 22% in β -sheets. Irregular loops connect the features of secondary structure, and one such loop, composed of residues 70-80, protrudes from the globular monomer and is involved in many intersubunit contacts.

This loop forms a hydrophobic pockets around the Met 14 residue of the adjacent subunit on the periphery of the subunit interface. The

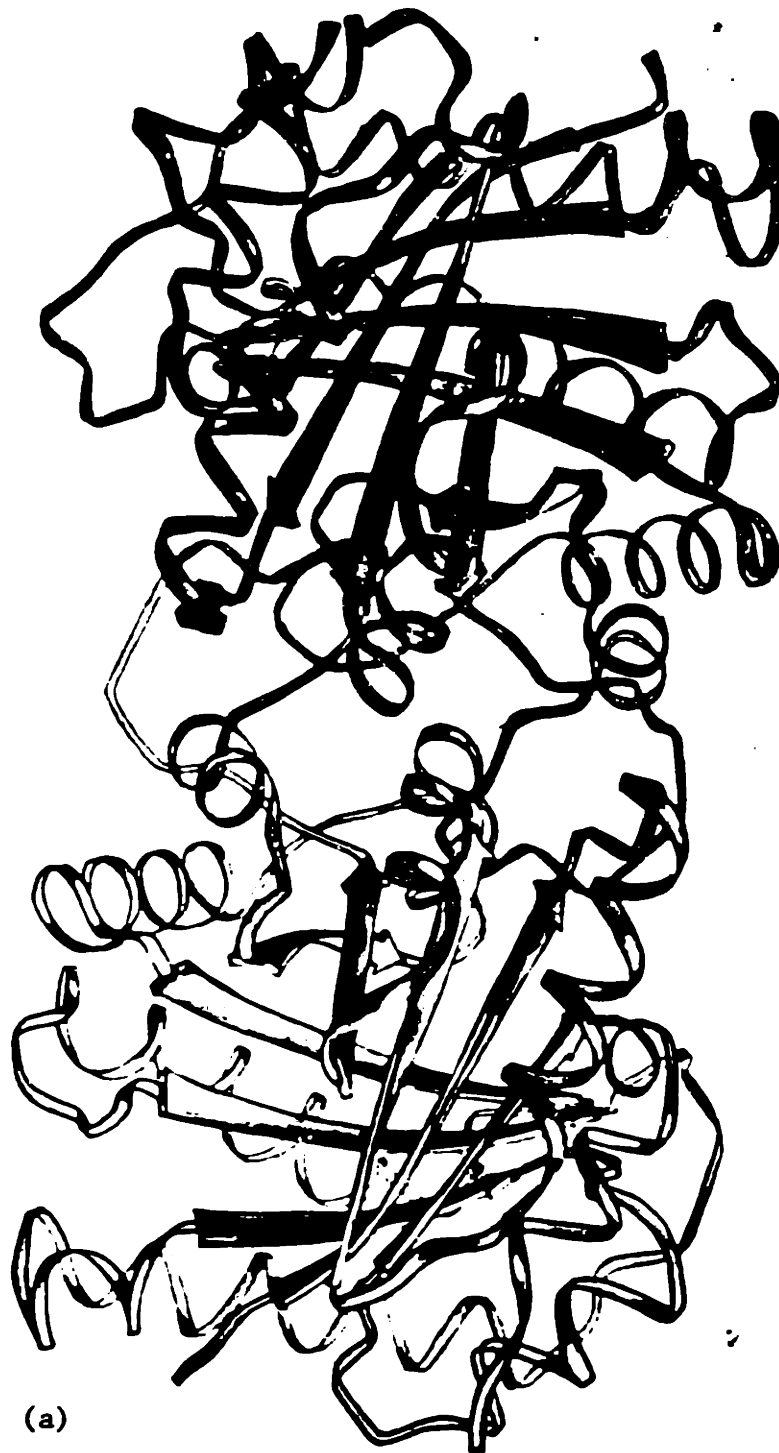
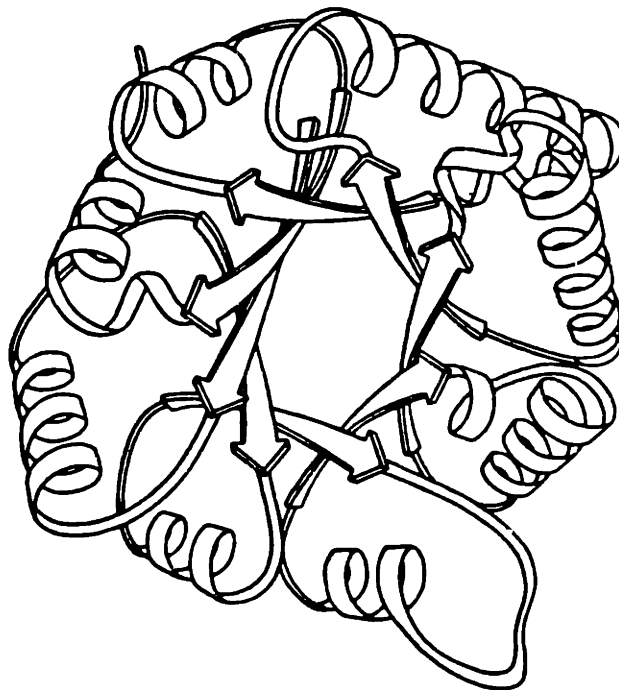
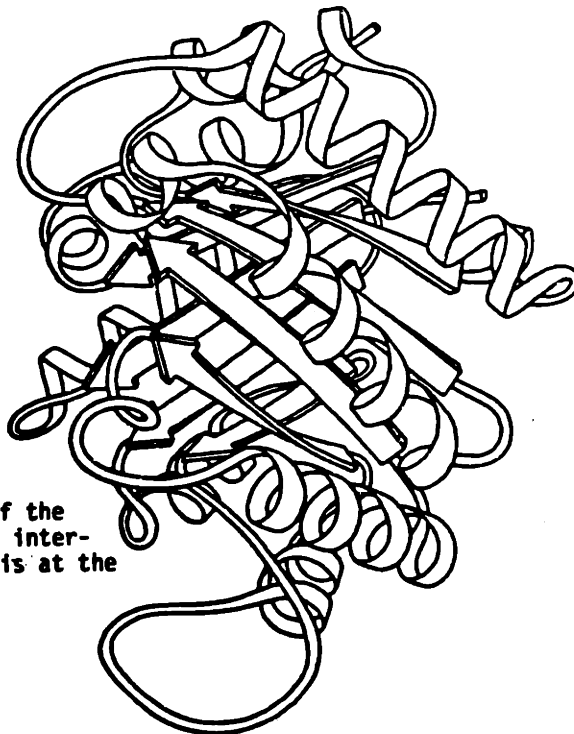


Figure 6. Ribbon diagrams of the structure of chicken muscle triose phosphate isomerase. Flat elongated features represent strands of β -sheet and curled regions represent α -helices. (a). A view of the dimer looking down the molecular two-fold rotation axis (27,28).

Figure 6. Continued . . .



(b) A view of the monomer looking down the axis of the beta-barrel.



(c) A side view of the monomer. The inter-subunit loop is at the bottom.

core of the subunit interface surrounding the molecular two-fold rotation axis is more polar. Since the dimer is the asymmetric unit, the subunits make different intermolecular contacts and adopt slightly different conformations in the crystals. This means that the molecular two-fold axis is inexact. The root mean square deviation in the positions of equivalent α -carbons calculated from the unrefined coordinates is 1.2 \AA (91).

The intersubunit loop (residues 70-80) is also involved in establishing the geometry of the active site, which is located in a cleft containing Glu 165, the catalytic base, at the carboxyl end of the β -barrel (Figure 7). Thus, each active site is formed by residues from both subunits. This is consistent with the idea that the dimer is the catalytically active species (83). The main chain amides Phe 74' and Thr 75' in the intersubunit loop are hydrogen bonded to the carboxyl group of Glu 97 (where prime denotes residues from the other monomer). Glu 97 is central in determining the geometry of the active site. This residue interacts closely with the ϵ -amino group of Lys 13 in the active site cleft, presumably forming an ion pair. The position of Lys 13 is also constrained by a hydrogen bond between the main chain amide of Met 14 and the main chain carboxyl group of Gly 72' in the intersubunit loop. Glu 97 is also in van der Waals contact with the imidazole side chain of His 95, but the geometry is not appropriate for the formation of a hydrogen bond. Instead, the $N_{\delta 1}$ of His 95, acting as an acceptor, is hydrogen bonded to the main chain amide of Glu 97. $N_{\delta 1}$ of His 95 is also in van der Waals contact with O_{γ} of Ser 96 which, in turn, makes a close contact with Glu 165.

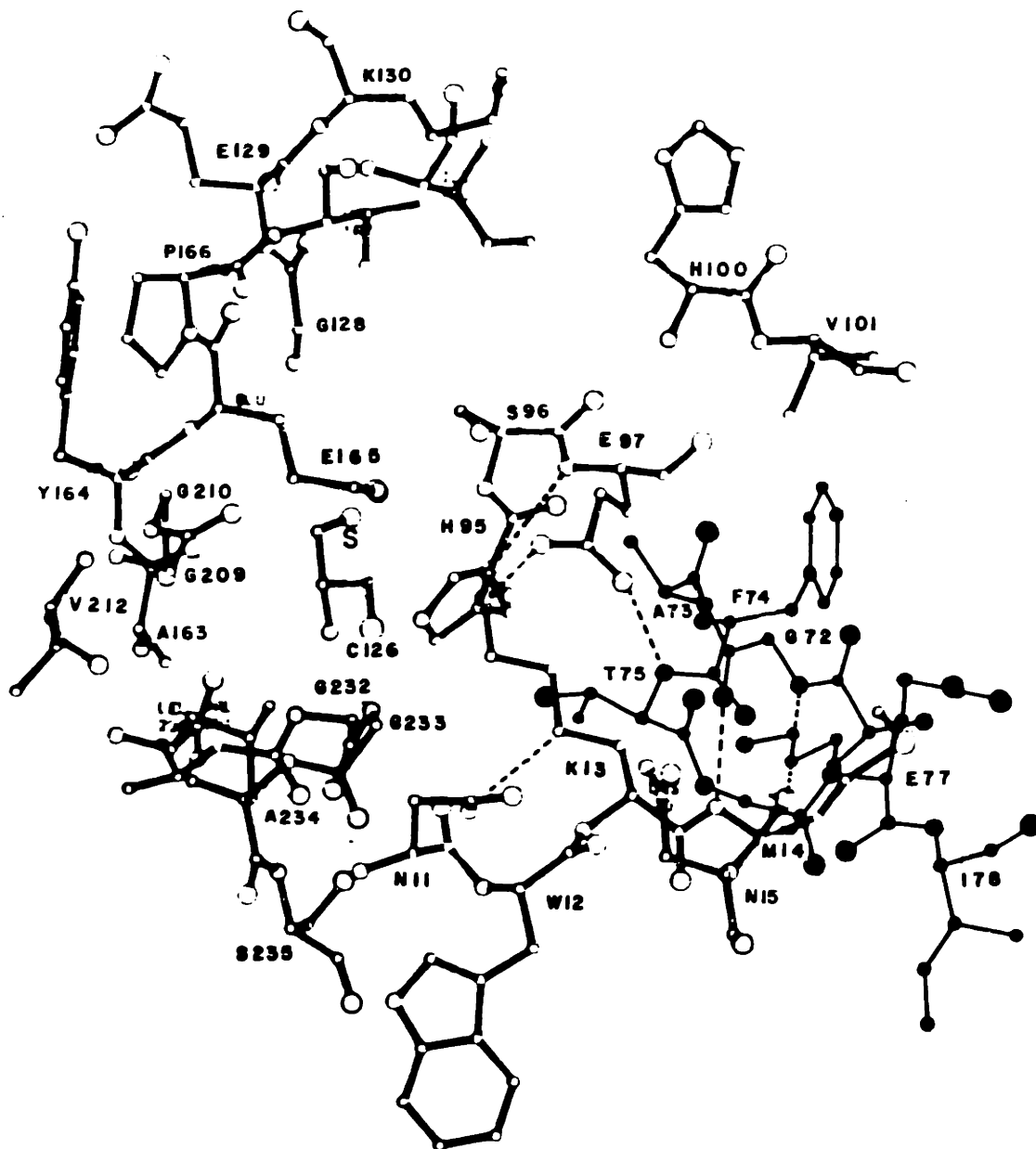


Figure 7. Active site region of one subunit of chicken muscle triose phosphate isomerase. Amino acid residues within 12 Å of the sulfate position are shown. Atoms drawn as solid balls are in the intersubunit loop of the second monomer. From Banner, *et al* 1975 (27).

Thus Glu 165, Ser 96, His 95 and Lys 13 line the active site pocket. Glu 97 is buried, and each active site involves residues from both subunits of the dimer.

The mother liquor of the chicken enzyme crystals is greater than 3M in sulfate, a competitive inhibitor, and a sulfate ion was found in the catalytic center. It appears to be bound to the amide nitrogen of Gly 232 and Gly 233, in a position to interact with the positive end of the dipole moment of the C-terminal helix (92). The sulfate is also in contact with the O_γ of Ser 211.

Studies were undertaken to probe the interactions of the enzyme with substrates and inhibitors. In general, these experiments were plagued by a number of disarming technical difficulties. 1) In early studies, the binding of ligands in the crystals caused large changes in the unit cell dimensions, thwarting the calculation of difference maps (93). 2) Under conditions which stabilized the unit cell dimensions, ligand binding disordered the crystals and frustrated data collection beyond 6Å resolution. 3) Since the crystals were stabilized in high concentrations of the competitive inhibitors sulfate and phosphate, the occupancy of substrates was low and difference density in the sulfate/phosphate site was non-existent. Enzymatic fragmentation of the substrates during data collection in sealed capillaries also reduced the occupancy of DHAP. This decomposition reaction also made cocrystallization problematic. 4) Ligands bound to only one subunit of the enzyme in the crystals; heterologous intermolecular contacts

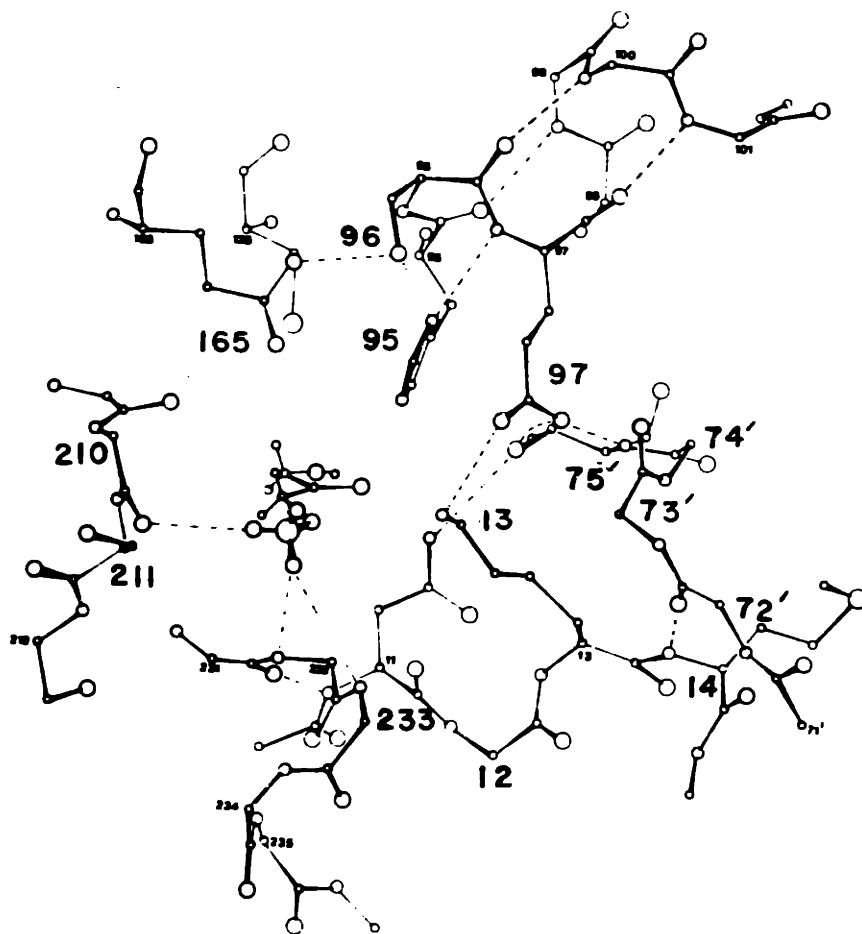


Figure 8. Dihydroxyacetone phosphate built into the active site of chicken TIM (D.C. Phillips and I.A. Wilson, unpublished results). The phosphate moiety is at the front. Residues 169 to 176, which move to interact with the phosphate, would lie across the center of the drawing and are omitted for clarity. The positions of the side chains of Glu 165, Glu 97, Ser 96, His 95 and Lys 13 in the active site cleft are indicated. Primed residues are in the other monomer.

apparently prevented ligand binding to the other subunit. Because of the clear possibility of perturbations of the enzyme-ligand interactions brought about by lattice forces, the level of confidence that could be ascribed to the results was reduced.

Nonetheless, several results of the ligand binding studies with the chicken enzyme crystals provided a provocative view of the catalytic events. Inorganic phosphate bound in the position of the sulfate ion in one active site of the dimer. A large movement of an external loop (residues 169-176) toward the active site accompanied phosphate binding (28,94). The low resolution difference electron density for the loop was consistent with an interaction between the main chain of Thr 172 and the phosphate oxygens. The change in the average position of Thr 172 was greater than 8\AA !

DHAP and phosphoglycolate bound in the active site of one subunit with their phosphate moieties presumably in the sulfate/phosphate position. Residues 169-176 moved to interact with the phosphate group, and other conformational changes were evident in both subunits. The difference density corresponding to bound DHAP stretched from the sulfate/phosphate position toward Glu 165 and His 95 in the active site pocket. Its length suggested that the substrates are bound in an extended conformation. Model building studies showed that DHAP and GAP could be built into the difference density such that the carbonyl oxygens interact with His 95 and Lys 13, and a proton from either C-1 or C-2 could be abstracted by Glu 165 (Figure 8) (28,94). These studies

did not illuminate the role, if any, of Ser 96 during the catalytic cycle.

Analysis of non-isomorphous crystals of the enzyme-DHAP complex at 3.0 Å resolution is currently under way at the University of Oxford (P.J. Artymuik and D.C. Phillips, unpublished results).

One feature of the triose phosphate isomerase structure that has emerged since its elucidation in 1975 is its extreme similarity to domains of other proteins. At least five enzymes which act on carbohydrates contain domains which mimick the fold of TIM. These proteins include cat muscle pyruvate kinase (95,96), spinach glycolate oxidase (97), Streptomyces albus xylose isomerase (M. Lewis, personal communication), 2-keto-3-deoxy-6-phosphogluconate (KDGP) aldolase (98,99) and Taka-amylase (100). In pyruvate kinase, glycolate oxidase and Taka-amylase, the active sites have been located at the mouth of the β -barrel. The amino acid sequences of these proteins are not known, so the basis of the structural homologies has not been determined. The gene for yeast pyruvate kinase has been cloned, and DNA sequencing studies have been initiated (G. Kawasaki, T. Alber and R. Davenport, unpublished results).

In summary, the crystal structure of chicken muscle TIM has revealed the fold of the molecule and the geometry of the active site residues in the free enzyme. Ligand binding studies have located the site of substrate binding and suggested that the enzyme undergoes substantial conformational changes during the catalytic cycle. The

identification of Glu 165 as the essential base is consistent with model building studies of the substrate in the active site. Ser 96, His 95 and Lys 13 are candidates for the catalytic electrophile(s). Because only low resolution data could be collected on the complexes, the details of the conformations of the substrate and enzyme could not be discerned. In addition, the low resolution electron density maps gave no information about the ratio of substrate species at the active site. Due to the technical difficulties with the chicken enzyme crystals described above, our attention was focused on the reaction and structure of yeast TIM.

References

1. Blake, C.C.F., Koenig, F., Mair, G.A., North, A.C.T., Phillips, D.C. and Sarma, V.R. (1975), *Nature*, 106, 757-763.
2. Doscher, M.S. and Richards, F.M. (1963), *J. Biol. Chem.*, 238, 2399.
3. Rossi, L.G. and Bernhard, S.A., (1970), *J. Mol. Biol.*, 49, 85-91.
4. Fink, A.T. and Ahmed, A.I. (1976), *Nature*, 263, 294-297.
5. Koshland, D.E., Jr. and Neet, K.E. (1968), *Ann. Rev. Bioch.*, 37, 359-410.
6. Johnson, L.N., Phillips, D.C. and Rupley, J.A. (1968), *Brookhaven Symp. Biol.*, 21, 120-138.
7. Phillips, D.C. (1966), *Sci. Am.*, 215, 78-90.
8. Blake, C.C.F., Johnson, L.N., Mair, G.A., North, A.C.T., Phillips, D.C. and Sarma, V.R., *Proc. Roy. Soc. Lond.* 167, 378-388.
9. Phillips, D.C. (1967), *Proc. Nat. Acad. Sci. USA*, 57, 484-495.
10. Warshel, A. (1978), *Proc. Nat. Acad. Sci. USA*, 75, 5250-5254.
11. Schindler, M. Assaf, Y., Sharon, N. and Chipman, D.M. (1977), *Biochemistry*, 16, 423-431.
12. Warshel, A. and Levitt, M. (1976), *J. Mol. Biol.*, 102, 227-249.
13. Reeke, G.N. Jr, Hartsuck, J.A., Ludwig, M.L., Quioco, F.A., Steitz, T.A. and Lipscomb, W.N. (1967), *Proc. Nat. Acad. Sci. USA*, 56, 2220.
14. Blow and Steitz, T.A. (1970), *Ann. Rev. Biochem.*, 39, 63-100.
15. Matthews, B.W. (1976), *Ann. Rev. Phys. Chem.*, 27, 493-523.
16. Eisenberg, D., "X-ray Crystallography and Enzyme Structure", in The Enzymes, ed. P. Boyer, 3rd. Edition, Vol. I, Academic Press, New York, 1970.
17. Douzou, P., Cryobiochemistry, An Introduction, Academic Press, London, 1977.
18. Petsko, G.A. (1975), *J. Mol. Biol.*, 96, 381-392.
19. Alber, T., Petsko, G.A. and Tsernoglou, D. (1976), *Nature*, 263, 297-300.

20. Fink, A.L. and Petsko, G.A. (1981), *Advances in Enzymology*, 52, 177-246.
21. Wolfenden, R., (1974), *Molec. Cell. Biochem.*, 3, 207-211.
22. Albery, W.J. and Knowles, J.R. (1976), *Biochemistry*, 15, 5631-5640.
23. Kolb, E., Harris, J.I. and Bridgen, J. (1974), *Biochem. J.*, 137, 185-197.
24. Furth, M.J., Milman, J.D., Priddle, J.D. and Offord, R.E., (1974), *Biochem.*, 139, 11-25.
25. Corran, P.H. and Waley, S.G. (1975), *Biochem, J.*, 145, 335-344.
26. Artavanis-Tsakonas and Harris, J.I. (1980), *Eur. J. Biochem.*, 108, 599-611.
27. Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., Pogson, C.I., Wilson, I.A., Corran, P.H., Furth, A.J., Milman, J.D., Offord, R.E., Priddle, J.D. and Waley, S.G. (1975), *Nature*, 255, 609-614.
28. Phillips, D.C., Rivers, P.S., Sternberg, M.J.E., Thornton, J.M. and Wilson, I.A. (1977), *Biochem. Soc. Trans*, 5, 642-647.
29. Artymuik, P.J., Blake, C.C.F., Grace, D.E.P., Oatley, S.J., Phillips, D.C. and Sternberg, M.J.G. (1979), *Nature*, 280, 563-568.
30. Alber, T., Hartman, F.C., Johnson, R.M., Petsko, G.A. and Tsernoglou, D. (1981), *J. Biol. Chem.*, 256, 1356-1361.
31. Meyerhof, O. and Lohmann, K., (1934), *Biochem. Z.*, 273, 413.
32. Meyerhof, O. and Kiessling, W. (1935), *Biochem. Z.*, 279, 40.
33. Bloom, B. and Topper, T.J. (1956), *Science*, 124, 982.
34. Reider, S.V. and Rose, I.A. (1956), *Fed. Proc.*, 15, 337.
35. Topper, Y.J. (1957), *J. Biol. Chem.*, 225, 419.
36. Reider, S.V. and Rose, I.A. (1959), *J. Biol. Chem.*, 234, 1007.
37. Herlihy, J.M., Maister, S.G., Albery, W.J. and Knowles, J.R. (1976), *Biochemistry*, 15, 5601-5607.
38. Albery, W.J. and Knowles, J.R. (1976), *Biochemistry*, 15, 5588-5600.
39. Knowles, J.R., Leadley, P.F. and Maister, S.G. (1971), *Cold Spring Harbor Symp. Quant. Biol.*, 36, 157-164.
40. Albery, W.J. and Knowles, J.R. (1976), *Biochemistry*, 15, 5627-5631.

41. Knowles, J.R. and Albery, W.J. (1977), *Acc. Chem. Res.*, 10, 105-111.
42. Fisher, L.M., Albery, W.J. and Knowles, J.R. (1976), *Biochemistry* 15, 5621-5626.
43. Maister, S.G., Pett, C.P., Albery, W.J. and Knowles, J.R. (1976), *Biochemistry* 15, 5607-5612.
44. Fletcher, S.J., Herlihy, J.M., Albery, W.J. and Knowles, J.R. (1976,) *Biochemistry*, 15, 5612-5617.
45. Leadlay, P.F., Albery, W.J. and Knowles, J.R. (1976), *Biochemistry* 15, 5617-5620.
46. Hall A. and Knowles, J.R. (1975). *Biochemistry*, 14, 4348-4352.
47. Quioco, F.A. and Richards, F.M. (1966), *Biochemistry*, 5, 4062.
48. Rose, I.A. (1981), *Phil. Trans, Roy. Soc. London B*, 263.
49. Iyengar, R. and Rose, I.A. (1981a), *Biochemistry*, in press.
50. Iyengar, R. and Rose, I.A. (1981b), *Biochemistry*, in press.
51. Reynolds, S.J., Yates, D.W. and Pogson, C.I. (1971), *Biochem. J.*, 122, 285-297.
52. Belasco, J.G. and Knowles, J.R. (1980), *Biochemistry*, 19, 472-477.
53. Fink A.L. and Geeves, M.A. (1979), *Methods in Enzymology*, 63, 336.
54. Makinen, M.W. and Fink, A.L. (1977), *Ann. Rev. Biophys. Bioeng.*, 6, 301.
55. Webb, M.R., Strandring, D.N. and Knowles, J.R.(1977), *Biochemistry*, 16, 2738-2741.
56. Deslongchamps, P., Atlani, P., Frehel, D. and Malaval, A. (1972), *Can. J. Chem.*, 50, 3405-3408.
57. Rose, I.A. and O'Connell, G.L. (1961), *J. Biol. Chem.*, 236, 3086.
58. Rose, I.A. (1962), *Brookhaven Symp. Biol.*, 15, 293.
59. Hartman, F.C. (1977), *Biochemistry*, 10, 146.
60. Miller, J.C. and Waley, S.G. (1971), *Biochem. J.*, 123, 163.
61. DeLaMare, S., Coulson, A.F.W., Knowles, J.R., Priddle, J.D. and Offord, R.E. (1972), *Eur. J. Bioch.*, 129, 321.
62. Schray, K.J., O'Connell, G.L. and Rose, I.A. (1973), *J. Biol. Chem.*, 248, 2214.

63. Norton, I.L. and Hartman, F.C. (1972), *Biochemistry*, 11, 4435.
64. Waley, S.G., Miller, J.C., Rose, I.A. and O'Connell (1970), *Nature*, 227, 181-182.
65. Hartman, F.C. and Gracy, R.W. (1973), *Biochem. Biophys. Res. Comm.* 52, 388.
66. Waley, S.G. (1972), *Biochem. J.*, 126, 255.
67. Hartman, F.C., La Muraglia, G.M., Tomozawa, Y. and Wolfenden, R. (1975), *Biochemistry*, 14, 5274-5279.
68. Webb, M.R. and Knowles, J.R. (1975), *Biochemistry*, 14, 4692-4698.
69. Noltmann, E.A., in *The Enzymes*, Vol. VI, P.D. Boyer, ed., Academic Press, New York, 1972, 271.
70. Trentham, D.R., McMurray, C.H., and Pogson, C. (1969), *Biochem. J.*, 114, 19-24.
71. Jones, R.B. and Waley, S.G. (1979), *Biochem. J.* 179, 623-630.
72. Campbell, I.D., Jones, R.B., Kiener, P.A. and Waley, S.G. (1979), *Biochem. J.*, 179, 607-621.
73. Belasco, J.G., Herlihy, J.M. and Knowles, J.R. (1978), *Biochemistry*, 15, 2971-2978.
74. Meyerhof, O. and Junowicz-Kocholaty, R. (1943), *J. Biol. Chem.*, 149, 71.
75. Lowry, O.H. and Passonneau, J.V. (1964), *J. Biol. Chem.*, 234, 31.
76. Burton, P.M. and Waley, S.G. (1968), *Biochem. Biophys. Acta*, 151, 714.
77. Veech, R.L., Raijman, L., Dalziel, K. and Krebs, H.A. (1969), *Biochem. J.*, 115, 837.
78. Plaut, B. and Knowles, J.R. (1972), *Biochem. J.*, 129, 311-320.
79. Hartman, F.C. and Ratrie, H. (1977), *Biochem. Biophys. Res. Comm.*, 77, 746-752.
80. Collins, K. (1974), *J. Biol. Chem.*, 249, 136-142.
81. Wolfenden, R. (1969), *Nature*, 233, 704.
82. Johnson, L.N. and Wolfenden, R. (1970), *J. Mol. Biol.*, 47, 93-100.
83. Waley, S.G. (1973), *Biochem. J.*, 135, 176-172.

84. White, C.J.B. and Fell, D.A. (1975), *Biochem. Biophys. Res. Comm.*, 67, 1013-1018.
85. White, C.J.B. and Fell, D.A. (1976), *Biochem. Soc. Trans.*, 4, 620-622.
86. Artavanis, S. (1974), Ph.D. Thesis, University of Cambridge.
87. Johnson, L.N. and Waley, S.G. (1967), *J. Mol. Bio.*, 29, 321-322.
88. Krietsch, K.G., Pentcher, P.G., Klingenburg, H., Hofstatter, T., and Bucher, T. (1970), *Eur. J. Biochem.*, 14, 289-
89. Meyer-Arendt, E., Beisenherz, G. and Bucher, T. (1953), *Naturwissenschaften*, 40, 59.
90. Rozacky, E.E., Sawyer, T.H., Barton, R.A. and Gracy, R.W. (1971), *Arch. Biochem. Biophys.*, 141, 312.
91. Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C. and Wilson, I.A. (1976), *Biochem. Biophys. Res. Comm.*, 72, 146-155.
92. Hol, W.G.J., van Duijnen, P.T. and Berendsen, H.J.C. (1978), *Nature*, 273, 443-446.
93. Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., and Pogson, C.I. (1971), *Cold Spring Harbor Symp. Quant. Biol.*, 36, 151-155.
94. Alber, T., Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., Rivers, P.S., and Wilson, I.A. (1981), *Phil. Trans. Roy. Soc. Lond. B.*, 293, 159-171.
95. Stammers, D.K. and Muirhead, H. (1977), *J. Mol. Biol.*, 112, 309-316.
96. Stammers, D.K., Levine, M., Stuart, D.I. and Muirhead, H. (1977), *Biochem. Soc. Trans.*, 5, 642-647.
97. Lindqvist, Y. and Branden, C.-I. (1980), *J. Mol. Biol.*, 143, 201-211.
98. Mavridis, I.M. and Tulinski, A.L. (1976), *Biochemistry*, 15, 4410-4417.
99. Richardson, J.S. (1981), *Adv. Prot. Chem.*, 34, 167-339.
100. Matsuura, Y., Kusunoki, M., Harada, W., Tanaka, N., Iga, Y., Yasuoka, N., Toda, H., Narita, K. and Kakudo, M. (1980), *J. Biochem.*, 87, 1555-1558.

CHAPTER II.

Nucleotide Sequence of the Yeast Triose Phosphate Isomerase Gene

I. Introduction

Though triose phosphate isomerases from four organisms had been sequenced (1-4), the absence of information about the primary structure of the yeast enzyme limited the interpretation of crystallographic and solution studies. Despite the similarities in their catalytic properties (5,6), the amino acid composition of triose phosphate isomerases from two yeast strains suggested that the yeast proteins do not share the 85% sequence homology of the enzymes from rabbit, chicken, and coelacanth (7,8). In order to ascertain the amino acid sequence of yeast TIM, the gene coding for the protein (TPI1) was sequenced. This work employed a new method of selecting eukaryotic promoter mutants developed by Dr. Glenn Kawasaki, and information about the expression of the protein in yeast and *E. coli* was obtained.

In organisms and tissues active in glycolysis, TIM is an abundant protein. By weight, it is 4% of chicken breast muscle and 2% of the soluble protein in wild type yeast strains grown on glucose (9). The transcript of the single yeast gene is present in large enough quantities to be readily detected by shotgun cloning of cDNA made from

partially purified yeast message (M. Holland, personal communication). A yeast mutant (GCR1) has been isolated in which the expression of genes coding for a number of glycolytic enzymes, including triose phosphate isomerase, is decreased (9).

The structural basis of efficient expression and regulation of yeast genes is not well understood. A number of yeast genes, including those coding for the glycolytic enzymes enolase and glyceraldehyde phosphate dehydrogenase (GAPDH), have been sequenced yet no regions of obligatory homology have been found among the flanking sequences and no sequences which would account for differences in levels of expression have been identified (10,11). By analyzing transcripts produced by deletion mutants of the yeast iso-1-cytochrome C gene (CYC1), Faye and coworkers have shown that the 5' flanking sequence may be sufficient to determine the pattern of transcription of the gene (12). They found that deletion of sequences between 9 and 242 bases upstream from the coding region affected the level and pattern of initiation of poly A containing CYC1 messages that could be detected by hybridization. Guarente and Ptashne constructed hybrid CYC1-lacZ genes and studied the level of β -galactosidase activity elicited by the fusion genes and several deletion mutants (13). They concluded that a sequence between 300 and 700 bases upstream from the coding region was important for the regulation of CYC1 by glucose and that a sequence between -300 and -242 may influence the level of gene expression.

In this chapter, the sequence of the yeast gene coding for triose phosphate isomerase and the sequence of a mutant which is expressed in

both yeast and *E. coli* is presented. A novel method of selecting eukaryotic promoter mutants is described. The sequences of the 5' and 3' flanking regions are compared with other yeast sequences, including those of four highly expressed genes coding for enolase and GAPDH. The choice of codons is analyzed and compared to that in other yeast genes. The gene sequence is examined in a search for clues concerning the evolution of the protein. The amino acid sequence deduced from the order of nucleotides is compared to the sequences of other triose phosphate isomerases.

II. Materials and Methods

The hosts and vectors used in this work are listed in Table 1. The cloning of a number of yeast genes coding for glycolytic enzymes, including triose phosphate isomerase, was carried out by Dr. Glenn Kawasaki, with whom I collaborated in these studies. In brief, the gene was isolated from a yeast library in pYE13 (14,15) by requiring complementation of the *tpi*⁻ and *leu*⁻ phenotypes of *S. cerevisiae* N582-4B. This was accomplished by growing the transformed N582-4B strain using glucose as the sole carbon source. The plasmid obtained by this process, pTPIC10, contained approximately 5 kilobase pairs (kb) of inserted yeast DNA.

pTPIC10 does not direct the synthesis of TIM in *E. coli* DF 502 grown on glucose. However, plating DF 502 which harbored pTPIC10 on minimal lactate agar led to the isolation of *tpi*⁺ transformants in which the

Table 1. Hosts and vectors.

Host strains

E. coli K12

- 1) DF 502: $\Delta(rha, pfkA, tpi)$ $pfkB1^+$, his^- , $pyrD^-$, edd^- , F^- , str^r .
- 2) RR1: F^- , $proA^-$, leu^- , thi^- , $lacY^-$, str^r , $r_{k^m k}^- EndoI^-$.

Saccharomyces cerevisiae

- 1) N582-4B: $atpi^-$, GAL^- , MAL^- , $LEU2^-$, can^r , cyh^r .

Vectors

- E. coli -- pBR322: amp^r , tet^r .
- Yeast -- CV13: 2μ , $LEU2$, pBR322.

Enzyme activity induced by TPI1 in yeast and E. coli in μM DHAP $min^{-1} mg^{-1}$ at 25° C.

Plasmid	Host			
	<u>E. coli</u>		<u>S. cerevisiae</u>	
	RT 142 (tpi^+)	DF 502	N501-1B (tpi^+)	N582-4B
no plasmid	6.309	0.000	18.30	0.00
pTPIC10		0.004		188.94
PTPIC10up		0.942		17 - 20
pFG1		+		-

plasmid had suffered an insertion or deletion. The smallest of these plasmids produced the highest level of TIM activity in E. coli DF 502, and it was named pTPIC10up. A 2.1 kb Eco R1 fragment of this plasmid containing 1.8 kb of yeast sequence was subcloned in the tet^r gene of pBR322, and the resulting plasmid was named pFG1.

pTPIC10 and pTPIC10up were isolated from yeast host strains as described by Nasmyth (15). Plasmids were purified from E. coli RR1 grown on M9 medium using the cleared lysate technique (16). In order to increase the number of copies of the plasmid in the host bacterium, chloramphenicol (100 mg/l) was added to cultures in log phase 10 to 15 hours prior to harvesting the cells. After equilibrium sedimentation in density gradients of CsCl, the purified plasmids were stored frozen in 1 mM tris-HCl, 1 mM EDTA at pH 7.9.

Restriction maps were obtained by reconciling the sizes of the products obtained by digesting the plasmids or purified fragments of them with single and multiple restriction enzymes. The method of Smith and Bernstein, involving partial restriction enzyme digestion, was also used (17). Fragments were visualized in 0.7 to 1.4% agarose gels or in 7 to 10% acrylamide gels by ultraviolet photography of gels stained with ethidium bromide (1 mg/l) or by autoradiography of fragments labelled with ³²P using the Klenow fragment of DNA polymerase (18-20).

Restriction enzymes used for mapping and sequencing experiments were obtained from New England Biolabs (NEB) and Bethesda Research Labs (BRL). Bacterial alkaline phosphatase was from Worthington Enzymes. T4

polynucleotide kinase and the large fragment of DNA polymerase were purchased from NEB, Boehringer Mannheim and the New England Nuclear Corporation (NEN). γ - ^{32}P -ATP and α - ^{32}P -NTP's were obtained from NEN.

DNA sequencing was carried out by the methods of Maxam and Gilbert (21) using reactions specific for cleavage at guanine, cytosine, purines, pyrimidines and adenine more than cytosine. Reagents used in the sequencing experiments were the highest grades available from the Aldrich, Eastman-Kodak and Sigma Chemical Companies. Agarose, acrylamides and urea were the highest grades available from Bio-Rad and BRL.

Analysis of the DNA sequences was aided by the use of a Digital Equipment Corporation VAX 11/780 computer using programs which were courteously provided by Dr. Roger Stader (22,23).

Enzyme activities in cell lysates were measured as described by Maitra and Lobo using a Gilford Recording Spectrophotometer (24).

III. Results

A plasmid that allows a yeast strain which lacks triose phosphate isomerase to grow on glucose-containing medium was isolated from the Nasmyth library of yeast genomic DNA in pYE13 (15). This plasmid was named pTPIC10. Extracts of the tpi⁻ strain (N582-4B) harboring pTPIC10 contained 9 to 13 times the triose phosphate isomerase activity of wild

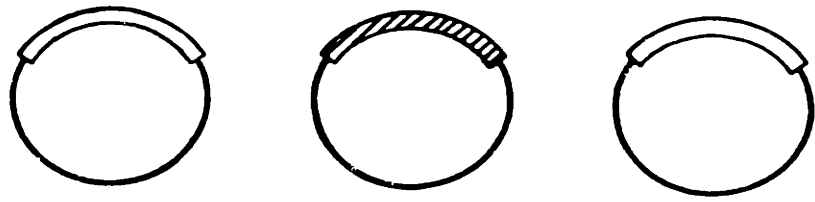
type strains, suggesting that the plasmid may encode the structural gene for the protein.

In order to locate the gene within the 5000 base pairs (bp) of yeast DNA inserted in pTPIC10 for sequencing studies, the following strategy was adopted. The plasmid was transfected into a strain of E. coli (DF502) in which the triose phosphate isomerase gene had been deleted, and the cells were plated on minimal agar containing lactate as the sole carbon source. This imposed a selection for at least a moderate level of expression of the enzyme, and it resulted in the isolation of three tpi^+ transformants. Restriction enzyme analysis of the plasmids from the tpi^+ transformants indicated that they had all suffered an insertion or deletion in the yeast sequence in vivo. The smallest plasmid which complemented the tpi^- allele of the host bacteria contained 2.2 kb of yeast DNA. It was named pTPIC10up. A 2.1 kb Eco R1 fragment of this plasmid containing 1.8 kb of the yeast sequence was inserted into the Eco R1 site of pBR322, and the resulting plasmid, pFG1, was also found to complement the tpi^- phenotype in E. coli. These experiments are summarized in Figure 1.

The levels of enzyme activity elicited by the insert of pTPIC10, pTPIC10up and pFG1 in tpi^- yeast and E. coli are shown in Table 1. Though the insert in pTPIC10up produces active enzyme in both organisms, the up mutation reduces expression approximately ten-fold in yeast.

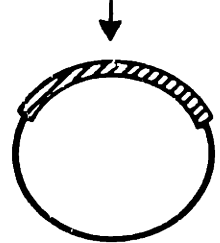
Partial restriction maps of pTPIC10 and pFG1 are shown in Figure 2. Since the region coding for the protein was expected to be a sizable

- Library of yeast genomic DNA partially digested with *Sau* 3A and cloned in the *Bam* HI site of CV13 (LEU2, 2 μ , pBR322)



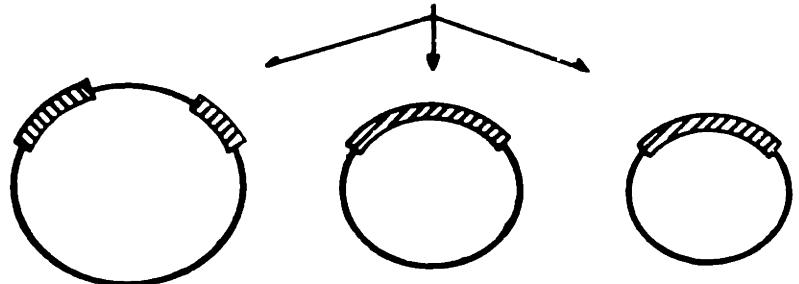
- Transform *S. cerevisiae* N582-4B (LEU2⁻, TPI⁻) and select for LEU2⁺ and TPI⁺.

pTPIC10



- Measure enzyme activity in extracts of N582-2D harboring the plasmid.
- Transform *E. coli* RR1, purify the plasmid and determine its restriction map.

- Transform *E. coli* DF502 (tpi⁻) and select for amp^r and tpi⁺.



pTPIC10up

- Measure enzyme activities and determine restriction maps of the plasmids from the tpi⁺ transformants. pTPIC10up produced the highest activity in *E. coli* and was expressed in yeast.

- Digest with *Eco* RI and subclone in pBR322 in *E. coli* DF502. Select for amp^r and tpi⁺.

- Transform *E. coli* RR1, purify the plasmid, determine its restriction map and DNA sequence.

pFGI

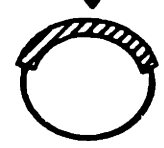


Figure 1. Isolation of the yeast triose phosphate isomerase gene, *TPI1*.

portion of the pFG1 insert, the gene was located by DNA sequencing. The sequence flanking the unique Bgl II site was determined and translated in both directions in all possible reading frames. One frame lacked stop codons and specified an amino acid sequence which was approximately 55% homologous with residues 33 to 173 in TIMs from rabbit, chicken, and coelacanth (1-3). The Bgl II site coded for amino acids 99 and 100 in the yeast TIM sequence. Restriction maps of pTPIC10 and pFG1 were used to establish the relationship between the two inserts, and regions of interest were sequenced using the strategy summarized in Figure 2.

The identity of 3000 bases from both strands of pTPIC10 and pFG1 was determined and compiled into a sequence of 2000 base pairs. On the basis of consistent restriction maps and the homology of the DNA sequences from -65 near the deletion in pFG1 to the Bgl II site at 300 in the coding region, it was assumed that the up mutation does not involve bases downstream from the Bgl II site. The presumption that the inserts in the two plasmids are identical downstream from the immediate vicinity of the up mutation is also supported by the ability of the up mutant to specify the production of active enzyme in yeast and E. coli.

About half of the sequence presented here was verified by sequencing both strands. The sequence of the coding region specifies an amino acid sequence which is consistent with the amino acid composition and the sequence of the N-terminal 16 amino acids of TIM from baker's yeast (7). The accuracy of the sequence of the coding region is also supported by the homologies, discussed in detail below, among the yeast TIM amino acid sequence and those of TIM's from several other organisms.

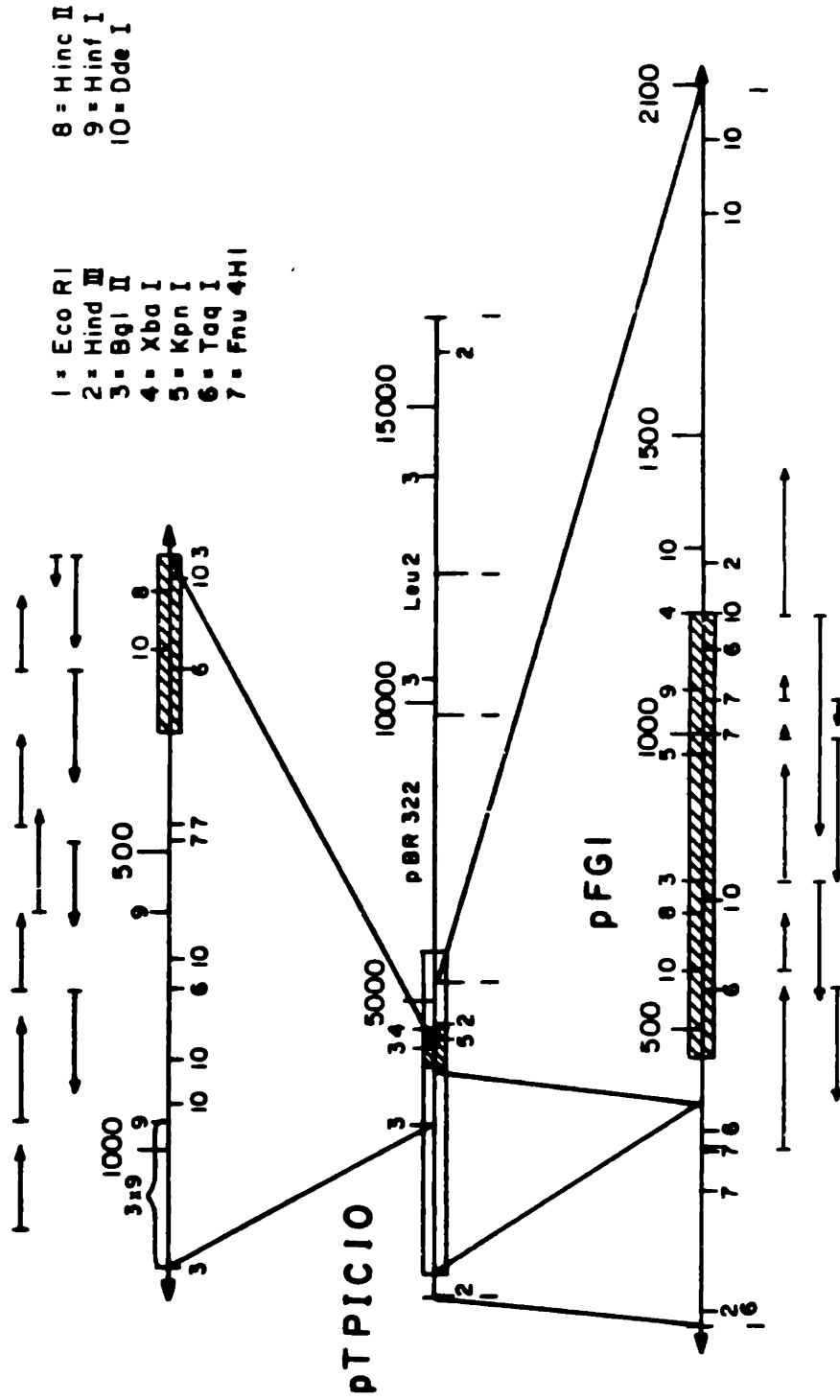


Figure 2. Restriction maps of pTPIC10 and the strategy used to sequence TP11. The center line shows pTPIC10 opened at the Eco RI site in pBR322. The boxed region is the yeast insert, and the shaded section within the box is the coding region of TP11. A fragment of pTPIC10 generated by digestion with Bgl II was used to sequence the 5' flanking region and the first 300 base pairs of the coding region (top line). The arrows indicate the direction and amount of sequence determined from each end-labeled restriction fragment. The bottom line shows the insert of pFGI, and the lines connecting it to pTPIC10 indicate the structural relationship of the two plasmids. The TP11 coding region is shaded.

Although systematic fitting of the amino acid sequence to the electron density map of the protein has not yet been completed, it is already apparent that many residues are credibly represented by the density. Crystallographic refinement at high resolution will provide a more stringent test of the coding region sequence. The Taq I site located at -436 in pTPI10 is the most likely location for a possible error of omission, since the assignment of fragments bordering the site was made on the basis of restriction mapping experiments rather than by sequencing an overlapping fragment. An omission larger than 15 bp at this site is unlikely.

The nucleotide sequence of TPI1 and the corresponding amino acid sequence of the protein is shown in Figure 3. The region coding for TIM is located in the only extensive open reading frame in the sequence and, on the basis of the amino acid sequence homologies discussed below, contains no introns. Since the N-terminal residue of the purified protein is alanine, it was of interest to determine the nature and extent of the post-translational cleavage. A comparison of the sequence of the N-terminal amino acids of the purified protein (7) with the sequence predicted from the nucleotide order indicates that only the initiating methionine is removed.

The pattern of codon usage in TPI1 and several other yeast genes is shown in Table 2 (10,11,25-27). Thirty three of the 64 possible codons are present in the yeast triose phosphate isomerase gene. Nine amino acids -- arg, cys, gln, glu, gly, his, tyr, met and trp -- are each represented by a single codon. Asn and pro are specified by a single

GAATCCATCAATAGATACGTCCTGAGGACCGTGCTACCCAAATGGACTGATTGTGAGGGA
 -458 -450 -440 -430 -420 -410 -400
 GACCTAACTACATACTGTTTAAGATTACGGATATTTAACTTACTTAGAATAATGCCATTT
 -390 -380 -370 -360 -350 -340
 TTTTGAGTTATAATAATCTACGTTAGTGTGAGCGGGATTTAACTGTGAGGACCTTAAT
 -330 -320 -310 -300 -290 -280
 ACATTCAGACACTTCTGACGGTATCACCCCTACTTATTCCTTCGAGATTATATCTAGGAA
 -270 -260 -250 -240 -230 -220
 CCCATCAGGTTGGTGGAAAGATTACCCGTTCTAAGACTTTTCAGCTTCTCTATIGATGTT
 -210 -200 -190 -180 -170 -160
 ACATTTGGACACCCCTTTTCTGGCATCCAGTTTTTAATCTTCAGTGGCATGTGAGATTTCT
 -150 -140 -130 -120 -110 -100
 CCGAAATTAATTAAGCAATCACACAATTTCTCTGGATACCACCTCGGTTGAAACTGACA
 -90 -80 -70 -60
 GGTGGTTTGTACGCATGCTAATGCAAGGAGCCTATATACCTTTGGCTCGGCTGCTGTA
 -50 -40 -30 -20 -10
 ACAGGGAATATAAAGGGCAGCATAATTTAGGAGTTTGTGAACTTGCAACATTTACTATT
 -1
 TCCCTTTCTTACGTAATAATTTTTCTTTTTAATTTCTAAATCAATCTTTTTCAATTTTTG
 -110 -100 -90 -80 -70 -60
 TTTGTATTTCTTTTCTTGTCTTAAATCTATAACTACAAAAAACACATACATAAACTAAAAAT
 -50 -40 -30 -20 -10 -1
 ALA ARG THR PHE PHE VAL GLY GLY ASN PHE LYS LEU ASN GLY SER LYS GLN SER ILE LYS
 GCCTAGAACTTTCTTTGTCGGTGGTAACTTTAATTAACGGTCCAAACAATCCATTA
 10 20 30 40 50 60
 GLU ILE VAL GLU ARG LEU ASN THR ALA SER ILE PRO GLU ASN VAL GLU VAL VAL ILE CYS
 GGAATTTGTTGAAAGATTGAACACTGCTTCTATCCCAGAAAAATGTCGAAGTTGTTATCTG
 70 80 90 100 110 120
 PRO PRO ALA THR TYR LEU ASP THR SER VAL SER LEU VAL LYS LYS PRO GLN VAL THR VAL
 TCCTCCAGCTACCTACTTAGACTACTCTGTCTCTTTGGTTAAGAAGCCACAAGTCACTGT
 130 140 150 160 170 180
 GLY ALA GLN ASN ALA TYR LEU LYS ALA SER GLY ALA PHE THR GLY GLU ASN SER VAL ASP
 CGGTGCTCAAAACGCCCTACTTGAAGGCTTCTGGTCTTTTACCAGGTAAGAACTCCGTTGA
 190 200 210 220 230 240
 GLN ILE LYS ASP VAL GLY ALA LYS TRP VAL ILE LEU GLY HIS SER GLU ARG ARG SER TYR
 CCAATCAAGGATGTTGGTGCTAAGTGGGTTATTTGGGTCACCTCCGAAAGAAGATCTTA
 250 260 270 280 290 300
 PHE HIS GLU ASP ASP LYS PHE ILE ALA ASP LYS THR LYS PHE ALA LEU GLY GLN GLY VAL
 CTTCCAGGAAGATGACAAGTTTCATTGCTGACAAGACCAAGTTTCGCTTTAGGTC AAGGTG
 310 320 330 340 350 360
 GLY VAL ILE LEU CYS ILE GLY GLU THR LEU GLU GLU LYS LYS ALA GLY LYS THR LEU ASP
 CGGTGTCATCTTGTGTATCGGTGAAACTTTGGAAGAAAAGAAGGCCGGCTAAGACTTTGGA
 370 380 390 400 410 420
 VAL VAL GLU ARG GLN LEU ASN ALA VAL LEU GLU GLU VAL LYS ASP TRP THR ASN VAL VAL
 TGTGTTGTAAGACAATTTGAACGCTGTCTTGGAAAGAAGTTAAGGACTGGACTAACGTCGT
 430 440 450 460 470 480
 VAL ALA THR GLU PRO VAL TRP ALA ILE GLY THR GLY LEU ALA ALA THR PRO GLU ASP ALA
 TGTGCTTACGAACCACTCTGGGCCATTGGTACCGGTTTGGCTCTACTCCAGAAGATGC
 490 500 510 520 530 540
 GLN ASP ILE HIS ALA SER ILE ARG LYS PHE LEU ALA SER LYS LEU GLY ASP LYS ALA ALA
 TCAAGATATTCACGCTTCCATCAGAAAGTTCTTGGCTTCCAAGTTGGGTGACAAGGCTGC
 550 560 570 580 590 600
 SER GLU LEU ARG ILE LEU TYR GLY GLY SER ALA ASN GLY SER ASN ALA VAL THR PHE LYS
 CAGCGAATTGAGAATCTTATACGGTGGTTCCGCTAACGGTAGCAACGCCGTTACCTTCAA
 610 620 630 640 650 660
 ASP LYS ALA ASP VAL ASP GLY PHE LEU VAL GLY ALA SER LEU LYS PRO GLU PHE VAL
 GCACAAGGCTGATGTCGATGGTTTCTTGGTCCGGTGGTCTTTCTTTTGAAGCCAGAATTTGT
 670 680 690 700 710 720
 ASP ILE ILE ASN SER ARG ASN ***
 TGATATCATCAACTTAGAACTAAGATTAATATAATTATATAAAAAATATTATCTTCTTT
 730 740 747 +1 +10 +20 +30
 TCTTTATATCTAGTGTATGTAATAAATTAATGATGACTACGGAAAGCTTTTTTATATTGT
 +40 +50 +60 +70 +80 +90
 TCTTTTTTCACTCTGAGCCACTTAATTTTCGTGAATGTTCTTGTAAAGGGACGGTAGATT
 +100 +110 +120 +130 +140 +150
 ACAAGTGATACAACAAAAAGCAAGGCGCTTTTTCTAATAAAAAGAAGAAAAGCATTTAAC
 +160 +170 +180 +190 +200 +210
 AATTGAACACCTCTATATCAACAGAAGA
 +220 +230 +240

Figure 3. The nucleotide sequence and corresponding amino acid sequence of TP11.

Table 2. Codon utilization in TPI1 and other yeast genes. The frequency at which each of the 64 codons is present per hundred codons in the genes for triose phosphate isomerase, enolase (ENO) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and isozymes 1 and 2 of cytochrome C (CYC) is listed.

		<u>TPI</u>	<u>ENO & GAPDH</u>	<u>CYC</u>			<u>TPI</u>	<u>ENO & GAPDH</u>	<u>CYC</u>
ALA	GCA	0	0	0.9	LYS	AAA	0.8	0.9	6.2
	T	8.0	9.5	2.7		G	7.6	7.3	8.5
	C	2.0	2.2	2.7	MET	ATG	0.4	1.8	3.1
	G	0	0	0.4					
ARG	CGA	0	0	0	PHE	TTT	1.2	0.2	2.2
	T	0	0.1	0		C	3.2	3.2	1.3
	C	0	0	0	PRO	CCA	2.4	3.2	2.2
	G	0	0	0		T	0.4	0.1	1.8
	AGA	3.2	3.2	2.2		C	0	0	0
	G	0	0	0.4		G	0	0	0
ASN	AAT	0.4	0.1	1.3	SER	TCA	0	0	0.9
	C	4.4	4.2	4.9		T	2.8	3.4	1.8
ASP	GAT	3.2	1.9	2.2		C	2.8	3.9	0.4
	C	2.8	5.1	2.2		G	0	0	0.4
CYS	TGT	0.2	0.4	1.8		AGT	0	0	0.9
	C	0	0	0.4		C	0.8	0	0
GLN	CAA	2.8	1.9	0.9	THR	ACA	0	0	2.2
	G	0	0	1.3		T	2.8	2.6	1.8
						C	2.0	3.2	1.8
				G		0	0	1.8	
GLU	GAA	6.8	5.4	3.6	TRP	TGG	1.2	1.1	0.9
	G	0	0	2.2					
GLY	GGA	0	0	0.9	TYR	TAT	0	0.1	2.2
	T	8.8	7.9	7.1		C	2.4	2.5	2.2
	C	0	0	1.3	VAL	GTA	0	0	0.4
	G	0	0	1.3		T	5.2	5.1	0.9
HIS	CAT	0	0	2.2	C	5.2	4.2	0.4	
	C	1.2	2.4	0.9	G	0	0	0.9	
ILE	ATA	0	0	0.4	STOP	TAA	0.4	0.3	0.4
	T	2.4	2.6	2.2		G	0	0	0.4
	C	3.6	2.8	1.3		TGA	0	0	0
LEU	TTA	1.6	0.3	1.3					
	G	6.0	7.5	3.6					
	CTA	0	0.1	0.4					
	T	0	0	0.4					
	C	0	0	0					
	G	0	0	0					

codon with one exception. Ala, leu, lys and phe are specified by a pair of codons, one of which is present much more frequently than the other, and asp, thr and ileu are coded by two triplets which are present equally often. Only ser is represented by three codons, and in this case two -- TCT and TCC -- predominate. This pattern of codon usage is similar to that found in the four yeast genes coding for enolase and GAPDH (10,11). On average, the first of the three bases in a codon is likely to be a purine, the second is unlikely to be cytosine and the third is likely to be a pyrimidine, especially in those cases where two or more codons specify an amino acid. Presumably because of codon choice, a number of oligonucleotides appear more frequently than would be expected in a statistically random sequence of bases (for example, AAGAAG is present four times in the coding region).

Since the 3-dimensional structure of triose phosphate isomerase contains approximate eight-fold rotational symmetry, it was of interest to see if the sequence of the coding region contains extensive repeats or other special structures (28). Two overlapping sequences of 15 and 21 nucleotides are inexactly re-iterated within the coding region. These sequences are shown in Figure 4. The sequence CGGTGAACTTTGGA beginning at position 384 has 13 bp in common with the 15 bp sequence CCGTAAGACTTTGGA beginning at 399. These sequences about a 14 base pair (bp) sequence, GGAAGAAAAGAAGG, in which the order of bases has mirror symmetry and is complementary to the sequence TTCTTTCTT found in both the 5' and 3' flanking regions at -52 and +31. As shown in Figure 4, sequences coding for amino acids 130-136, 139-143 and 151-156 also have a high degree of homology. The homologous sequences are in the same

379	<u>ATC</u>	<u>GGT</u>	<u>GAA</u>	<u>ACT</u>	<u>TTG</u>	<u>GAA</u>	<u>GAA</u>	AAG	<u>AAG</u>	<u>GCC</u>	408	
126	ILE	GLY	GLU	THR	LEU	GLU	GLU	LYS	LYS	ALA	135	
406	<u>GCC</u>	<u>GGT</u>	<u>AAG</u>	<u>ACT</u>	<u>TTG</u>	<u>GAT</u>	<u>GTT</u>					426
135	ALA	GLY	LYS	THR	LEU	ASP	VAL					141
		451	<u>TTG</u>	<u>GAA</u>	<u>GAA</u>	<u>GTT</u>	<u>AAG</u>	<u>GAC</u>				468
		150	LEU	GLU	GLU	VAL	LYS	ASP				155

Figure 4. Similar oligonucleotide sequences within the coding region of the yeast triose phosphate isomerase gene are shown with their corresponding amino acid sequences. Homologous nucleotides are underlined.

reading frame and specify amino acids which adopt similar, but not identical, secondary structures in the protein. In addition to these direct repeats, the coding region contains a 14 bp palindrome beginning at position 219. Inverted repeats larger than 11 bp do not occur in the sequence, and large direct repeats which would suggest duplications of entire domains during the evolution of the protein have not been found.

The sequence of TPI1 also provides the basis for experimentally testable hypotheses about the expression of the gene. The triose phosphate isomerase mRNA is abundant in yeast grown on glucose (D. Fraenkel, unpublished results), and, as noted above, the experiments of Faye et al and Guarente and Ptashne have suggested that the efficiency of transcription of the yeast CYC1 gene is determined by the 5' noncoding sequence (12,13). We have sequenced 658 bp on the 5' side of the triose phosphate isomerase coding region and an additional 145 bp which, according to restriction mapping experiments, is separated from the sequence shown in Figure 3 by a gap of 30 to 50 bp (data not shown).

The sequence of the 5' flanking region has a number of notable features (Figure 5). Base sequences with mirror symmetry 22 and 16 nucleotides long begin at -391 and -297. Inexact palindromes spanning 22 and 21 bp begin at -214 and -209. The sequence TATAAA, which has been found to precede the 5' end of a number of eukaryotic messages by 30 to 31 bp (29), is present at -170. Similar oligonucleotides are present at -204 (TATATA) and -33 (TATAA). Between -142 and the beginning of the coding region, almost 80% of the bases are A or T, and this region is distinctly divided into two parts. Between -142 and -38,

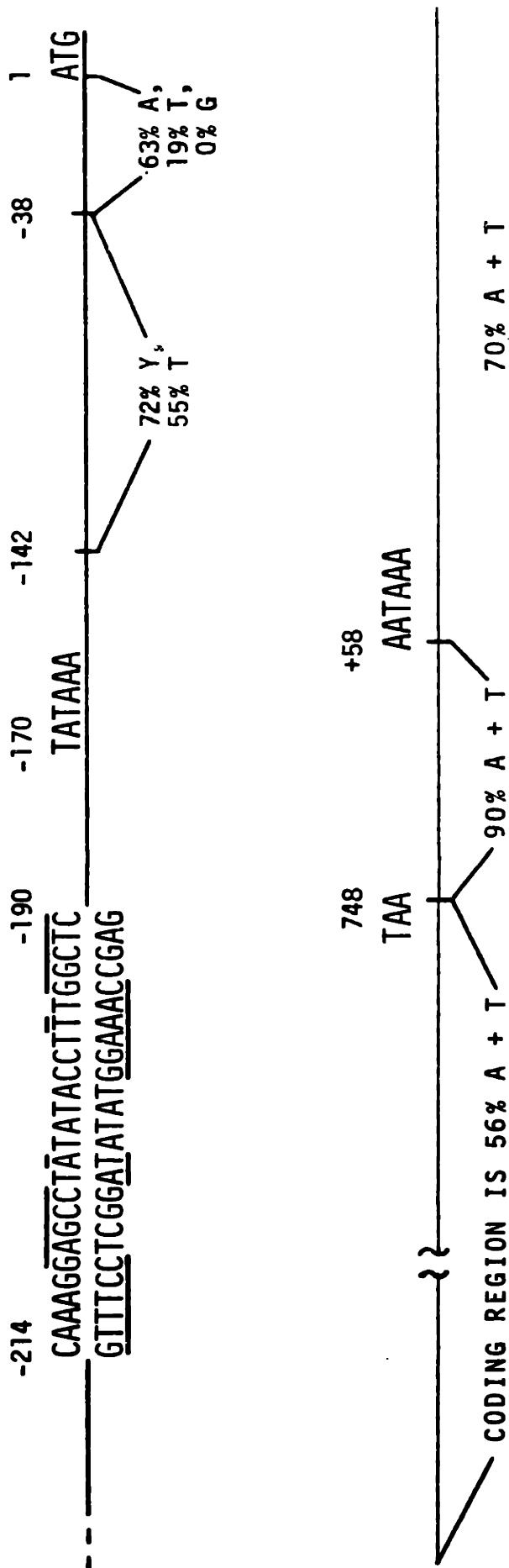


Figure 5. Organization of the yeast triose phosphate isomerase gene. The 5' flanking region contains Hogness box sequences at -204, -202, -170 and -33. The region between -214 and -190 contains two overlapping palindromes (underlined and overlined). The coding region of 747 base pairs contains no introns. The 3' flanking region is extremely A-T rich.

72% of the residues in the message strand are pyrimidines with 55% of the bases being thymine. These are distributed in pyrimidine-rich stretches -- including oligonucleotides of 11 and 12 consecutive pyrimidines -- separated by purine-containing clusters. In contrast, 63% of the residues between -38 and 1 are adenine, and the proportion of thymine in this region drops to 19%, the lowest in the entire sequence. There are no guanine residues between -38 and 1. The ATG which initiates the coding region is the first such sequence in any reading frame downstream from the TATAAA sequence at -170 and the TATATA sequence at -204. The distribution of bases in the yeast triose phosphate isomerase gene is summarized in Figure 6.

Holland and Holland have noted that the 5' flanking sequences of a number of yeast genes show a high degree of homology (10,11). As illustrated in Figure 7, the region between the TATAAA box and the yeast triose phosphate isomerase coding region is reminiscent of the 5' noncoding sequences of the yeast genes for enolase and GAPDH (10,11).

In contrast, no consistent sequence homology has been observed among the 3' flanking sequences of yeast genes, including TPI1. An AATAAA sequence, thought to be a signal for polyadenylation and/or transcription termination in higher eukaryotes, is present at +58 (30,31). The sequence from +1 to +63 is almost 90% A-T, and beyond +63 the A-T content averages about 70%. The 3' flanking region is characterized by alternating purine-rich and pyrimidine-rich oligonucleotide clusters. Presumably because of this clustering pattern and the relative paucity of guanine and cytosine residues, there are few

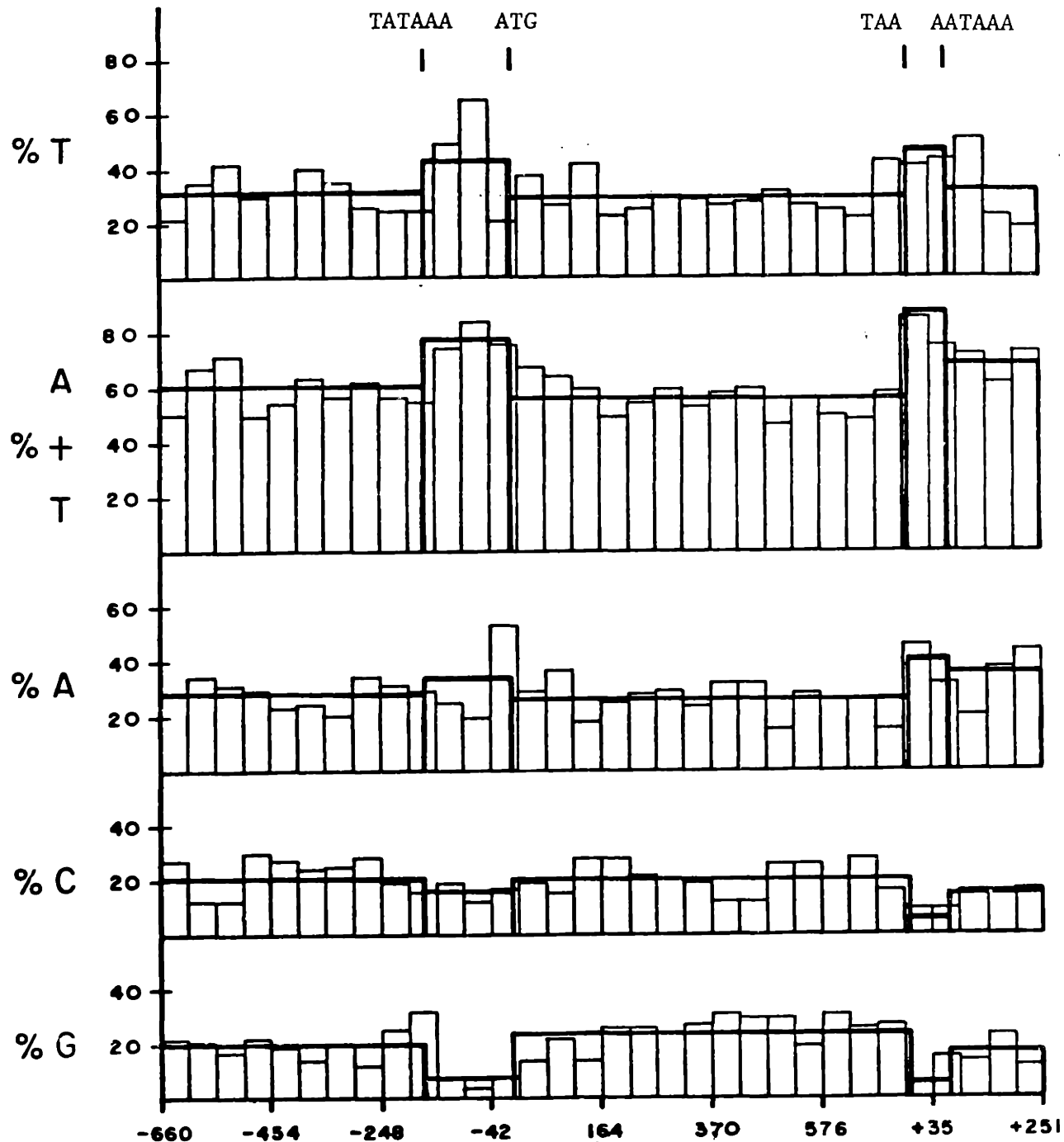


Figure 6. The base composition of the coding strand of TP11 is plotted in blocks of approximately 50 nucleotides. The heavy lines show the composition of several larger divisions of the gene. These divisions are: -660 to the Hogness box at -170, -170 to the start of the coding region, the coding region, +1 to +63, and +63 to the end of the region which was sequenced.

```

-170      -160      -150      -140      -130      -120
TATAAAGGGC AGCATAATTT AGGAGTTTAG TGAACCTGCA ACATTTACTA TTTTCCCTTC
TAAAAACG-- ----TA--TT AT-AGC-AAA CGCAATTGTA A--TTAATTC TTATTTT---
TATAAATAGC -----TTT TGATAATGGC AATATTAATC AAATTTA--- TTTTAC-TTC
TATAAAGAAC GG-----T AGGTATTGAT TGTAATTCTG TAAAT--CTA -----TTTC
TATAAAGGTT GCACTCATTG AAGA---TAG -----

```

```

-110      -100      -90      -80      -70      -60
TTACGTAAAT ATTTTCTTT TTAATTCTAA ATCAATCTTT TTCAATTTT TGTTTGTATT
----G----T ATCTTTTCTT CCC-TTGTC- -TCAATCTTT T---ATTTT ATTTTATTT
TTT-GTAA-C ATTCTCTCTT GTAATTCC-- ----- TT--ATTCT TCTAGCTATT
TT----AAA- -----CTTC T-----TAA A ACTTTTA TA--GTTA-- -GTCT---TT
-----T TTTTTTCTTG TGTGTCT--A TTCA-TTTTA TT--ATT--- -----GT---

```

```

-50      -40      -30      -20      -10      -1 1
CTTTTCTTGC TTAAATCTAT AACTACAAAA AACACATACA TAACTAAAA ATG TPI 1
CTTTTCTTAG TTTCTTTCAT AACCACTAATA CTATAACATA CAA--TAATA ATG pENO 8
          ACCAAGCAA
TTT----- ----- ---CATAAAA AACACACACA TAAATCAA- ATG pENO 46
          CAAGCAACTGCTTATCA
TTTTTTAGTT TAAAAACACC AAGAACAATA AACACACATA AATAAACAAA ATG pGAP491
-----TTGT TTAAATGTT- -----AAAAA AACACACAAA CAAAC CAAA ATG pGAP 63
          CCAAGAACTTAGTTTCAAATTAATTCAT AAAAA

```

Figure 7. Comparison of the 5' flanking regions of yeast genes for triose phosphate isomerase, enolase and glyceraldehyde-3-phosphate dehydrogenase. The sequences are depicted with insertions and deletions to improve the homology. Regions of extensive homology are present from -1 to -25, -90 to -100 and -108 to -130. The sequence numbers refer to TPI1.

restriction sites in this region. Digests with over 30 enzymes, including frequent and infrequent cutters, have revealed an unusually small number of restriction sites in the 700 bp following the end of the coding region (Figure 2). Hind III cuts within an 8 bp palindrome, AAAGCTTT, at +77, and a 12 bp palindrome, ATATAATTATAT, begins at +6.

As in the yeast genes for enolase and GAPDH, there are homologies between the sequences flanking the coding region (10,11). A 10 bp direct repeat begins at positions -52 and +30, and the decamers beginning at -101 and +192 are complementary.

Since the 5' and 3' flanking sequences are grossly similar and since the rate of evolution of the protein is apparently low (3,4), it was of interest to see whether the structure of the yeast triose phosphate isomerase gene might be compatible with transcription and translation in both directions. The sequence of the complementary strand contains two TATAAA sequences, 43 and 91 bp upstream from the end of the protein coding region and contains no AATAAA hexanucleotide. The first ATG downstream from the TATAAA sequences occurs 17 bp into the triose phosphate isomerase coding region, and this is followed by an open reading frame of 174 bp. This is the largest open reading frame initiated by a methionine codon in the entire complementary strand. The codons specified by this sequence are not limited to those which are characteristic of yeast genes. In addition, the bases between the TATAAA sequences and the first ATG codon are not strikingly homologous with the sequence of the analogous region of the triose phosphate isomerase gene.

As described above, we have isolated a number of mutant plasmids which allow the triose phosphate isomerase gene to be expressed in both yeast and E. coli. In order to better understand the nature of these mutations, the sequence of the junction between pBR322 DNA and the yeast insert on the 5' side from the coding region in pFG1 was determined. The sequence of this region is shown in Figure 8 aligned with the corresponding sequence of the intact yeast insert in pTPIC10. The inserts are identical between -65 and 300. Upstream from -65, there is a sequence of 19 bp in pFG1 which is not related to the corresponding sequence in pTPIC10. This sequence ends at a site which could have been created by the ligation of Bam H1 and Sau 3A ends during the construction of the parent plasmid, and presumably, it represents the extreme 5' end of the intact yeast insert. This suggests that the up mutation involved a deletion of about 3400 bp from entirely within the yeast insert.

IV. Discussion

We have isolated and sequenced the gene coding for yeast triose phosphate isomerase and isolated three mutant plasmids which allow the gene to be expressed in both yeast and E. coli. These plasmids contain large deletions or insertions in the yeast sequence and presumably act by circumventing the effect of a region which stops transcription in E. coli. The deletions may remove such a sequence, allowing transcription which begins in the tet^r gene of pBR322 to proceed through the yeast

-170	-160	-150	-140	-130	
TATAAAGGGC	AGCATAATTT	AGGAGTTT	TGAACTTGCA	ACATTTACTA	
ATATTTCCCG	TCGTATTAAA	TCCTCAAATC	ACTTGAACGT	TGTAAATGAT	
CCGCTTTGGC	CGCCGCCTAG	TCCTGCTCGC	TTCGCTACTT	GGAGCCACTA	
GGCGAAACCG	GCGGCGGATC	AGGACGAGCG	AAGCGATGAA	CCTCGGTGAT	
-120	-110	-100	-90	-80	
TTTTCCCTTC	TTACGTAAAT	ATTTTTCTTT	TTAATTCTAA	ATCAATCTTT	
AAAAGGGAAG	AATGCATTTA	TAAAAGAAA	AATTAAGATT	TAGTTAGAAA	
TCGACTACSC	GGTCATGGCG	ACCACACCCG	TCCTGTGGAT	CAAACGAAAA	
AGCTGATGCG	CCAGTACCGC	TGGTGTGGGC	AGGACACCTA	GTTTGCTTTT	
			← pBR322	← Yeast	
				→ Sequence	→
-70	-60	-50	-40	-30	
TTCAATTTTT	TGTTTGTATT	CTTTTCTTGC	TTAAATCTAT	AACTACAAAA	
AAGTAAAAA	ACAAACATAA	GAAAAGAACG	AATTTAGATA	TTGATGTTTT	
<hr/>					
ACCTAATTTTT	TGTTTGTATT	CTTTTCTTGC	TTAAATCTAT	AACTACAAAA	
TGGATAAAA	ACAAACATAA	GAAAAGAACG	AATTTAGATA	TTGATGTTTT	
Deletion boundaries.					
-20	-10	-1 1			
AACACATACA	TAAACTAAAA	ATG			
TTGTGTATGT	ATTTGATTTT				pTPIC10
<hr/>					
AACACATACA	TAAACTAAAA	ATG			
TTGTGTATGT	ATTTGATTTT				pFG 1

Figure 8. Comparison of the intact 5' flanking sequence of TP11 and the sequence of a deletion mutant that allows the protein to be produced in *E. coli*. The DNA sequences of the wild type mutant plasmids are identical between -65 and 300. At -80 in pFG1, the remnants of the Bam HI site of pBR322 are encountered. The presumed deletion boundaries are indicated.

triose phosphate isomerase gene. The insertion may be a transposable element which has hopped into the yeast sequence so as to place a bacterial promoter between the transcription terminator and the yeast triose phosphate isomerase gene.

In either event, the sequence where translation is initiated in E. coli in the messages produced by these mutants has not been clearly identified. In the one mutant plasmid which was sequenced, there is no extensive homology between the bases adjacent to the initiator methionine codon and the 3' end of the E. coli 16S RNA. Translation which begins at the initiator AUG of the tet^r gene would not lead to the production of an active fusion protein, because the two coding regions are not in the same reading frame. The correct frame is blocked by stop codons upstream from the yeast coding region. Further studies on the products of the yeast gene in E. coli are under way to shed light on the mechanism by which the foreign gene is expressed.

The strategy of selecting for expression of a dormant cloned gene in E. coli may provide a general method of obtaining deletion and insertion mutants of eukaryotic promoters. A similar class of deletion mutants which are missing unique restriction sites have been selected by digesting plasmids with the appropriate restriction enzyme prior to transfection in E. coli (10). If the mutant genes can be put back into the cells of their origin, the effects of the genetic lesions can be probed.

One of the deletion mutants of TPI1 produced in vivo was sequenced and found to contain only 65 bp of the 5' flanking region -- approximately 3.4 kb had been deleted upstream from the coding region. Since this deletion reduced by ten-fold the TIM activity in extracts of a transformed tpi⁻ yeast strain, we can conclude that a sequence which is important for the expression of the gene has been deleted. However, until the structure and translational efficiency of the mutant mRNA's have been characterized, the basis for the reduced expression will not be understood.

The intact yeast sequence from which the mutants were produced in vivo does not facilitate the production of triose phosphate isomerase in E. coli, but it is efficiently expressed in yeast. Triose phosphate isomerase is the most abundant protein in SDS polyacrylamide gels of extracts of S. cerevisiae N582-4B which harbors pTPIC10 (G. Kawasaki and D. Fraenkel, unpublished results). Enzyme activity in such a strain is 10 times higher than in wild type strains in which the protein is approximately 2% by weight of the soluble protein in the cell (9). The sequence of bases which is likely to be responsible for such efficient expression has been determined, though the elements necessary for the control and expression of the gene have not been pinpointed.

TPI1 has several features in common with other genes. A so-called Hogness box (TATAAA) is present at -170 and is at the beginning of a region of higher-than-average A-T content. There are few guanines between the TATAAA sequence and the coding region (13/170). As in several other yeast genes, the coding region is initiated by the first

ATG downstream from the Hogness box. This is consistent with a model in which translation is initiated at the first AUG on a messenger RNA molecule (32). The coding region of 747 bases is followed by an extremely A-T rich region containing the sequence AATAAA at +58. This hexanucleotide has been found downstream from a number of eukaryotic genes and may function as a signal for polyadenylation and/or transcription termination (30,31).

As shown in Figure 7, the 5' flanking regions of a number of efficiently expressed yeast glycolytic genes are strikingly similar. The sense strand contains both purine-rich and pyrimidine-rich segments between the TATAAA box and the start of the coding region. A stretch of five to six A's followed by several A-Y doublets and several more A's are located in the 38 nucleotides which immediately precede the coding region. The most consistent homologies are present between the 5' flanking regions of the yeast triose phosphate isomerase gene and the enolase gene or pENO 8 isolated and sequenced by Holland and Holland (11). The homologies suggest that a TATAAA sequence may be present in pENO 8 just upstream from the end of the sequence reported by those workers.

The homologous segments, however, do not have a consistent linear relationship; it is necessary to represent the sequences with insertions and deletions in order to maximize the homologies. This has also been found in the 5' flanking sequences of the human globin genes (33). Since the number of bases between conserved segments varies from sequence to sequence, their functional relationship is likely to be

complex, involving more than the preservation of identical structures in the DNA. In addition, the differences in the 5' flanking sequences which may account for different levels of transcription are not understood. Analysis of mutants will shed light on the functional roles of the conserved segments and the role of the spacings between them.

Similarly, the relationship between the structure and function of the 3' flanking sequence is not well understood. Kroenberg et al exposed the 3' flanking sequences of globin mRNA to exonucleolytic digestion and found that this treatment does not affect the efficiency of translation of the messages in vitro (34). A number of experimentors have found that the sequence of the 3' flanking region is not important in determining the efficiency of transcription of genes in a number of organisms, including yeast (12,13,35). The coordinately regulated yeast genes for enolase, GAPDH and triose phosphate isomerase have 3' flanking sequences which are not extensively homologous (10,11). The 3' flanking sequences of the human α -globin genes and in mouse dihydrofolate reductase mRNA's have also been found to be unexpectedly heterogeneous (36,37). However, the peculiar structure of this region suggests that there is selective pressure on the sequence in vivo. The 3' flanking sequence of the yeast triose phosphate isomerase gene is almost 90% A-T up to the AATAAA sequence at +58 and contains alternating regions of purine-rich and pyrimidine-rich clusters. Possible functions of this region include: 1) providing the recognition sites for polyadenylation and transcription termination, 2) contributing to mRNA stability and 3) contributing to the stability of the coding region sequence over

evolutionary time by affecting the ability of the sequence to recombine with other genomic sequences (30,31,38).

Determination of the sequence of the yeast triose phosphate isomerase gene has allowed us to deduce the amino acid sequence of the enzyme. The protein is coded in a single open reading frame containing no introns. The amino acid sequence is consistent with the composition of the purified protein determined by Hartman and coworkers (7). Figure 9 shows a comparison of the amino acid sequences of TIM's from yeast, rabbit, chicken, coelacanth and B. stearrowthermophilus (1-4). The yeast sequence is 50% homologous with the three vertebrate triose phosphate isomerases and 37% homologous with the B. stearrowthermophilus enzyme. (In comparison, the vertebrate TIM's are 85% homologous and 30-40% homologous with the thermophilic protein.) To maximize the correspondence, it is necessary to represent the yeast sequence with a single amino acid insertion at residue 56.

Though the overall level of homology is modest for proteins with such similar properties (5,6) and structures (see Chapter III), the residues thought to be catalytically important are conserved. This includes the active site base, Glu 165, the candidates for the catalytic electrophiles, His 95 and Lys 13, and the sequences which bind to the phosphate moiety of the substrate, residues 171-173, 209-211 and the C-terminal helix (39). There is no apparent tendency for the differences to be confined to those which would result from single base changes in the codons. Many of the differences in the yeast enzyme represent a conservation of the type of residue, though the sign of

Figure 9. Page 1 of 2.

YEAST	(MET)ALA ARG THR PHE PHE PHE VAL GLY GLY ASN PHE LYS LEU ASN GLY SER LYS CLN SER ILE LYS GLU ILE VAL GLU ARG LEU ASN THR	10	20	30
Rabbit	Ala Pro Ser	Lys	Arg	Lys Asn Leu Gly
Chicken	Ala Pro -	Lys	Lys Arg Lys	Leu Ile Thr Thr
Coel.	Ala Pro -	Lys	Asp	Leu Ile His Thr
<u>B. S.</u>	Met	Lys Pro Ile Ile Ala	Met His Lys Thr Leu Ala Glu Ala Val Asn Phe	Leu Ile Thr Thr Leu Ile Gln Thr Asp Val Lys Gly
YEAST	ALA SER ILE PRO GLU ASN VAL GLU VAL VAL ILE CYS PRO PRO ALA THR TYR LEU ASP TYR SER VAL SER LEU VAL LYS PRO GLN VAL THR	40	50	60
Rabbit	Lys Val	Ala Asp Thr	Cys Ala	Phe Ala Arg Gln Lys Leu Asp
Chicken	Lys Leu Ser	Ala Asp Thr	Cys Gly Ala	Phe Ala Arg Gln Lys Leu Asp
Coel.	Lys Val	Phe Thr Gly	Cys Ala	Phe Ala Arg Leu Lys Asp
<u>B. S.</u>	His Val Pro	Ala Asp Glu Val Asp Ser Val Val Ala	Phe Leu Phe Arg Leu Val Gln Ala Ala Asp Gly Thr Asp Leu Gln Lys	Lys Lys Lys Phe Gly Lys Lys Lys
YEAST	VAL GLY ALA GLN ASN ALA TYR LEU LYS ALA SER GLY ALA PHE THR GLY GLU ASN SER VAL ASP GLN ILE LYS ASP VAL GLY ALA LYS TRP	70	80	90
Rabbit	Ala	Cys	Lys Val Thr Asn	Ile
Chicken	Ala	Cys	Lys Val Pro Lys	Ile
Coel.	Ala	Cys	Lys Val Ser Lys	Ile
<u>B. S.</u>	Ile	Thr Met His Phe Ala Asx Glx	Tyr	Val
YEAST	VAL ILE LEU GLY HIS SER GLU ARG ARG SER TYR PHE HIS GLU ASP ASP LYS PHE ILE ALA ASP LYS THR LYS PHE ALA LEU GLY GLN GLY	100	110	120
Rabbit	Val	His Val	His Val	Gly Val
Chicken		His Val	His Val	Gly Leu
Coel.		His Val	His Val	Gly Leu
<u>B. S.</u>		His Met	Ala Glx Thr	Glx The Val Asx Lys
YEAST	VAL GLY VAL ILE LEU CYS ILE GLY GLU THR LEU LEU GLU LYS LYS ALA GLY LYS THR LEU ASP VAL VAL ARG GLN LEU ASN ALA VAL	130	140	150
Rabbit	Leu	Ala	Lys Asp	Arg Glu
Chicken	Leu	Ala	Lys Asp	Arg Glu
Coel.	Leu	Val Ala	Lys Asp	Arg Glu
<u>B. S.</u>	Leu Ile Pro	Ile	Cys	Leu Ile Pro Ile
YEAST	LEU GLU GLU VAL LYS ASP TRP THR ASN VAL VAL VAL ALA TYR GLU PRO VAL TRP ALA ILE GLY THR GLY LEU ALA ALA THR PRO GLU ASP	160	170	180
Rabbit	Ala Asp Asn	Ser Lys	Ser Lys	Ser Lys
Chicken	Ala Asp Asn	Ser Lys	Ser Lys	Ser Lys
Coel.	Ala Asp Asp	Ser Lys	Ser Lys	Ser Lys
<u>B. S.</u>	Ala, Gly Leu Thr Pro Gln Glu Val Lys Ile Ile Leu			Leu

charged residues is often inverted. Differences in the sign or the number of charged residues occur most frequently in helices in the protein and often occur such that residues at $n \pm 3$ or $n \pm 4$ have the opposite sign. (This arrangement of charged residues may stabilize nascent helices during the process of protein folding (P. Kim, personal communication).)

Since all five triose phosphate isomerases have similar catalytic properties (4-6), the sequences of the yeast and B. stearothermophilus enzymes suggest that the extreme lack of variability among the enzymes from rabbit, chicken, and coelacanth cannot be explained solely on the basis of selection for maximum catalytic efficiency. Other selective pressures at the level of protein structure -- including, perhaps, the interaction of triose phosphate isomerase with other cellular components (S. Mowbray and G.A. Petsko, unpublished results) -- or efficiency of translation, transcription or DNA replication probably underly the observed primary structural homologies. Examination of both strands of the yeast triose phosphate isomerase gene does not support the hypothesis that the restriction of variability may arise from the presence of multiple overlapping genes, though this hypothesis is directly testable by DNA-RNA hybridization.

As described above, the coding sequence contains several inexact repeats up to 21 nucleotides in length. They are in the same reading frame and specify amino acids in several of the helices in the protein. Since the repeats code for similar short amino acid sequences in similar structures, they may have arisen and been maintained as a consequence of

selection for protein primary structure. It is equally possible that they arose through internal gene duplication events during the evolution of the protein. Consequently, their presence, while extremely unlikely in a random sequence of nucleotides, does not provide strong evidence for either hypothesis. However, the sequence of the coding region clearly does not support the hypothesis that the approximate 8-fold rotational symmetry of the triose phosphate isomerase molecule arose through the duplication of entire domains of super-secondary structure during the evolution of the protein. This conclusion is in agreement with those of Phillips et al, who analyzed the homologies within the primary, secondary and tertiary structures of the chicken TIM monomer (28).

The amino acid sequence of yeast triose phosphate isomerase is specified by a restricted set of codons characteristic of highly expressed yeast genes. This set of codons is a subset of those in less efficiently expressed yeast genes and is distinctly different from the sets of codons characteristic of other organisms (40). A similar relationship has been observed between the codons in the genes for abundant and less abundant proteins in E. coli. Ikemura has found that in E. coli there is a strong correlation between the frequency of codon use and the intracellular concentration of tRNA's with the complementary anticodon (41). This supports the idea that codon choice arises from optimization of translational efficiency. (However, studies which have found that codon context can influence the rate of translation suggest that the optimization of translational efficiency is likely to be a complex process, involving more than just codon choice (42).) In order

to shed light on this problem, we have placed a non-homologous coding region, derived from a cDNA copy of the message for human parathyroid hormone, into the yeast triose phosphate isomerase gene in pFG1 (data not shown). This hybrid gene will be used to test the relationship between translational efficiency and codon choice in E. coli and yeast.

By itself, the nucleotide sequence of TPI1 raises more questions than it answers. Knowledge of the sequence will facilitate the rational construction of mutants which may yield information about the structural basis of the regulation of the gene, the structure of the promoter and the catalytic mechanism of the enzyme. In addition, the determination of the amino acid sequence of the protein will aid in the interpretation of mechanistic studies and facilitate the crystallographic refinement of the enzyme structure at high resolution.

REFERENCES

1. Corran, P.H. and Waley, S.G. (1975), *Biochem. J.*, 145, 335-344.
2. Furth, M.J., Milman, J.D., Priddle, J.D. and Offord, R. (1974), *Biochem. J.*, 139, 11-25.
3. Kolb, E., Harris, J.I. and Bridger, J. (1974), *Biochem. J.*, 137, 185-197.
4. Artavanis-Tsakonas, S. and Harris, J.I. (1980), *Eur. J. Biochem.*, 108, 599-611.
5. Noltmann, E.A., in The Enzymes, Vol. VI, P.D. Boyer, ed., Academic Press, New York, 1972, 271.
6. Belasco, J.G. and Knowles, J.R. (1980), *Biochemistry*, 19, 472-477.
7. Alber, T., Hartman, F.C., Johnson, R.M., Petsko, G.A. and Tsernoglou, D. (1981), *J. Biol. Chem.*,
8. Krietsch, W.K.G., Pentchev, P.G., Klingenburg, H., Hofstatter, T. and Bucher, T. (1970), *Eur. J. Biochem.*, 14, 289-300.
9. Clifton, D., Weinstock, S.B. and Fraenkel, D.G. (1980), *Genetics*, 88, 1-11.
10. Holland, J.P. and Holland, M.J. (1980), *J. Biol. Chem.*, 255, 2596-2605.
11. Holland, M.J., Holland, J.P., Thill, G.P. and Jackson, K.A. (1981), *J. Biol. Chem.*, 256, 1385-1395.
12. Faye, G., Leung, D.W., Tatchell, K., Hall, B.D. and Smith, M. (1981), *Proc. Nat. Acad. Sci., USA*, 78, 2258-2262.
13. Guarente, L. and Ptashne, M. (1981), *Proc. Nat. Acad. Sci., USA*, 78, 2199-2203.
14. Broach, J.R., Strathern, J.N. and Hicks, J.B. (1979), *Gene*, 8, 121-133.
15. Nasmyth, K.A. and Reed, S.I. (1980), *Proc. Nat. Acad. Sci. USA*, 77, 2119-2123.
16. Clewell, D.B. and Helsinki, D.R. (1970), *Biochemistry*, 9, 4428-4440.
17. Smith, H.O. and Bernstein, M.L. (1976), *Nuc. Acids Res.*, 3, 2387-2398.

18. Maniatis, T., Jeffrey, A. and van deSande, (1975), *Biochemistry*, 14, 3787-3794.
19. Vogelstein, B., Strayer, D. and Gillespie, D. (1979), *Proc. Nat. Acad. Sci. USA*, 76, 615.
20. Kronenberg, H.M., McDevitt, B.E., Majzoub, J.A., Nathans, J., Sharp, P.A., Potts, J.T. Jr. and Rich, A. (1979), *Proc. Nat. Acad. Sci. USA*, 76, 4981-4985.
21. Maxam, A.M. and Gilbert, W. (1980), *Methods in Enzymology*, 65, 499-560.
22. Staden, R. (1977), *Nuc. Acids Res.*, 4, 4037-4051.
23. Staden, R. (1978), *Nuc. Acids Res.*, 5, 1013-1015.
24. Maitra, P.K. and Lobo, Z. (1971), *J. Biol. Chem.*, 246, 475-488.
25. Smith, M., Leung, D.W., Gillam, S., Astell, C.R., Montgomery, D.L. and Hall, B.D. (1979), *Cell*, 16, 753-761.
26. Williamson, V.M., Bennetze, J., Young, E.T., Nasmyth, K. and Hall, B.D. (1980), *Nature*, 283, 214-216.
27. Montgomery, D.L., Leung, D.W., Smith, M., Shalit, P., Faye, G., and Hall, B.D. (1980), *Proc. Nat. Acad. Sci. USA*, 77, 541-545.
28. Phillips, D.C., Sternberg, M.J.E, Thornton, J.M. and Wilson, I.A. (1978), *J. Mol. Biol.*, 119, 329-351.
29. Benoist, C. and Chambon, P. (1981), *Nature*, 290, 304-310.
30. Proudfoot, N.J. and Brownlee, G.G. (1976), *Nature*, 263, 211-214.
31. Fitzgerald, M. and Shenk, T. (1981), *Cell*, 24, 251-260.
32. Kozak, M. (1980), *Cell*, 22, 7-8.
33. Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., DeRiel, J.K., Forget, B.G. Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980), *Cell*, 21, 653-668.
34. Kronenberg, H.K., Roberts, B.E. and Efstratiadis, A. (1979), *Nuc. Acids Res.*, 6, 153-166.
35. Rose, M., Casadaban, M.J. and Botstein, D. (1981), *Proc. Nat. Acad. Sci. USA*, 78, 2460-2464.
36. Michelson, A.M. and Orkin, S.H. (1980), *Cell*, 22, 371-377.

37. Setzer, D.R., McGrogan, M., Nunberg, J.H. and Schimke, R.T. (1980), *Cell*, 22, 361-370.
38. Slightom, J.L., Blechl, A.E. and Smithies, O. (1980), *Cell*, 21, 627-638.
39. Alber, T., Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., Rivers, P.S. and Wilson, I.A. (1981), *Phil. Trans. Roy. Soc. Lond. B*, 293, 159-171.
40. Grantham, R. Gautier, C. and Gouy, M. (1980), *Nuc. Acids Res.*, 8, 1893-1912.
41. Ikemura, T. (1981), *J. Mol. Biol.*, 146, 1-21.
42. Bossi, L. and Roth, J.R. (1980), *Nature*, 286, 123-127.

CHAPTER III.

Crystal Structure of Yeast Triose Phosphate Isomerase at 3.0 Å Resolution

The first steps in the determination of the structure of yeast TIM were the production of crystals suitable for analysis and the solution of the phase problem to yield an interpretable electron density map.

I. Crystals

The first of these problems was solved by Drs. Petsko, Tsernoglou and Hartman (1). Extremely good crystals of the enzyme were produced in the presence of 10% saturated ammonium sulfate using polyethylene glycol (PEG) 4000 as the precipitant (Figure 1a). The crystals are grown by inducing a phase separation with ammonium sulfate and 60% PEG and then diluting the solution approximately threefold until only a few droplets of the protein phase remain. The crystals appear to grow from these droplets due to competition between the protein and PEG for water of hydration (1) (Figure 1b).

The crystals diffract to at least 1.5 Å resolution and can be stabilized in solutions of 40% PEG 4000 in the absence of sulfate, a



Figure 1a) Crystals of yeast triose phosphate isomerase grown from polyethylene glycol. The larger crystal is 1.5 mm long.

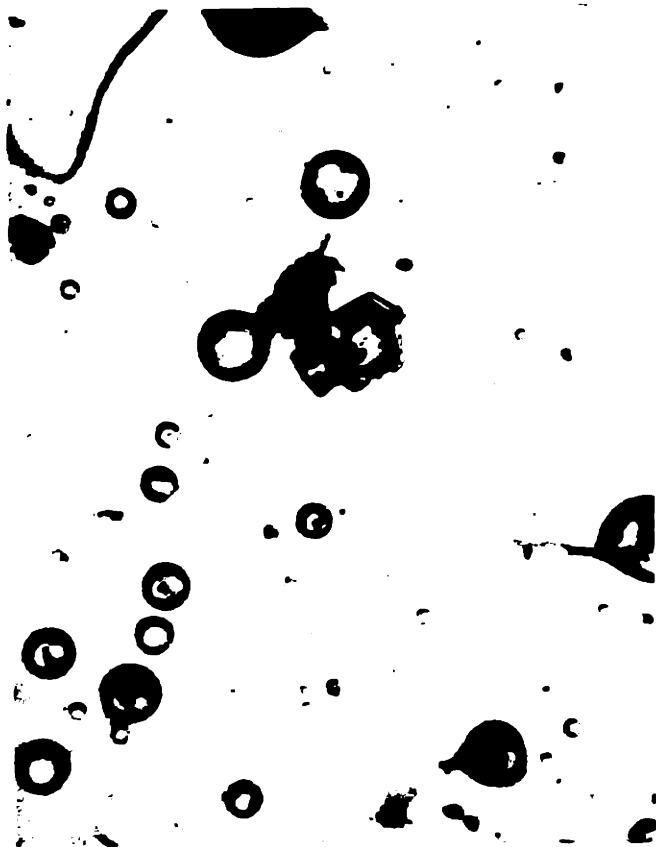


Figure 1b) The crystals grow after the protein is "oiled out". Nucleation apparently occurs in the oil, or protein, phase. The dense material present inside the droplets may be precipitated protein.

competitive inhibitor of the enzyme. They have the symmetry of space group $P2_1$ with one dimeric molecule in the asymmetric unit. The crystallization conditions, space groups and unit cell dimensions of crystals of triose phosphate isomerases from a number of organisms are summarized in Table 1 (1-4).

II. The Phase Problem

Given crystals of such high quality, a strategy for determining the protein structure needed to be chosen. Crystallographic experiments are directed toward evaluating the electron density, ρ , at each point in the unit cell, given in fractional coordinates x , y and z , using the expression:

$$\rho(x,y,z) = \frac{1}{V} \sum_{hkl} F(hkl) \exp(i\alpha(hkl)) \exp(-2\pi(hx+ky+lz)) \quad (1)$$

where V is the volume of the unit cell, h , k and l are the Miller indices of a given reflection, $F(hkl)$ is the magnitude of the scattered wave and $\alpha(hkl)$ is its phase. X-ray detectors are sensitive only to the energy of each scattered wave, proportional to $F(hkl)^2$. As a result, information about the phase of the reflections is lost in the process of measuring the data. To evaluate the electron density function, however, the phase of each reflection must be estimated. This is the "phase problem" in crystallography, and it arises because measurements of the energy of the scattered waves do not yield direct information about their phases.

Table 1. Crystallographic parameters for triose phosphate isomerase crystals (1).

Source	Space Group	Molecules/ Asymmetric Unit	Cell Constants			
			a	b	c	β
Rabbit muscle	$P2_1$	1	36.6	76.4	89.6	100.45°
Rabbit muscle	$P6_122$ or $P6_522$	1/2	99.4	99.4	105.8	90.0°
Chicken muscle	$P2_12_12_1$	1	106.0	74.8	61.7	90.0°
Yeast	$P2_12_12_1$	1	161.2	62.2	47.3	90.0°
Yeast	$P2_1$	1	61.3	98.4	49.7	90.9°
Yeast + DHAP	$P2_1$	1	61.7	82.4	46.1	92.6°
(soaked)*						
Yeast + PGA	$P2_1$	1	74.2	82.9	37.6	102.0°
(cocrySTALLIZED)*						

(The crystal forms marked with an asterisk are described in the next chapter.)

Two methods can be used to determine the phases of the reflections from a crystal of a molecule as large as triose phosphate isomerase. Since the structure of chicken muscle TIM is known (5), one possibility is to use the molecular replacement method to calculate initial phases for the yeast TIM data (6). This approach involves finding the orientation of a model structure (in this case, chicken TIM) which would produce x-ray diffraction data (formally, a Patterson synthesis) most like the actual data from the unknown structure. This method was ultimately rejected for two reasons. 1) It can be very difficult to remove a bias toward the model structure from the phases for the unknown structure. Images of the model structure can feed back throughout the analysis, and true differences will be difficult to detect. As a result, it is hard to show that the two structures are truly independent, and comparisons between them can not be made with confidence. 2) The approximate eight-fold rotational symmetry of the chicken TIM monomer (7) could give rise to many orientations of this model which would approximate the diffraction data obtained from the yeast TIM crystals. Finding the correct orientation promised to be a lengthy trial-and-error process. (Indeed, after the yeast TIM structure had been solved, it was found that the orientation of the chicken enzyme which best approximated the yeast TIM Patterson to 6 Å resolution did not, in fact, correspond to the correct position of the yeast TIM molecule (G.A. Petsko, unpublished results).

III. Isomorphous Replacement

Instead the structure was solved by the method of isomorphous replacement (8). Since no new techniques were used, the method will be described only briefly. It is treated comprehensively in a number of reviews, monographs and books (9-12). For the technical details of the structure determination, the interested reader should consult the references to the crystallographic literature which are cited below.

The isomorphous replacement method has the advantage that the resulting phases have no inherent bias toward any known structure. However, calculating phases depends on producing one or more derivatives of the enzyme crystals which contain heavy atoms bound specifically to the protein molecules. Ideally, the only difference between the native and derivative crystals should be the addition of the heavy atom; no change in the conformation of the protein or in the dimensions of the unit cell should occur. That is, the derivative and native structures should be isomorphous. In practice, this situation can be quite difficult to achieve.

A. Phase Information from Isomorphous Differences

The changes in the diffraction pattern which result from the binding of heavy atoms to the protein can be exploited to produce a three dimensional map of the vectors between the metal atoms in the unit cell. Such a map is calculated using the difference Patterson synthesis:

$$P(u,v,w) = \frac{1}{V} \sum_{hkl} |F_{PH}(hkl) - F_P(hkl)|^2 \exp(2\pi i(hu + kv + lw)) \quad (2)$$

where u, v and w are the fractional coordinates along the unit cell edges, $F_{PH}(hkl)$ is the amplitude of the wave scattered from the derivative crystal and $F_P(hkl)$ is the analogous amplitude of the wave scattered from the native crystal. The Patterson function represents the interatomic vectors weighted by the product of the number of electrons ($z_i \cdot z_j$) for each atom pair (13). It requires no information about the phase of each reflection.

The difference Patterson synthesis is solved to yield a set of heavy atom positions which accounts for all the major peaks on the map. These heavy atom positions are then refined using a least squares method. The degree to which they account for the observed changes in the diffracted intensities is evaluated to check the validity of the model. Problems can arise from a number of sources, including nonisomorphism of the derivative and native protein crystals, poor (imprecise) data, incorrect scaling of the data sets, and complex patterns of heavy atom substitution which are difficult to model. In addition, the assumption is made that $|F_{PH} - F_P|$ is a good approximation of F_H , the amplitude of the scattering due to the heavy atoms alone. This condition is rarely satisfied, and this leads to spurious peaks and ripples in the difference Patterson map.

Once the heavy atom positions have been estimated, the contribution to the diffraction pattern of the derivative crystals which arises from the heavy atoms alone, $\overline{F_H}$, can be calculated

$$\overline{F_H(hkl)} = \sum_j f_j \exp(2\pi i(hx+ky+lz)) \quad (3)$$

where f_j is the atomic scattering factor for the j th heavy atom. For each reflection, expression 3 is summed over all the heavy atoms in the unit cell to give both the amplitude and the phase of $\overline{F_H}$.

The amplitude, $F(hkl)$, and the phase, $\alpha(hkl)$, of a scattered wave can be represented as a vector, $\overline{F(hkl)}$ (Figure 2a). This vector is called a structure factor. When $F_p(hkl)$ and $F_{pH}(hkl)$ are much larger than $F_H(hkl)$, the structure factor of the isomorphous derivative crystal, $\overline{F_{pH}(hkl)}$, can be represented as the vector sum of the scattering from the protein atoms and the scattering from the heavy atoms:

$$\overline{F_{pH}(hkl)} = \overline{F_p(hkl)} + \overline{F_H(hkl)} \quad (4)$$

This sum is illustrated in Figure 2b. In practice, $\overline{F_H(hkl)}$ is calculated from the model of the heavy atom positions using expression 3, but only the magnitudes of the scattered waves, $F_{pH}(hkl)$ and $F_p(hkl)$ are derived directly from the X-ray data. For any given reflection, these three quantities will be consistent with only two values for the phase of $\overline{F_p(hkl)}$. This is illustrated graphically in Figure 2c (14). The ambiguity in the phase determination can be vitiated by using a second heavy atom derivative (Figure 2d) or by exploiting anomalous scattering effects, or both.

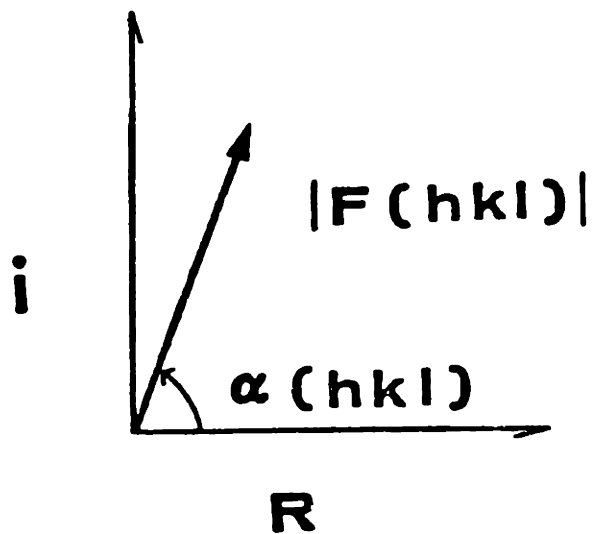


Figure 2a) The structure factor can be represented as a vector in a complex plane.

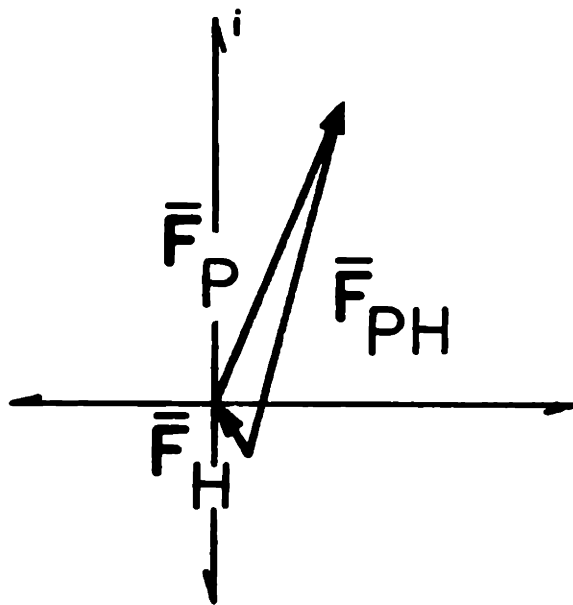


Figure 2b) The vector sum of \overline{F}_P and \overline{F}_H .

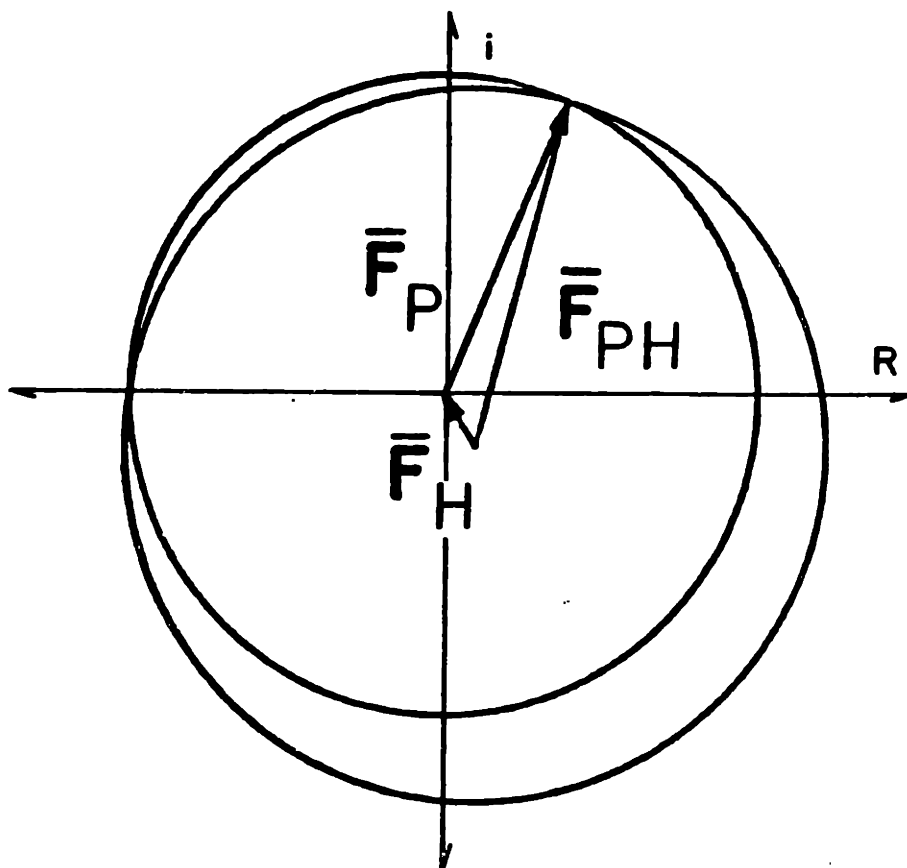


Figure 2c) Phase circles for a single isomorphous derivative (Harker construction). \overline{F}_H is drawn as a vector in the complex plane, and a circle of radius F_P is drawn with its center at the head of \overline{F}_H . Since the phase of \overline{F}_P is unknown, the circle represents \overline{F}_P with all possible phase angles. A second circle of radius F_{PH} (representing all possible orientations of \overline{F}_{PH}) is drawn centered on the beginning of the tail of \overline{F}_H . When \overline{F}_P and \overline{F}_H are not colinear, the circles will intersect in two points. The points of intersection are symmetrical about \overline{F}_H and give values for α_P and α_{PH} which are consistent with Equation 4:

$$\overline{F}_{PH} = \overline{F}_P + \overline{F}_H .$$

In the special case where \overline{F}_P and \overline{F}_H are colinear, the circle will touch at only one point, and the phase determination is unambiguous.

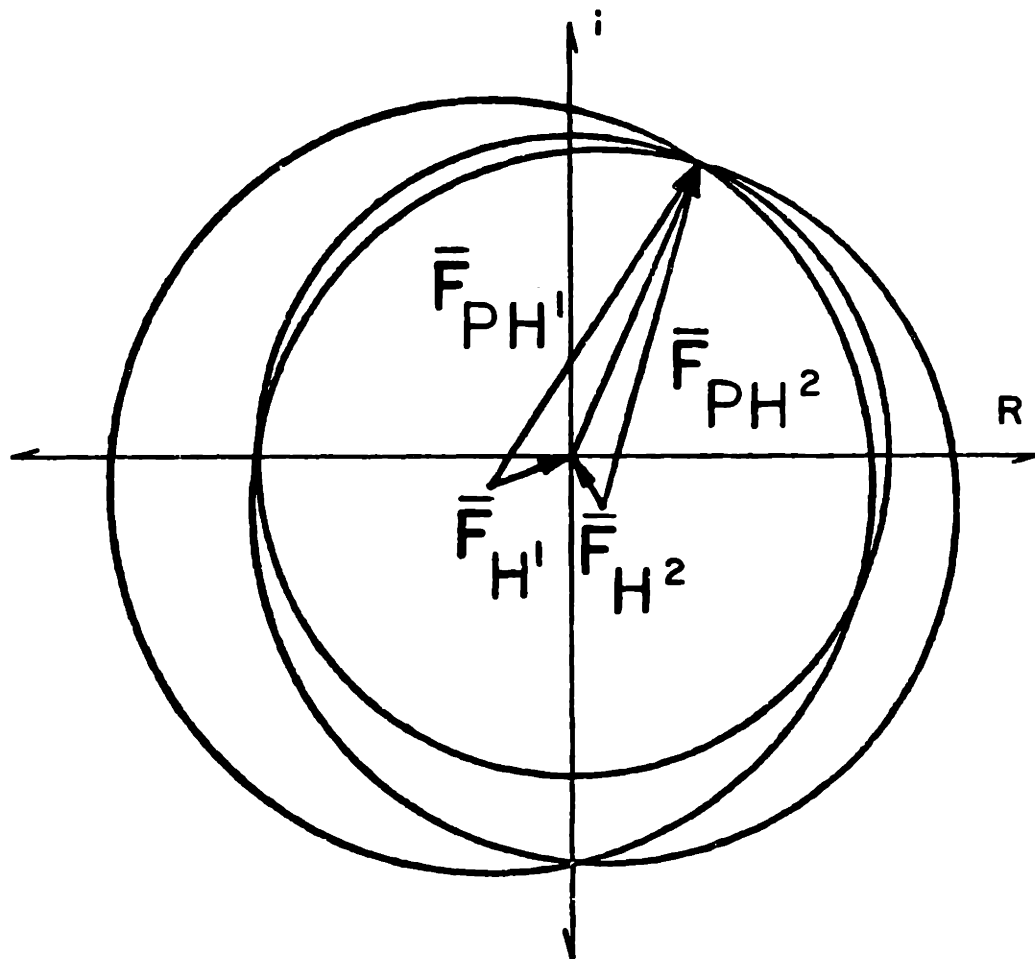


Figure 2d) Phase circle for two isomorphous derivatives, each constructed as described in Figure 2c. In principle, this is sufficient to resolve the phase ambiguity.

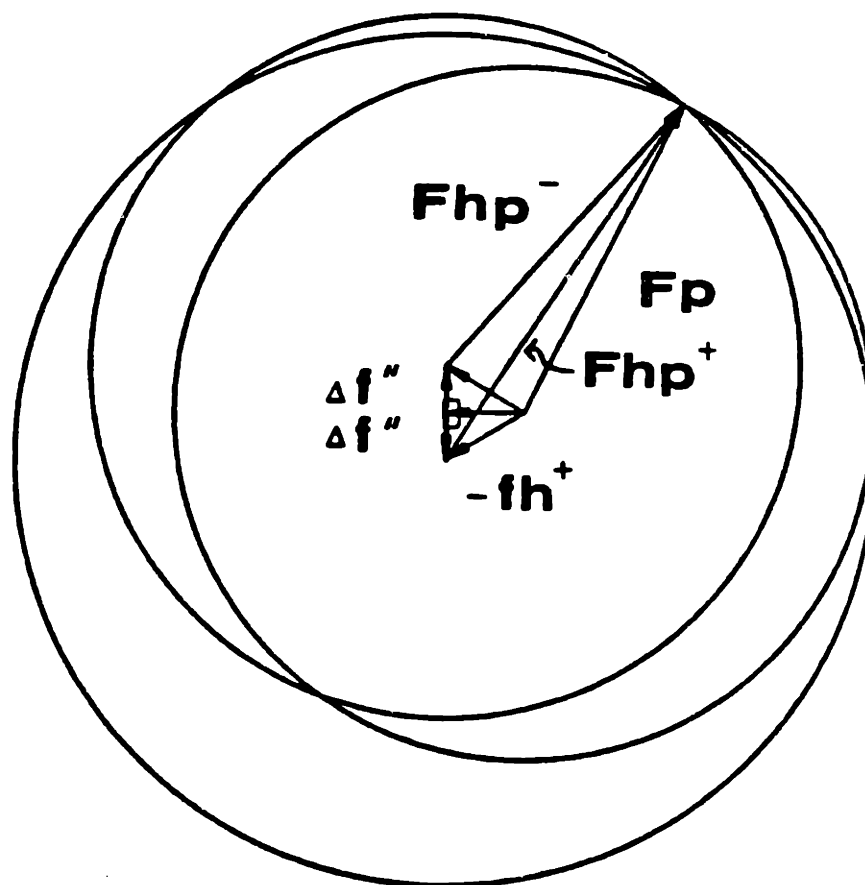


Figure 3. Phase circle for a single derivative with significant anomalous scattering. The imaginary component ($\Delta f''$) of the heavy atom structure factor is 90° out of phase with the real component. This results in the breakdown of Friedel's law. The difference between $F_{PH}(hkl)$ and $F_{PH}(-h-k-l)$ can yield information about the ϕ_{PH} phases and the pattern of heavy atom binding.

B. Anomalous Scattering

Anomalous scattering is caused by the absorption of the incident radiation by atoms in the crystal. Typically, the heavy atoms in the derivative crystal exhibit the largest anomalous scattering effects. These are manifested by the breakdown of Friedel's law, which predicts that $F(hkl)$ and $F(-h-k-l)$ are equal. Anomalous scattering effects can be used to provide a better estimate of the magnitude of $F_H(hkl)$ and to provide independent information about the phases of the structure factors of the native protein crystal, $\overline{F_P(hkl)}$.

The phase circles for an anomalous scatterer are shown in Figure 3. This information is independent of, and complementary to, the phase information derived from the isomorphous differences. Jensen et al have shown that, in favorable cases, a protein structure can be solved by exploiting the isomorphous and anomalous differences of a single heavy atom derivative (15). In general, however, the small magnitude of anomalous scattering effects makes them hard to measure accurately, and several different derivatives are needed to produce phases which can be used to calculate an interpretable electron density map.

The accuracy of the phase determination depends critically on the accuracy of the heavy atom structure factors, $\overline{F_H(hkl)}$, and this, in turn is a reflection of the fidelity of the model of heavy atom binding (Equation 3). The heavy atom parameters are tested and adjusted by their ability to predict $F_H(hkl)$ which, as described above, is only rarely well approximated by the isomorphous difference.

$|F_{PH}(hkl) - F_P(hkl)|$. Thus, the errors in the estimate of $F_H(hkl)$ are propagated through the entire phase calculation.

The isomorphous and anomalous differences can be combined to give a more accurate estimate of $F_H(hkl)$ than that provided by the isomorphous differences alone (16). This quantity is called the lower estimate of F_H , or F_{HLE} :

$$F_{HLE}^2 = F_P^2 + F_{PH}^2 - 2F_P F_{PH} (1 - (k\Delta I / 2F_P)^2) \quad (5)$$

where k is the ratio of the real and imaginary components of F_H (Figure 3) and ΔI is the anomalous difference, $|F_{PH}(hkl) - F_{PH}(-h-k-l)|$. In practice, k is determined empirically to compensate for the complex environment of the heavy atoms and errors in the structure factor amplitudes:

$$k_{emp} = \frac{2 \sum \sum \sum |F_{PH} - F_P|}{\sum \sum \sum |\Delta I|} \quad (6).$$

F_{HLE} is a reasonable approximation of F_H , usually introducing errors which are smaller than the errors in the measurements of $F_P(hkl)$ and $F_{PH}(hkl)$. However, Equation 5 breaks down when $F_P(hkl)$ is small, when the angle between $\overline{F_P(hkl)}$ and $\overline{F_{PH}(hkl)}$ is obtuse, or when the anomalous difference is measured extremely poorly. Reflections which may satisfy any of these conditions are excluded from the analysis. Values of F_{HLE} are used to improve the model of heavy atom binding, thus improving the values of $\overline{F_H(hkl)}$ calculated using Equation 3.

C. The Treatment of Errors in Phase Determination

In theory, all the phase circles shown in Figures 2d and 3 should intersect in one point, and the ambiguity in the phase of the reflection should be completely resolved. In practice, errors from a number of sources frustrate the exact determination of phases. The amplitudes $F_P(hkl)$ and $F_{PH}(hkl)$ are inexact and the calculated structure factors of the heavy atoms, $\overline{F_H(hkl)}$, reflect the imperfections in the difference Patterson synthesis and the heavy atom model which have been discussed above. Problems can arise from the inaccuracy of Equation 4 due to non-isomorphism and centrosymmetry in the heavy atom constellations. In addition, in cases where $F_H(hkl)$ is small, small errors in the structure factor amplitudes can lead to large errors in $\alpha_P(hkl)$. In such cases, the phases are essentially indeterminate. A central problem of phase determination, then, is to assess and minimize the effects of the errors on the calculated electron density, $\rho(x,y,z)$ (Equation 1).

The standard methods for the treatment of errors in the isomorphous replacement method were developed by Blow and Crick (17). Their approach was extended to treat the errors in the anomalous differences by North (18) and Matthews (19). In general, the probability distribution for the phase of a given reflection is calculated and the centroid of this distribution is chosen as the phase. This so-called "best phase" has been shown to produce better electron density maps than the phase with the highest probability (20).

Blow and Crick suggested the simplifying assumptions that F_H and F_P are known accurately and that all errors lie in the measurement of F_{PH} . For any arbitrary phase angle, α_p , the triangle formed by F_P , F_H and F_{PH} in a complex plane will not close exactly (Figure 4). If F_{PH}^{calc} is the vector sum of F_H and $F_P \exp(i\alpha)$, then the lack of closure of the triangle is given by:

$$F_{PH}^{calc} - F_{PH} = \epsilon \quad (7)$$

If the distribution of errors is assumed to be Gaussian, the probability, $P(\alpha)$, that a given phase is correct is:

$$P(\alpha) = N \exp(-\epsilon^2/2E^2) \quad (8)$$

where N is a normalizing factor such that the sum of all probabilities is unity, and E is a statistical estimate of the total error in the derivative measurements. E is usually calculated in one of two different ways:

$$E^2 = (\sum \sum \sum W |F_{HLE} - F_H^{calc}|_2^2) / n \quad (9)$$

where W is a weighting factor and n is the number of reflections in the triple sum, and:

$$E^2 = \sum \sum \sum \epsilon^2 / n \quad (10).$$

The phase information (probability distributions) from j derivatives can be combined using the expression:

$$P(\alpha) = \prod_j P_j(\alpha) = N \exp(-(\sum \epsilon_j^2 / 2E_j^2)) \quad (11).$$

Blow and Crick defined the "best phase" as that which produces a Fourier which has the minimum mean square difference from the Fourier calculated with the true F_P s (17). Dickerson, Kendrew and Strandberg

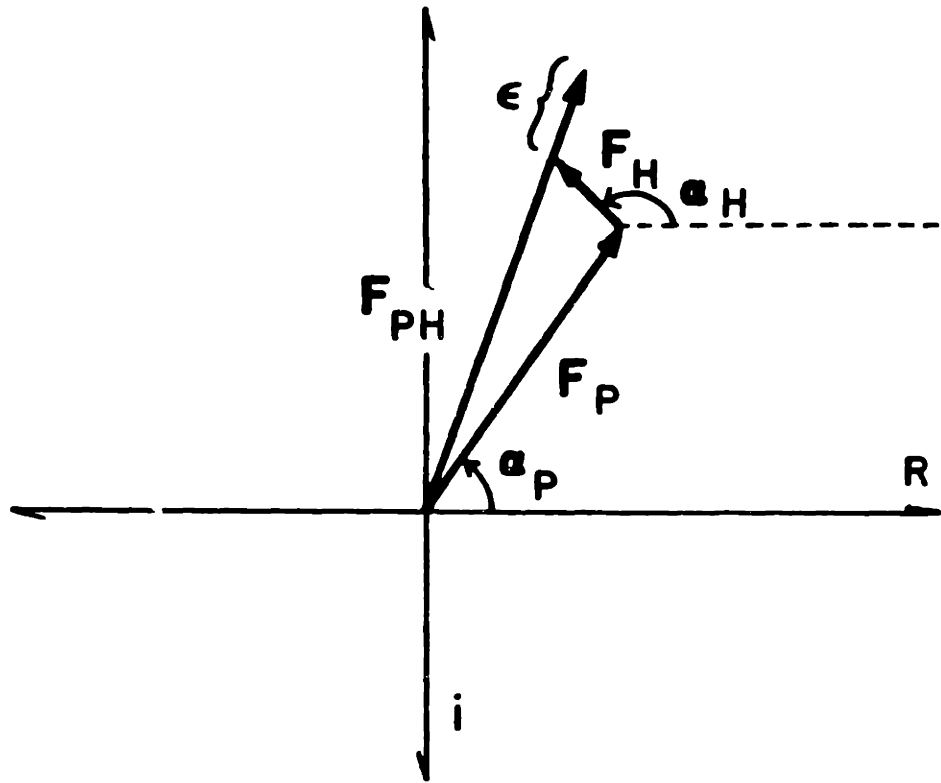


Figure 4. Calculation of the lack-of-closure error, ϵ .

showed that, in polar coordinates, the vector to the centroid of the phase probability distribution has a magnitude of $mF_p(hkl)$ and an angle of α_{pbest} (21). m is called the figure of merit, and it is the mean value of the cosine of the error in phase angle. It varies from zero for randomly determined phases to one for unambiguously determined phases. Thus, it is a measure of the reliability of the phase determination. It is used as a weighting factor for $F_p(hkl)$ in the calculation of the electron density, ensuring that reflections with well determined phases will contribute relatively more to the electron density map than reflections for which the phase information is inconsistent.

Once an electron density map is obtained, using weighted F_p 's in Equation 1, the protein structure is discerned by locating or "tracing" the polypeptide chain in the electron density. This process depends for its success on the quality of the electron density map and therefore on the accuracy of the phases and structure factor amplitudes. The tracing is validated by establishing a correspondence between the known amino acid sequence of the protein and the electron density of the putative side chains.

The crystal structure of yeast triose phosphate isomerase was solved at 3.0 Å resolution by the method of multiple isomorphous replacement. The isomorphous and anomalous differences of two heavy atom derivatives were used to calculate phases. This chapter describes the solution of the x-ray structure.

IV. Materials and Methods

A. Crystallization

The yeast triose phosphate isomerase used in this study was the generous gift of Dr. Fred Hartman of Oak Ridge National Laboratory. It was purified from Saccharomyces cerevisiae Hansen and stored at 4°C as a precipitate in 95% saturated ammonium sulfate. Crystals of the free enzyme were grown from 50mM Tris-HCl, 1mM EDTA, 1mM 2-mercaptoethanol 10% saturated ammonium sulfate at pH 7.1-7.4 using PEG 4000 as the precipitant. The starting protein concentration was 20 mg/ml, but this was diluted to 2-3 mg/ml in the final crystallization solution. As described above, crystal growth followed phase separation. The crystallization experiments were carried out at room temperature. Prior to data collection, the crystals were stabilized in sulfate-free mother liquor containing 40% PEG 4000 at 4°C.

PEG 4000 was from Baker Chemical Co., and Tris base and ammonium sulfate were electrophoresis grade reagents from Biorad. Efforts to produce crystals using reagents of lower purity were not successful.

B. Heavy Atom Derivatives

To diffuse heavy atoms into the crystal lattice, crystals were transferred to sulfate-free mother liquor at 4°C and then to identical mother liquor containing the heavy atom compound. The compounds that were used were from the collections of heavy metal reagents available in

detector of the diffractometer were alligned before measuring each data set.

Diffractometer data were collected by the ω -step scan method of Wyckoff et al (24). Peak intensities were estimated by integrating over 40% of the width of the reflections. Using a take-off angle of 5° , the counts of seven of eleven steps of 0.03° in ω were summed. On average, the full peak width was 0.5° in ω . Intensities below 2σ were not included in the data sets.

D. Data reduction

Backgrounds were estimated using a radial bakckground curve which was dependent on ϕ and 2θ . Background curves were measured for each crystal at the ϕ value of minimum transmission (ϕ_{\min}) and at $\phi_{\min} + 90^\circ$. Absorbtion corrections were made by the semi-empirical method of North, Phillips and Mathews (25). A single reflection near $\chi = 270^\circ$ was measured at 24 values of ϕ to construct the empirical absorbtion curve. Five reflections in different regions of reciprocal space were measured periodically to monitor radiation damage. In general, radiation damage was linear as a function of time, monotonic, and independent of resolution. Both linear and non-linear correction functions were calculated using a least squares procedure. In all cases, the linear correction provided a better fit to the observed data. Lorentz and polarization corrections were applied (26).

our laboratory and in the Laboratory of Molecular Biophysics at the University of Oxford. Dimercuric acetate was a gift from Dr. B.W. Matthews. The soaks were usually carried out in the dark at 4°C to inhibit any photochemical reactions of the heavy atom compounds.

Crystals were mounted for precession photography in sealed glass capillary tubes containing a slug of mother liquor. Precession photographs were taken on Buerger precession cameras (Charles Supper Co. and Enraf-Nonius, Inc.) using nickel-filtered copper K_α x-rays. X-ray sources included an Elliot rotating anode and an Enraf-Nonius sealed tube X-ray generator.

C. Data Collection

Crystals were mounted for data collection in sealed glass capillary tubes flanked by slugs of mother liquor and mineral oil to insure constant hydration at low temperature. Data were collected at -5 to +5°C to reduce the rate of radiation damage (22). Native data with a minimum interplanar spacing of 2.0 Å were collected on a single crystal using the multiwire area detector at the University of California at San Diego (23). Using two crystals, native data were also collected to 1.9 Å resolution with a single-counter Nicolet P3 diffractometer equipped with a low temperature device. Diffractometer measurements were made on heavy atom derivatives to 3.0 Å resolution, and equivalent reflections were measured as h,k,l and h,-k,l. Nickel filtered copper K_α radiation was used for all intensity measurements. The telescope, source and

To correct systematic errors in Friedel equivalent reflections, scale factors were applied to the more highly absorbed equivalent in blocks of reciprocal space so that $F_{PH}(hkl)/F_{PH}(-h-k-l)$ approached unity.

Data sets were analyzed and scaled by reiterative application of the local scaling procedure of Matthews (27). Analysis and scaling were performed, as needed, as a function of $h, k, l, F(hkl), \sin\theta/\lambda$ and in blocks of reciprocal space. The data were never scaled as a function of $F(hkl)$, because reflections which are weak in the native data should, on average, be stronger in the derivative data. This situation arises because changes which decrease the intensity of weak reflections will tend to cause them to fall below 2σ and be dropped from the derivative data set. Consequently, these reflections are not used in the calculation of the scale factor. As a result, the weak reflections which are present in the derivative data will be, on average, more intense than their native counterparts.

The method of Patterson origin comparison was used to improve the scaling of the derivative to the native data.

E. Heavy Atom Parameters

Patterson maps were calculated using the isomorphous, anomalous and "combined" differences as coefficients. The F_{HLE} method of Matthews (16) and Dodson and Vijayan (28) was used to integrate the information from the isomorphous and anomalous differences. F_{HLE}^2 Patterson maps

were solved by inspection, and the putative heavy atom positions were refined by minimizing the overall difference between F_{HLE} and $F_H(hkl)$.

The absolute configuration of the heavy atoms in the Baker's (mercury meso-2,3-dimethoxytetramethylene bis-acetate) and platinum nitrite derivatives was checked by calculating difference Fourier maps with the sign of the anomalous contributions reversed (16). Maps with the correct heavy atom enantiomorph were uniformly cleaner and had higher peaks in the expected positions. A difference Fourier calculated with coefficients $F_{Pt(NO_2)_4}(hkl) - F_P(hkl)$ and phase angles derived from the Baker's derivative alone was used to establish the relative y coordinates of the heavy atoms and to adjust the platinum positions for further refinement. Using both the Baker's and $Pt(NO_2)_4$ derivatives, several cycles of phase refinement (21) were carried out to improve the heavy atom model.

(These phases were used to calculate a difference Fourier map using the structure factor amplitudes $F_{DMA}(hkl) - F_P(hkl)$. This map showed two mercury peaks at the Baker's/EMP site and indicated that DMA could be used to produce a third heavy atom derivative. However, the DMA data were not used in the phase calculation, because, by the end of data collection, radiation damage had reduced the intensities of the check reflections to less than 40% of their initial values.)

F. Calculation of Phases

Phases were calculated by the methods of Blow and Crick (17) and Matthews (19). Centric isomorphous E values divided by root 2 were used, and the E's for the anomalous differences were set at approximately 1/3 of those for the isomorphous differences.

G. Electron Density Maps

Electron density maps were calculated with the fast Fourier program of R. Fischer. They were contoured empirically to establish continuous chain connections with a minimum of ambiguity. This was accomplished by setting the lowest contour level at about $0.25 \text{ e}^-/\text{\AA}^3$ and contouring in steps of $0.1 \text{ e}^-/\text{\AA}^3$. Fourier maps were calculated using phases based on the Baker's derivative alone as well as phases based on both Baker's and platinum nitrite derivatives. The latter map was easier to interpret, though the former clearly showed many features of the molecule. The electron density maps were improved by using native and derivative data sets which were all collected on the P3 diffractometer with the crystals mounted in the same orientation and using identical algorithms for peak integration and data reduction. Inclusion of the data from the ethyl mercuric phosphate derivative (EMP) did not improve the electron density.

The chain was traced in a minimap by the method of Hill et al (29). Arbitrarily choosing an obvious α -carbon, the electron density at a radius of 3.6\AA was searched for a branch point which could represent the

next α -carbon in the polypeptide chain. This procedure was reiterated and dots were placed on the α -carbon positions. When the direction of the chain was ambiguous, the search radius was increased. This often resolved the course of the backbone, and residue $n+1$ could be identified after locating residues n and $n+2$. When the chain ended or became impossible to follow, the tracing direction was reversed. Two factors were essential for the ultimate success of this procedure. 1) A single dimer was isolated in the minimap; trying to follow the fold of the backbone across the edges of the map proved too difficult. 2) The tracing was carried out systematically and sequentially. Skipping from one easily identified feature to the next was counterproductive. Among other things, it led to a confusion of the subunits and made it impossible to consistently connect the features which had been traced. On the other hand, the sequential method assured that all the dots on the map were in the same subunit.

After 210 of an estimated 248 α -carbons in one subunit had been located, the fold of the chicken TIM monomer could be discerned in the yeast TIM electron density map. The coordinates of 10 to 20 α -carbons were estimated and rotation and translation matrices were calculated which transformed the unrefined chicken TIM coordinates (30) into the yeast TIM unit cell. This procedure was repeated until the root mean square (r.m.s.) difference between the α -carbon coordinate sets was not improved by adding more atoms.

The transformed coordinates were used to complete the tracing of the first monomer and to follow the polypeptide chain of the second from

beginning to end. The latter procedure allowed the yeast and chicken enzymes to be compared qualitatively.

V. Results

A. Heavy Atom Derivatives

The conditions of the heavy atom soaking experiments are listed in Table 2. In general, precession photographs were compared to assess the results of the soaks. Figure 5 shows the analogous zones from crystals of the native protein and the derivatives made with Baker's dimercurial and potassium platinochloride. Dimercuric acetate (DMA) and uranyl nitrate soaks were assessed by taking three dimensional data and calculating difference Patterson or difference Fourier maps. DMA produced an independent derivative but the data were not used in the phase calculation, because severe radiation damage undermined their reliability.

B. Data Collection

Data collection statistics are given in Table 3. The agreement residual between equivalent reflections in the 2.0 Å native data set collected using the multiwire area detector was a respectable 6% on F's. The data set collected on the crystal soaked in uranyl nitrate provides a measure of the precision of the diffractometer data. Since the heavy atom did not bind specifically in the crystals, these data may be

Table 2. Preparation of heavy atom derivatives. All soaking experiments were carried out in the dark at 4° C in sulfate-free mother liquor.

Reagent	Concentration	Time of Soak	Number of Heavy Atoms Bound/ Asymmetric Unit
Ethyl mercuric phosphate	3 mM	44 & 89 hours	2
Mercury meso-2,3-dimethoxytetramethylene bis-acetate, (Baker's dimercurial)	1.8 mM	14 - 22 days	2
Dimercuric acetate	2 mM	2 & 7 days	4
$K_2Pt(NO_2)_4$	5 mM	31 days	4
K_2PtCl_4	4.5 & 7.3 mM	4 days	Diffraction pattern showed changes, 2-4 sites
p-toluenyl mercuric chloride	4.3 mM	2 days	0
$PdCl_2$	2.5 mM	4 days	?, crystal turned orange and disordered
$UO_2(NO_3)_2$	0.5 mM	65 hours	0
p-hydroxy-mercuri-benzoate	1 mM	6 & 18 days	0

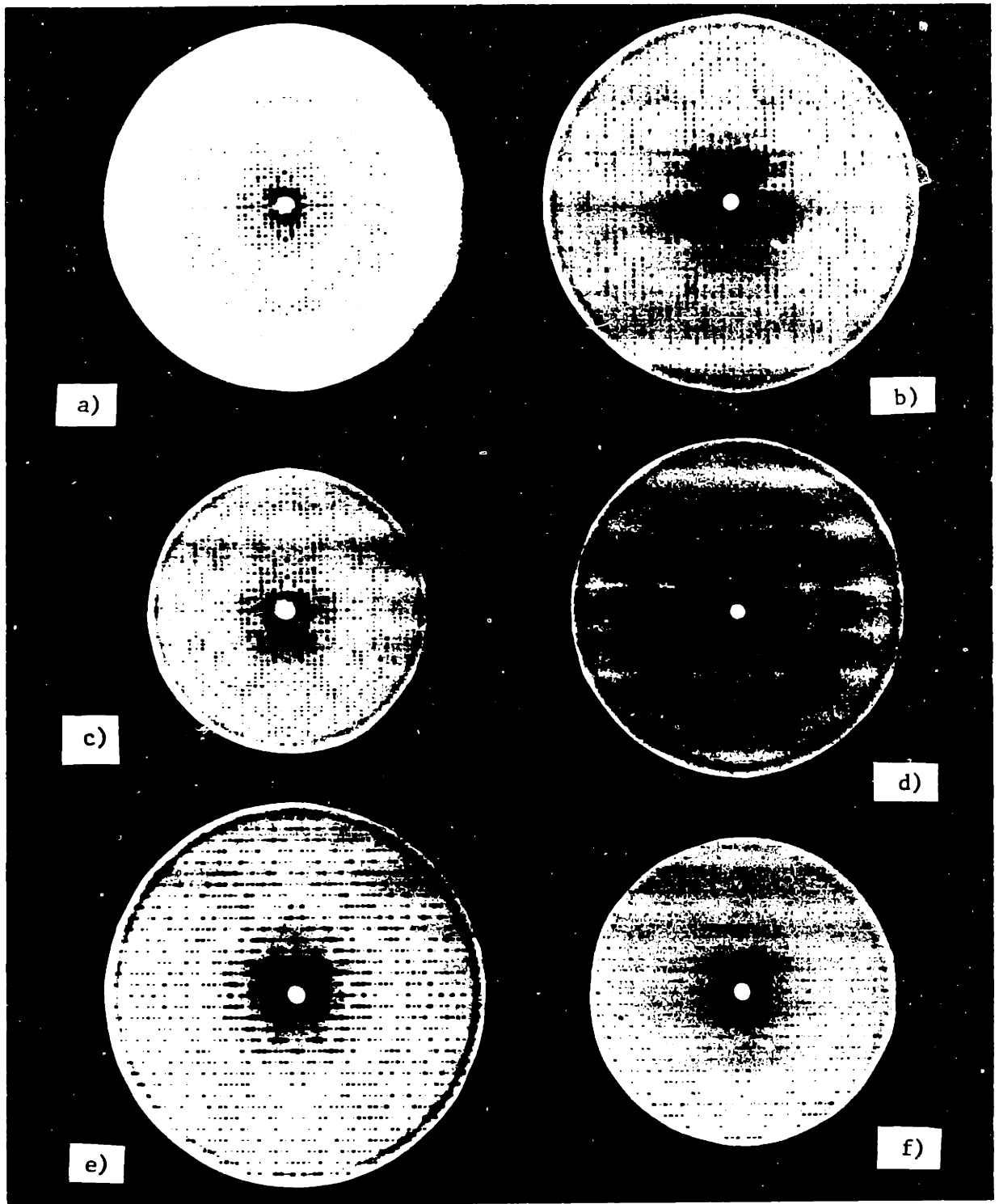


Figure 5. Precession photographs of yeast TIM native and derivative crystals. a) Native $hk0$, $\mu=21^\circ$. b) Baker's $hk0$, $\mu=15^\circ$. c) EMP $hk0$, $\mu=15^\circ$. d) $K_2Cr_2O_7$, $\mu=18^\circ$. e) Native $Ok1$, $\mu=15^\circ$. f) Baker's $Ok1$, $\mu=15^\circ$.

Table 3. Data collection statistics for diffractometer data from 60 to 3 Å resolution.

	<u>Crystal</u>				
	Native	EMP	BAKER'S	Pt(NO ₂) ₄	PtCl ₄
Measurements	12,162	15,049	23,620	23,379	19,975
Number Accepted	11,862	14,772	23,486	22,907	18,942
Mean F	25.5	34.1	30.4	20.1	29.6
Acentric Friedel R after scaling*		0.06	0.05	0.05	0.03
Mean fractional isomorphous change ⁺		0.17	0.21	0.11	0.11
Maximum radiation damage	17%	39%	26%	34%	37%

$$* = \frac{\sum_{hkl} (|F(hkl) - F(-h-k-l)|)}{\sum_{hkl} ((F(hkl) + F(-h-k-l))/2)}$$

$$+ = \frac{\sum_{hkl} (|F_{PH}(hkl) - F_P(hkl)|)}{\sum_{hkl} F_P(hkl)}$$

regarded as a measure of the differences in equivalent reflections which arise from factors such as machine and crystal misalignment, mounting asymmetry, fluctuations in the intensity of the beam, improper scaling and statistical fluctuations in the scattered energies. The agreement between equivalent reflections of 2% to 3.0 Å resolution may represent the base level of error in the diffractometer data.

C. Heavy Atom Parameters

The Baker's and EMP difference Patterson maps were very clean and were consistent with binding of both compounds to the same single site per monomer. The mercury atoms bound in the active site, probably to Cys 125. Baker's was hydrolyzed during the soak and bound as a monomeric compound (Figure 6).

The $\text{Pt}(\text{NO}_2)_4$ and PtCl_4 data also produced similar difference Patterson maps, but these were harder to interpret. The peaks on the Harker section overlapped and the cross vectors clustered around the origin (Figure 7). The heavy atom sites were adjacent to the noncrystallographic two-fold rotation axis and were only partially occupied. These factors made it difficult to accurately model the pattern of platinum binding. The refined heavy atom parameters are given in Table 4.

The heavy atom occupancies were refined in 10 shells of F_{HLE}^s of increasing resolution and were found to be constant. This demonstrated that the temperature factors were correctly estimated.

SECTION 33
YTIM-BAKER'S 3A 3-D MAP V SECTIONS FILES PATT
0 500

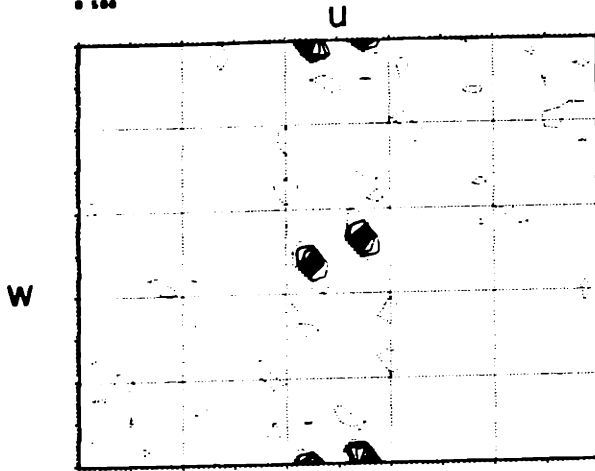
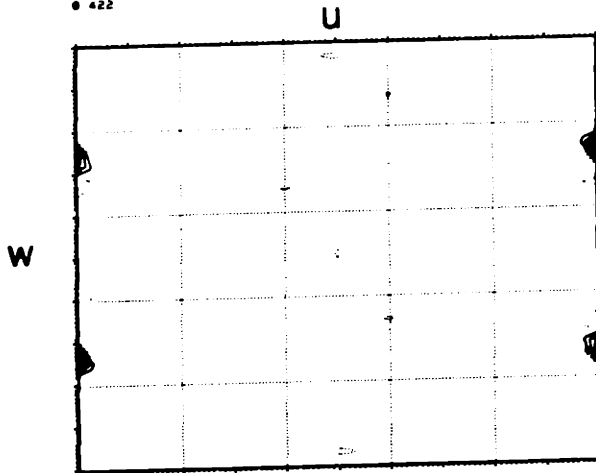


Figure 6. a-c) TIM + Baker's F_{HLE}^2 Patterson synthesis.

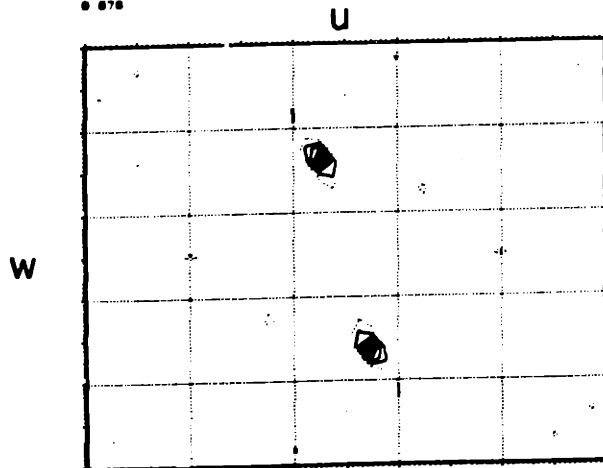
a) Harker section, $v = 0.500$

SECTION 28
YTIM-BAKER'S 3A 3-D MAP V SECTIONS FILES PATT
0 422



b) Cross vectors at $v = 0.422$

SECTION 4
YTIM-BAKER'S 3A 3-D MAP V SECTIONS FILES PATT
0 070



c) Cross vectors at $v = 0.78$. The rest of the map was featureless.

SECTION 33
TIM + BAKERS UNSCALED ANOMALOUS DIFF PATTERSON. 3A 3-D. 6-3-88
0 588

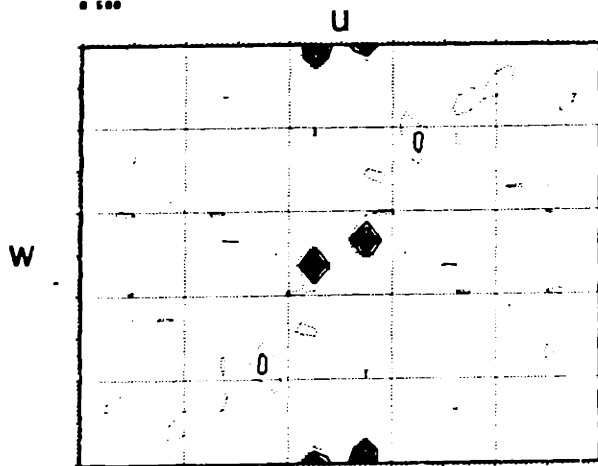
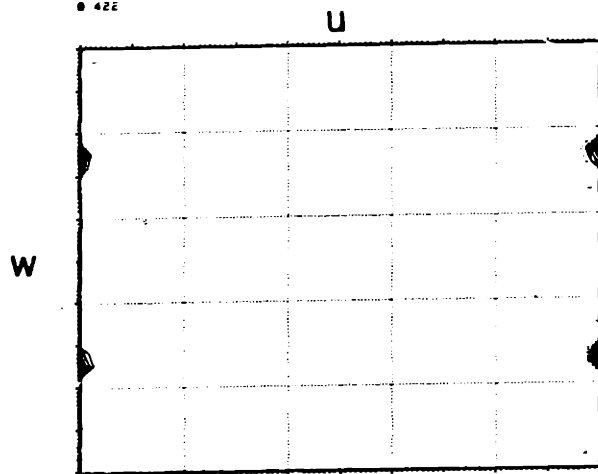


Figure 6. d-f) TIM + Baker's anomalous difference Patterson synthesis.

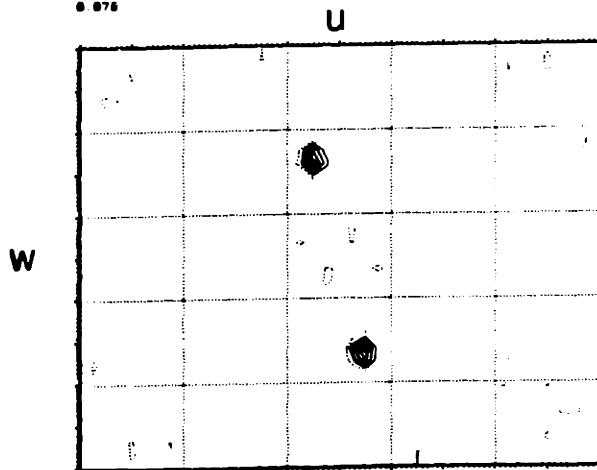
d) Harker section, $v = 0.500$

SECTION 26
TIM + BAKERS UNSCALED ANOMALOUS DIFF PATTERSON. 3A 3-D. 6-3-88
0 422



e) Cross vectors at $v = 0.422$

SECTION 2
TIM + BAKERS UNSCALED ANOMALOUS DIFF PATTERSON. 3A 3-D. 6-3-88
0 078



f) Cross vectors at $v = 0.78$.

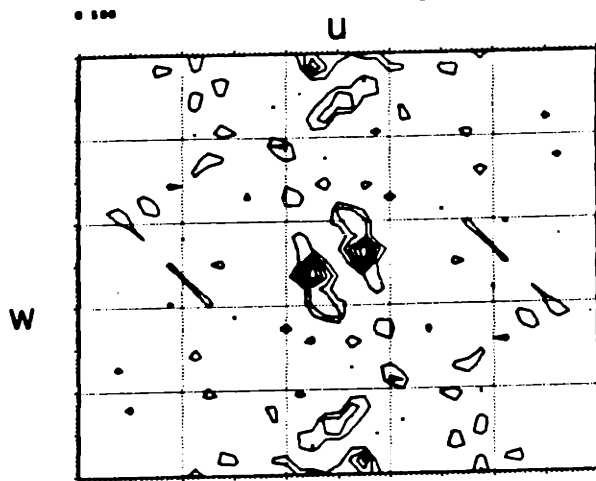
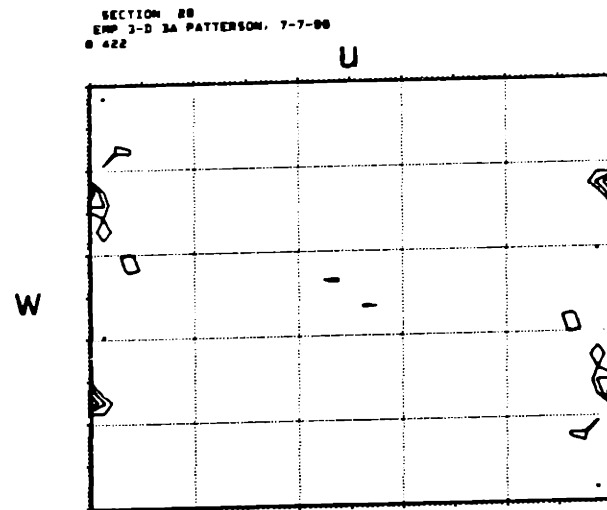
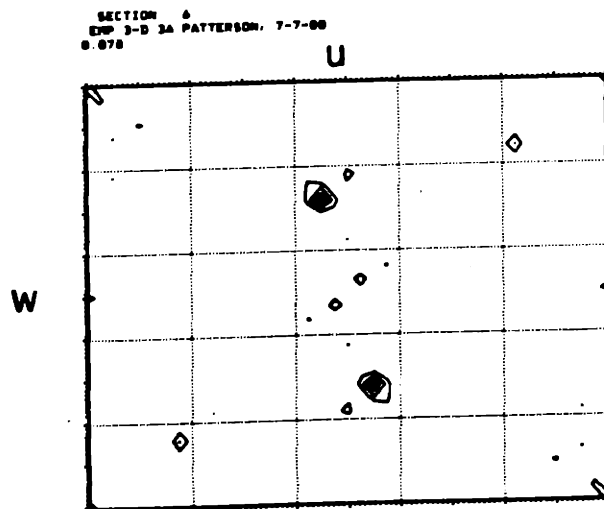


Figure 6 g-i) TIM + EMP isomorphous difference Patterson synthesis.

g) Harker section, $v = 0.500$



h) Cross vectors at $v = 0.422$



i) Cross vectors at $v = 0.78$. This map is similar to that of the TIM + Baker's differences.

SECTION 1
 YEAST TIM NO. 8 5 ISOMORPHOUS DIFFERENCE PATTERSON PTC14 6A
 0 500

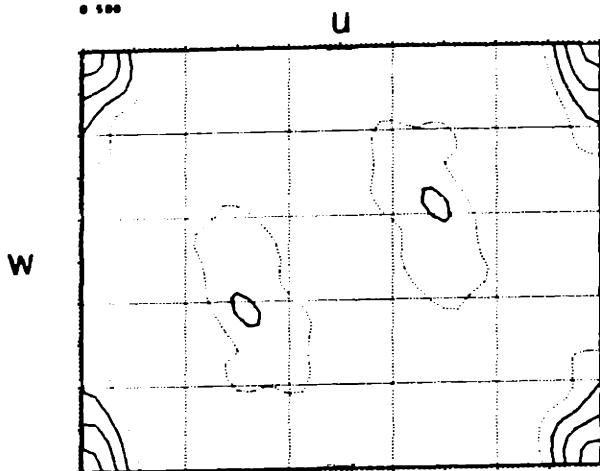
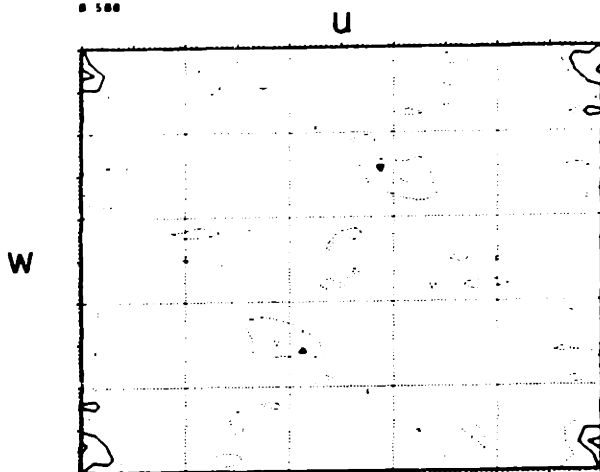


Figure 7. a-c) Platinum binding to yeast TIM.

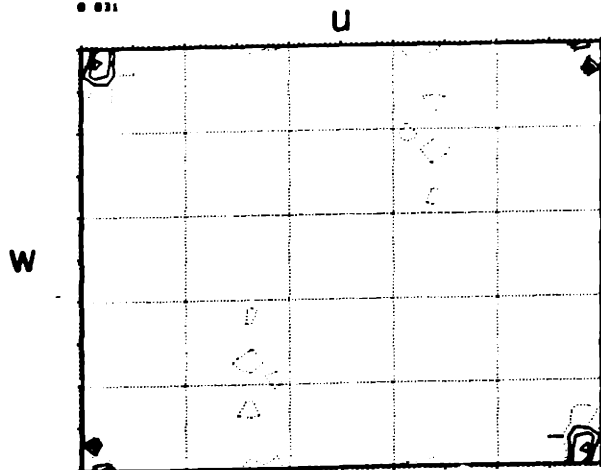
a) PtCl_4 isomorphous difference Patterson. Harker section, $v = 0.500$, at 6.0 \AA resolution.

SECTION 33
 YTIM+PTIND214 3A 3-D MAP V SECTIONS FILES PATT. 6-25-88
 0 500



b) $\text{Pt}(\text{NO}_2)_4 \text{ F}_{\text{HLE}}^2$ Patterson synthesis. Harker section, $v = 0.500$, at 3.0 \AA resolution.

SECTION 3
 YTIM+PTIND214 3A 3-D MAP V SECTIONS FILES PATT. 6-25-88
 0 031



c) $\text{Pt}(\text{NO}_2)_4 \text{ F}_{\text{HLE}}^2$ Patterson synthesis. Cross vectors fall near the origin, $v = 0.031$.

Table 4. Heavy atom parameters.

Derivative	Kraut Scale Factor		x	y	z	Occ.	B
EMP	1.03	1	0.22295	0.00346	0.26721	0.47	30.3
		2	0.22446	0.41621	0.00179	0.38	11.1
Baker's	1.045	1	0.22655	0.00000	0.26639	0.64	11.2
		2	0.22579	0.41522	0.00552	0.79	16.9
K ₂ Pt(NO ₂) ₄	1.02	1	0.02262	0.03575	0.03141	0.53	93.1
		2	0.03125	0.99716	0.99421	0.40	66.2
		3	0.96994	0.02765	0.03091	0.19	-13.5
		4	0.96373	0.04362	0.11079	0.39	101.0

EMP

sin ² R/g ²	\bar{R}	\overline{Wiso}	\overline{Wanom}	$\overline{F_{calc}}$	k _{emp}	R _{FHLE}	$\langle a_P - a_H \rangle$
0-0.003	13.02	0.14		6.3			
0.006	7.7	0.16		5.9			
0.008	5.99	0.18		5.4			
0.011	5.07	0.17		5.1			
0.014	4.48	0.15		4.6			
0.017	4.05	0.16	1.58	4.4	5.2		
0.019	3.73	0.16	1.48	4.1	5.5		
0.022	3.47	0.17	1.45	3.9	5.9		
0.025	3.26	0.19	1.41	3.6	5.9		
0.028	3.00	0.19	1.38	3.5	5.4		
Overall	60-3.0	0.17		4.5	5.6		

Table 4. Heavy atom parameters, continued . . .

$\sin^2 \theta / \lambda^2$	\bar{R}	Baker's			k_{emp}	R_{FHLE}	$\langle \alpha_{\text{P}} - \alpha_{\text{H}} \rangle$
		$\overline{\Delta_{\text{iso}}}$	$\overline{\Delta_{\text{anom}}}$	$\overline{F_{\text{calc}}}$			
0-0.003	13.02	0.18	1.86	10.8	6.8	0.41	
0.006	7.70	0.21	1.63	10.3	7.2	0.33	
0.008	5.99	0.25	1.31	9.7	7.8	0.33	
0.011	5.07	0.21	1.27	9.2	8.6	0.36	
0.014	4.48	0.19	1.32	8.7	8.2	0.37	
0.017	4.05	0.19	1.15	8.2	9.0	0.38	
0.019	3.73	0.19	1.14	7.8	8.4	0.37	
0.022	3.47	0.21	1.13	7.4	8.0	0.37	
0.025	3.26	0.21	1.20	7.1	7.0	0.38	
0.028	3.00	0.22	1.17	6.7	6.4	0.39	
Overall	60.0-3.0	0.21	1.25	8.2	7.8	0.37	87.4°

$\sin^2 \theta / \lambda^2$	\bar{R}	Pt(NO ₂) ₄			k_{emp}	R_{FHLE}	$\langle \alpha_{\text{P}} - \alpha_{\text{H}} \rangle$
		$\overline{\Delta_{\text{iso}}}$	$\overline{\Delta_{\text{anom}}}$	$\overline{F_{\text{calc}}}$			
0-0.003	13.02	0.17	1.45	9.7	7.0	0.36	
0.006	7.70	0.11	1.04	5.3	5.3	0.43	
0.008	5.99	0.10	1.00	3.8	4.3	0.45	
0.011	5.07	0.08	1.00	3.4	3.6	0.49	
0.014	4.48	0.08	1.12	3.2	2.9	0.48	
0.017	4.05	0.08	1.09	3.1	2.8	0.44	
0.019	3.73	0.09	1.13	2.9	2.8	0.43	
0.022	3.47	0.10	1.13	2.7	2.8	0.44	
0.025	3.26	0.12	1.11	2.6	2.6	0.47	
0.028	3.00	0.13	1.11	2.6	2.7	0.50	
Overall	60.0-3.0	0.10	1.11	3.6	3.7	0.45	87.4°

$\overline{\Delta_{\text{iso}}}$ = mean fractional isomorphous difference

$\overline{\Delta_{\text{anom}}}$ = mean anomalous difference

$$k_{\text{emp}} = \frac{2 \sum_{\text{hkl}} |F_{\text{PH}} - F_{\text{P}}|}{\sum_{\text{hkl}} |\Delta_{\text{anom}}|}$$

$$R_{\text{FHLE}} = \frac{\sum_{\text{hkl}} |F_{\text{HLE}} - F_{\text{H}}^{\text{calc}}|}{\sum_{\text{hkl}} F_{\text{HLE}}}$$

D. Phase Calculation

The E values for the centric and acentric reflections are given in Table 5 as a function of resolution for the Baker's, EMP and platinum nitrite derivatives. Figure 8 shows the figure of merit histogram and the dependence of the figure of merit on resolution. The average figure of merit was 0.76.

E. Electron Density Maps

11,828 of the approximately 12,000 reflections which occur between 60 and 3 Å resolution were included in the Fourier summation. The electron density map was contoured perpendicular to the b and c axes, transferred to acetate sheets and stacked. The triose phosphate isomerase dimers were seen immediately, surrounded by large channels of solvent. The solvent channels were up to 35 Å in diameter. The heavy atom positions were used to locate the intramolecular two fold rotation axes to verify the identification of the molecules.

The chain was traced as described above. Several sections of the electron density map stacked perpendicular to y are shown in Figure 9. The dots represent α-carbon positions. Figure 9 shows a cross-section of both monomers. The obvious parallel strands of chain in the upper half of the density correspond to several strands in the β-barrel. The large solvent channels are also apparent.

Table 5. E values for the Baker's, platinum nitrite and ethyl mercuric phosphate derivatives of yeast TIM.

Number of reflections = 11,828

Resolution = 60.0 - 3.0 Å

Mean figure of merit = 0.76

Baker's				
<u>sin²θ/λ²</u>	<u>Å</u>	Isomorphous		<u>Anomalous</u>
		<u>Centric</u>	<u>Acentric</u>	
0-0.003	13.02	50.05	35.52	18.70
0.006	7.70	41.26	26.13	13.00
0.008	5.99	42.49	23.26	9.30
0.011	5.07	38.10	26.68	8.90
0.014	4.48	51.71	32.09	9.10
0.017	4.05	57.51	35.12	8.20
0.019	3.73	54.85	29.15	8.20
0.022	3.47	51.89	27.83	8.90
0.025	3.26	48.48	27.13	9.80
0.028	3.00	48.82	26.47	9.90
Overall	60.0-3.0	48.71	28.86	9.68

Pt(NO ₂) ₄				
<u>sin²θ/λ²</u>	<u>Å</u>	Isomorphous		<u>Anomalous</u>
		<u>Centric</u>	<u>Acentric</u>	
0-0.003	13.02	48.36	29.26	14.40
0.006	7.70	33.38	24.60	13.30
0.008	5.99	33.25	21.88	13.00
0.011	5.07	32.70	23.49	13.60
0.014	4.48	38.63	23.70	15.10
0.017	4.05	30.00	22.50	14.90
0.019	3.73	38.94	23.26	15.50
0.022	3.47	30.30	22.89	15.30
0.025	3.26	30.83	21.38	15.50
0.028	3.00	30.77	20.29	15.00
Overall	60.0-3.0	35.40	22.77	14.66

Table 5. Continued . . .

<u>$\sin^2 \theta / \lambda^2$</u>	<u>R</u>	Isomorphous	
		<u>Centric</u>	<u>Acentric</u>
0-0.003	13.02	42.59	31.83
0.006	7.70	36.38	22.77
0.008	5.99	34.74	22.41
0.011	5.07	34.33	26.35
0.014	4.48	38.90	30.03
0.017	4.05	51.39	30.33
0.019	3.73	43.86	28.37
0.022	3.47	42.08	28.78
0.025	3.26	49.16	29.29
0.028	3.00	45.97	27.93
Overall	60.0-3.0	42.10	28.14

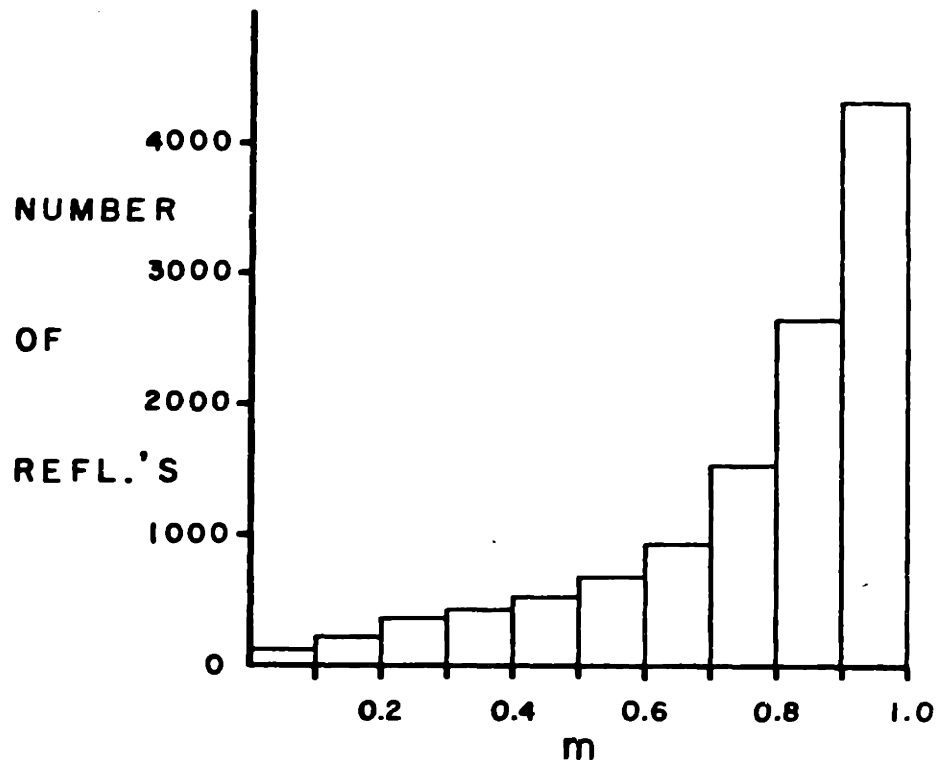
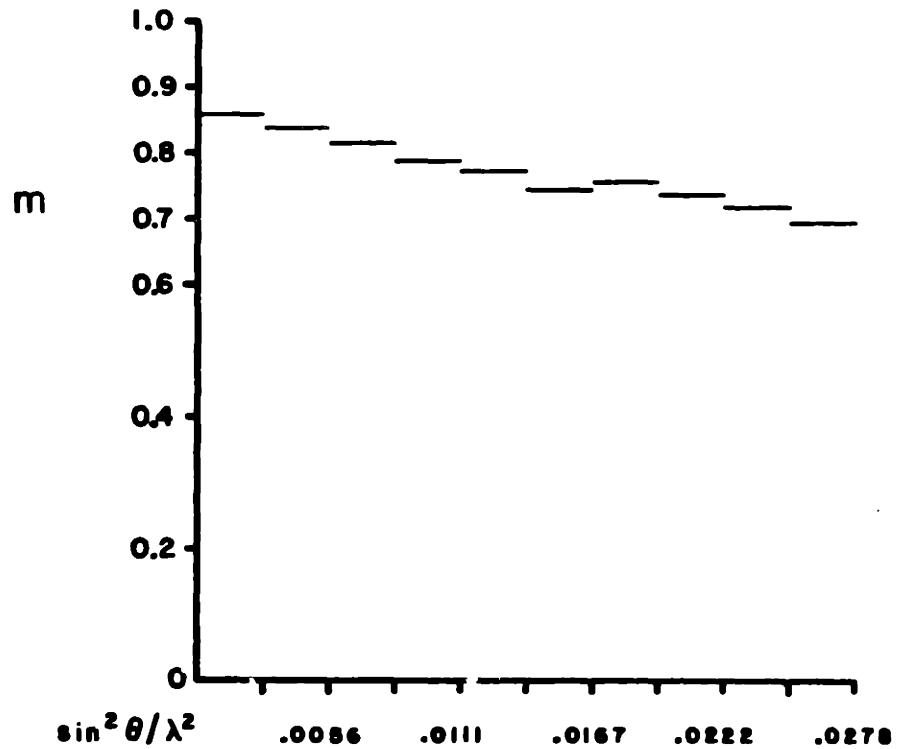


Figure 8. a) Figure of merit histogram for the phases calculated using data from the Baker's and platinum nitrite derivatives.



b) The figure of merit as a function of resolution.



Figure 9. Several sections of the yeast TIM 3.0 Å electron density map. One dimer is shown. The parallel strands of density in the upper half of the figure are part of the β -barrel of one subunit. The density corresponding to the disordered loop, residues 168 to 177 is in the upper right of the molecule. The solvent channel are apparent. Dots were placed on the α -carbon positions during the tracing process.

Using coordinates estimated from the locations of the dots on the minimaps, the r.m.s. difference in the positions of 15 equivalent α -carbons from conserved regions in the yeast and chicken monomers was 0.9 Å. This is comparable to the r.m.s. difference of 1.2 Å in the positions of equivalent atoms in the subunits of the chicken TIM dimer, although the latter r.m.s. difference was derived from a comparison of all the non-hydrogen atoms in the structure (30).

Using the RING system of programs on a Vector General graphics display system (31), coordinates of eight amino acids (31 atoms) in conserved regions of the structure were measured. Including side chain atoms, the r.m.s. difference between the coordinates of these atoms in the yeast and chicken monomers was 0.5 Å. The matrices which relate the chicken and yeast proteins are:

$$\begin{aligned}
 R &= \begin{pmatrix} -0.27750 & 0.71122 & -0.64588 \\ -0.91200 & -0.40640 & -0.05569 \\ -0.30209 & 0.57358 & 0.76141 \end{pmatrix} \\
 T &= \begin{pmatrix} 39.02902 & 94.41542 & 24.97730 \end{pmatrix}
 \end{aligned}$$

The transformed coordinates of the chicken enzyme were used to trace the second subunit of the yeast TIM dimer. In general, the features of secondary structure coincided; large differences in the path of the polypeptide chain were confined to the loops of irregular structure which connect the α -helices and strands of β -sheet. The extra lysine residue at position 56 is accommodated by a bulge in one of the loops.

The density corresponding to a loop containing residues 168 to 177 was all but absent in one subunit and stronger, but jumbled, in the

other. The weakness of the density suggests that these residues are disordered in the yeast TIM crystals. The upper right corner of Figure 10 shows this feature. Since the lowest contour level was set at $0.25 \text{ e}^-/\text{\AA}^3$ and the electron density of the polypeptide chain is about $1 \text{ e}^-/\text{\AA}^3$, the absence of continuous density for residues 168 to 177 implies that this loop adopts a minimum of four conformations in the crystal.

VI. Discussion

The yeast triose phosphate isomerase structure was solved at 3.0 \AA resolution by exploiting the isomorphous and anomalous differences of two heavy atom derivatives. The Baker's derivative provided the bulk of the phasing power. The resulting electron density map was quite clear. For the most part, the path of the polypeptide chain was easily followed and many of the amino acid side chains are credibly represented by the density.

Despite differences in their amino acid sequences and in the nature of the intermolecular contacts in the crystals, the overall conformations of yeast and chicken TIM are very similar. The secondary structural features of the chicken enzyme monomer include a β -barrel of eight strands surrounded by 12 α -helices. These features are also present in yeast TIM. In several regions of the molecules where the amino acid sequence is conserved, the r.m.s. difference in the positions of equivalent atoms is smaller than the error inherent in the yeast TIM electron density map. The most obvious differences in the backbone

conformation occur in loops on the surface of the molecules which connect strands in secondary structures.

The coordinates which were used to trace the polypeptide chain in the second monomer of yeast TIM were calculated using a matrix based on α -carbon positions in the first monomer. Since these coordinates predicted the α -carbon positions in the second subunit quite accurately, it can be concluded that the subunits in the chicken and yeast proteins have the same spatial relationship. This finding is somewhat enigmatic, because several of the amino acid sequence differences occur in the subunit interface. In particular, Met 14 in chicken TIM, a residue which fits into a hydrophobic pocket formed by residues 70' to 80' on the periphery of the subunit interface, is replaced by a leucine residue in the yeast enzyme. Other changes in the subunit interface include: Ser 45 in the chicken enzyme to ala, Ile 46 to thr, Ile 48 to leu, Gln 53 to ser, Val 69 to lys, Ile 78 to asn, Pro 80 to val, Met 82 to gln and Ile 86 to val (5).

The close similarity of the structures of TIM's from yeast and chicken has a number of implications. It is apparent that a number of primary structures can form a stable and active TIM fold. Indeed, active heterodimers of the chicken and B. stearothermophilus enzymes, which are only 37% homologous in sequence, have been reported (32). Only the residues which are thought to form the catalytic features of the molecule are conserved in all five sequences which have been determined (Chapter II). A similar relationship has been observed among members of a number of protein families, including the serine proteases,

immunoglobulins, cytochrome C and myoglobins (33). In the sperm whale and seal myoglobins, the amino acid differences which occur on the helix-helix contact surfaces are all pairwise changes which preserve the helix packing geometry (34,35). A detailed interpretation of the yeast TIM electron density map and refinement against high resolution data will facilitate a residue-by-residue comparison of the yeast and chicken TIM structures and illuminate the ways in which common conformations are built from different sequences. This work is in progress.

One important function of the primary sequence is to direct the folding of the polypeptide chain. The similarity of the structures of triose phosphate isomerases from different organisms suggests that the folding pathway of the molecule is consistent with wide variations in amino acid sequence. Amino acid differences in the homologous proteins bovine pancreatic trypsin inhibitor and black mamba toxin (36,37) have been found to affect the kinetics of the disulfide interchange reactions on the folding pathway in vitro but not their sequence or the nature of the cysteine pairs in the folded structure (T.E. Creighton, unpublished results). Thus, either the information which specifies the fold is concentrated in conserved regions, or alternatively, residues throughout the sequence may influence folding, but different pathways and structures may accommodate different amounts of variation. Another possibility is that changes in primary structure must be coupled to preserve interactions which occur during folding. Mapping of genetic lesions which affect the folding of the protein without reducing its stability may help distinguish between these models (38,39).

The close similarity of the structures of yeast and chicken triose phosphate isomerases which is already apparent provides a structural understanding of the similar kinetic and thermodynamic properties of these enzymes. This work sets the stage for structural studies of the interactions of yeast triose phosphate isomerase with substrates and inhibitors.

References

1. Alber, T., Hartman, F.C., Johnson, R.M., Petsko, G.A. and Tsernoglou, D. (1981), *J. Biol. Chem.*, 256, 1356-1361.
2. Johnson, L.N. and Waley, S.G. (1967), *J. Mol. Biol.*, 29, 321-322.
3. Banner, D.W., Bloomer, A.C., Petsko, G.A. and Phillips, D.C. (1971), *Cold Spring Harbor Symp. Quant. Biol.*, 36, 151-153.
4. Norton, I.L. and Hartman, F.C. (1972), *Biochemistry*, 11, 4435-4441.
5. Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., Pogson, C.I., Wilson, I.A., Corran, P.H., Furth, A.J., Milman, J.D., Offord, R.E., Priddle, J.D. and Waley, S.G. (1975), *Nature*, 255, 601-614.
6. Rossman, M.G., ed., The Molecular Replacement Method, Gordon and Breach, Science Publishers, New York, 1972.
7. Phillips, D.C., Sternberg, M.J.E., Thornton, J.M. and Wilson, I.A. (1977), *J. Mol. Biol.*, 119, 329-351.
8. Crick, F.H.C. and Magdoff, B.S. (1956), *Acta Cryst.*, 9, 901-908.
9. Eisenberg, D., in The Enzymes, ed. P.D. Boyer, Academic Press, New York, 1970.
10. Holmes, K.C. and Blow, D.M., The Use of X-Ray Diffraction in the Study of Protein and Nucleic Acid Structure, John Wiley and Sons, New York, 1965.
11. Stout, G.H. and Jensen, L.H., X-Ray Structure Determination, MacMillan Publishing Co., New York, 1968.
12. Blundell, T.L. and Johnson, L.N., Protein Crystallography, Academic Press, New York, 1976.
13. Patterson, A.L. (1934), *Phys. Rev.*, 46, 372.
14. Harker, D. (1956), *Acta Cryst.*, 9, 1.
15. Sieker, L.C., Adman, G. and Jensen, L.H. (1972), *Nature*, 235, 40-42.
16. Matthews, B.W. (1966), *Acta Cryst. A.*, 20, 230-239.
17. Blow, D.M. and Crick, F.H.C. (1959), *Acta Cryst.*, 12, 794-803.
18. North, A.C.T. (1965) *Acta. Cryst.*, 18, 212-216.
19. Matthews, B.W. (1966), *Acta Cryst. A.*, 20, 82-86.

20. Cullis, A.F., Muirhead, H., Perutz, M.F., Rossmann, M.G. and North, A.C.T. (1961), Proc. Roy. Soc. A, 265, 161.
21. Dickerson, R.E., Kendrew, J.C. and Strandberg, B.E. (1961), Acta Cryst., 14, 1188-1195.
22. Petsko, G.A. (1975), J. Mol. Biol., 96, 381-392.
23. Matthews, D.A., Alden, R.A., Bolin, J.T., Freer, S.T., Hamlin, R., Xuong, N., Kraut, J., Poe, M., Williams, M. and Hoogsteen, K. (1977), Science, 197, 452-455.
24. Wyckoff, H.W., Doscher, M., Tsernoglou, D., Inagami, T., Johnson, L.N., Hardman, K.D., Allewell, N.M., Kelly, D.M. and Richards, F.M. (1967), J. Mol. Biol., 27, 563-578.
25. North, A.C.T., Phillips, D.C. and Mathews, F.S. (1968), Acta Cryst. A., 24, 351-359.
26. Bragg, W.L., James, R.W. and Bosanquet, C.M. (1921), Phil. Mag., 42, 1.
27. Matthews, B.W. and Czerwinski, G.W. (1975), Acta Cryst. A., 31, 480-487.
28. Dodson, E.J. and Vijayan, M. (1971), Acta Cryst. B., 27, 2402-2411.
29. Hill, E., Tsernoglou, D., Webb, L. and Banazak, L.J. (1972), J. Mol. Biol., 72, 577-591.
30. Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C. and Wilson, I.A. (1976), Biochem. Biophys. Res. Comm., 72, 146-155.
31. Jones, T.A. (1978), J. Appl. Cryst., 11, 268-272.
32. Artavanis-Tsakonas, S. and Harris, J.I. (1980), Eur. J. Biochem., 108, 599-611.
33. Richardson, J.S. (1981), Adv. Prot. Chem., 34, 167-339.
34. Takano, T. (1977), J. Mol. Biol., 110, 537-568.
35. Scouloudi, H. (1979), J. Mol. Bio., 126, 661-671.
36. Deisenhofer, J. and Steigemann, W. (1975), Acta Cryst. B., 31, 238-250.
37. Strydom, D.J. (1977), Biochem. Biophys. Acta, 491, 361.
38. Goldenberg, D.P. and King, J. (1981), J. Mol. Biol., 145, 633-651.
39. Smith, D.H. and King, J. (1981), J. Mol. Biol., 145, 653-676.

Chapter IV.

Crystal Structure of a Michaelis Complex:

Interactions of Substrates with Triose Phosphate Isomerase

I. Introduction

Substantial evidence for changes in the conformation of the TIM molecule during the catalytic cycle has been reported. For example, the binding of ligands in the active site changes the absorption spectrum of the protein (1,2) and results in disordering and alteration of the unit cell dimensions of crystals of the chicken muscle enzyme (2,3). The binding of DHAP to crystalline chicken TIM was observed at 6.0 Å resolution by Phillips and coworkers (4). Difference Fouriers showed the substrate bound to only one subunit of the dimer and indicated that a loop of the polypeptide chain composed of residues 168 to 176 moves over 8 Å towards the active site to interact with the substrate. Difference density was apparent in other regions of both subunits, even though only one active site was occupied. Phosphoglycolate (PGA), a possible transition state analogue (2), and inorganic phosphate also caused the loop movement associated with substrate binding (4), again in only one subunit of the crystalline chicken TIM dimer.

Unfortunately, the disordering of the chicken enzyme crystals caused by ligand binding has frustrated efforts to take data which would reveal the structure of the enzyme-substrate complex at high resolution. Our efforts to demonstrate that the crystals of yeast TIM are suitable for such studies are the subject of this chapter.

II. Materials and Methods

A. Crystalline complexes

Triose phosphate isomerase from *S. cerevisiae* Hansen was generously provided by F. Hartman. Crystals of the free enzyme were grown as described in Chapter III, and the crystals were transferred to mother liquor containing substrate or inhibitor as described above for the preparation of heavy atom derivatives. The concentration of ligands used in the transfer experiments ranged from 10^{-2} mM to 5 mM for PGA and 1 to 50 mM for DHAP.

PGA was obtained from the Sigma Chemical Co. as the trimonocyclohexylamine salt. In the early crystallization trials, it was converted to the free acid form prior to use as recommended by the manufacturer. DHAP (Sigma) was used as the lithium salt or the free acid. PEG 4000 was from the Baker Chemical Co., Tris base and 2-mercaptoethanol were the electrophoresis grade reagents from Biorad.

Yeast TIM was cocrystallized with PGA at room temperature by batch methods and by vapor diffusion. The protein concentration was varied between 10 and 20 mg/ml in 50 mM Tris-HCl, 1 mM EDTA, 1 mM 2-mercaptoethanol at pH 7.1 to 7.4. The PGA concentration was 0.3 to 1.9 times the active site concentration. In the batch experiments, solid PEG 4000 was added until the TIM-PGA solution became barely turbid. Distilled water (5 to 30 μ l/100 μ l) was added until the turbidity disappeared, and crystals grew in 12 to 24 hours. Similar experiments at 4^o C using DHAP instead of PGA produced poorly formed microcrystals in 12 hours.

Crystals were mounted for x-ray photography in sealed glass capillary tubes containing a slug of mother liquor. X-ray photographs were taken on Buerger precession cameras (Charles Supper Co. and Enraf-Nonius, Inc.) using nickel-filtered copper K _{α} radiation. X-ray sources included an Elliot rotating anode and an Enraf-Nonius sealed tube x-ray generator.

B. Data Collection

For diffractometer measurements of the TIM-DHAP complex, a crystal of the free enzyme was mounted in a flow cell (Figure 1) (5) in sulfate-free mother liquor and placed on a Nicolet P3 diffractometer at -5^o C. A Nicolet LT-1 low temperature device was used to control the temperature.

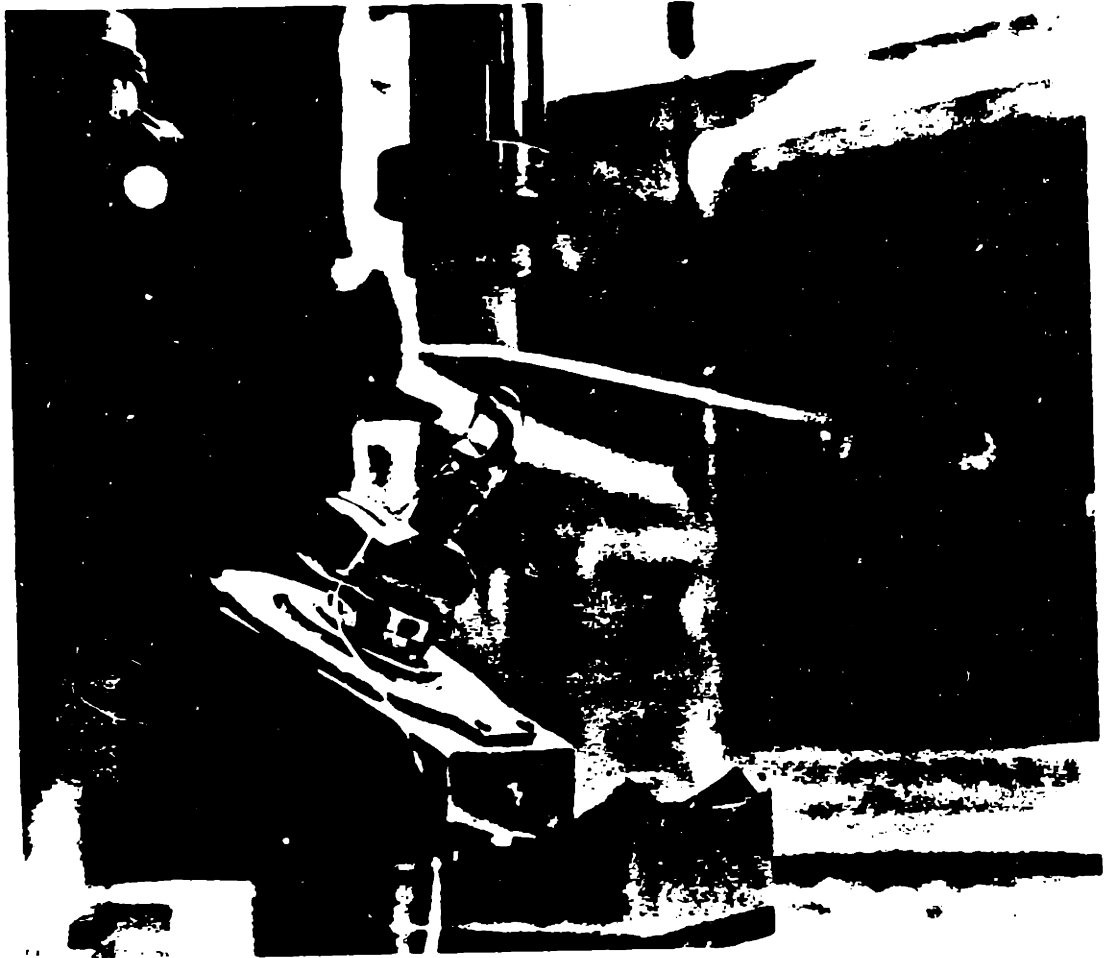


Figure 1. Flow cell and goniostat. The crystal is mounted in a bed of pipe cleaner fibers in a 0.7 mm capillary tube on the C-shaped yoke. The substrate can be flowed into the capillary through the tubes coming in the top and going out the bottom. The nozzle of the cooling device is above the flow cell, and the x-ray source is on the left.

The crystal was aligned and equilibrated with mother liquor containing 3 mM DHAP for 18 hours. The flow rate of this solution was approximately 0.5-1.0 ml/day.

Data were collected using the stationary counter/moving crystal method with the diffractometer in the ω -scan mode. Seventeen steps covering 2° in ω were summed and background measurements made for each reflection were subtracted to yield the integrated intensity. The peak profiles were inspected for evidence of interference of the diffracted beam by the brass yoke of the flow cell. Reflections where blocking occurred were removed from the data set. Nickel filtered copper K_α x-rays were used. The diffractometer was aligned prior to use. Reflections representing interplanar spacings of 60 to 3.5 \AA were measured.

After data collection was completed, substrate-free mother liquor was flowed over the crystal for 24 hours, and the 3-dimensional data to 6 \AA resolution were remeasured.

C. Data Reduction, Scaling and Fourier

Radiation damage was monitored by periodic measurement of 5 standard reflections. A linear radiation damage correction was applied to the data. No absorption correction was made (the flow cell makes it superfluous) and standard Lorentz and polarization corrections were applied (6).

Data were scaled to the native structure factor amplitudes by the local scaling method of Matthews (7).

Fouriers were calculated using the phases of the free enzyme obtained by the method of multiple isomorphous replacement as described in Chapter III. Maps were calculated using the coefficients $F_{\text{DHAP}}(\text{hkl}) - F_{\text{P}}(\text{hkl})$ and $3F_{\text{DHAP}}(\text{hkl}) - 2F_{\text{P}}(\text{hkl})$.

The difference density was contoured on minimaps, and the $3F_{\text{DHAP}} - 2F_{\text{P}}$ map was analyzed using the program FRODO (8) on a Vector General graphics system. All calculations were performed using a Digital Equipment Corporation PDP 11/60 computer.

III. Results

A. Crystalline Complexes of Yeast TIM with DHAP and PGA

In order to investigate the feasibility of x-ray structural studies of the complexes of DHAP and PGA with yeast TIM, the free enzyme crystals were soaked in mother liquor containing these ligands. When the DHAP concentration was greater than 5 mM, the free enzyme crystals tended to crack and disorder. However, the crystals appeared to be stable in 1 to 3 mM DHAP, and diffraction could be observed to 1.8 Å resolution on still photographs (Figure 2). Unfortunately, the diffraction pattern decayed rapidly; the crystals showed a marked tendency to dissolve when exposed to x-rays at room temperature. The use of a cooling device (9) to keep the temperature between +5 and +10°

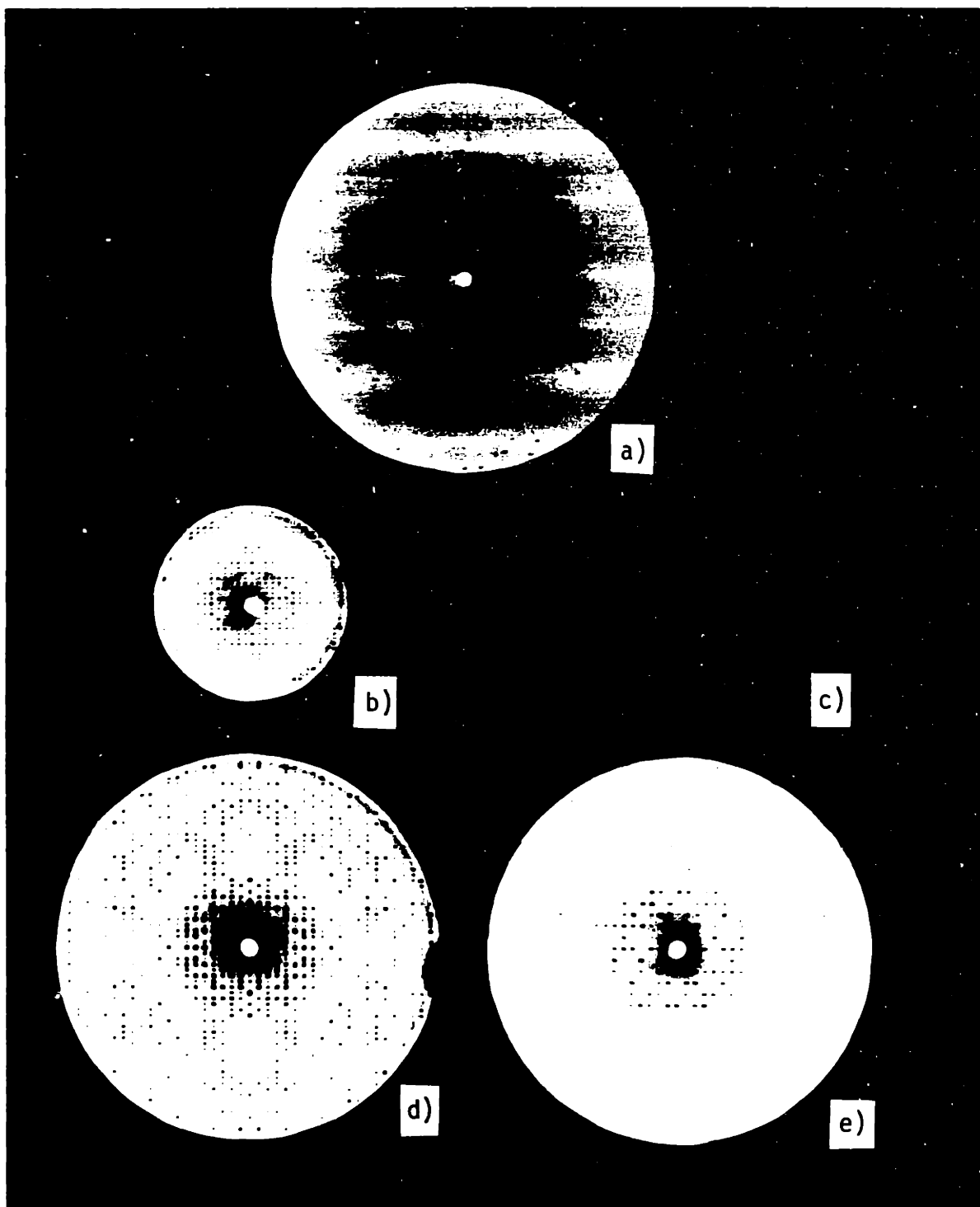


Figure 2. X-ray photographs of crystals of the free enzyme soaked in 5 mM DHAP for 24 hours. a) Initial still photographs diffract to at least 1.8 Å. b) Exposure to the x-ray beam results in a rapid decay of the strength of the diffraction pattern, changes in the unit cell dimensions and, ultimately dissolution of the crystal. hk0 precession photograph, $\mu = 10^0$. c) h01 zone, $\mu = 10^0$. d) Native hk0 precession photograph for comparison. e) Native h01 precession photograph for comparison.

C slowed this process but did not stop it. Two precession photographs showed highly mosaic reflections and altered unit cell dimensions, suggesting that the crystal lattice may have changed slowly before the crystals dissolved (Figure 2). The new unit cell dimensions calculated from the precession photographs shown in Figure 2 were $a = 61.2 \text{ \AA}$, $b = 82.4 \text{ \AA}$, $c = 37.64 \text{ \AA}$ and $\beta = 92.6^\circ$.

The crystals of the free enzyme also dissolved when the mother liquor contained even submillimolar concentrations of PGA. The lowest concentration of PGA used was, however, 10^{-2} mM or approximately $10 \times K_i$.

Cocrystallization experiments yielded well formed crystals with two different habits (Figure 3a,b). The diffraction of x-rays from the large elongated plates shown in Figure 3a could not be observed. The smaller chunky crystals, on the other hand, diffract to at least 1.8 \AA resolution (Figure 3c). Precession photographs indicated that the crystals are monoclinic, with the symmetry of space group $P2_1$. The unit cell dimensions are quite different from those of the crystals of the free enzyme; $a = 74.2 \text{ \AA}$, $b = 82.9 \text{ \AA}$, $c = 37.6 \text{ \AA}$ and $\beta = 102^\circ$ (Figure 3d,e,f). A comparison of the unit cell volume and the molecular weight of the protein indicates that there is one dimeric molecule per asymmetric unit.

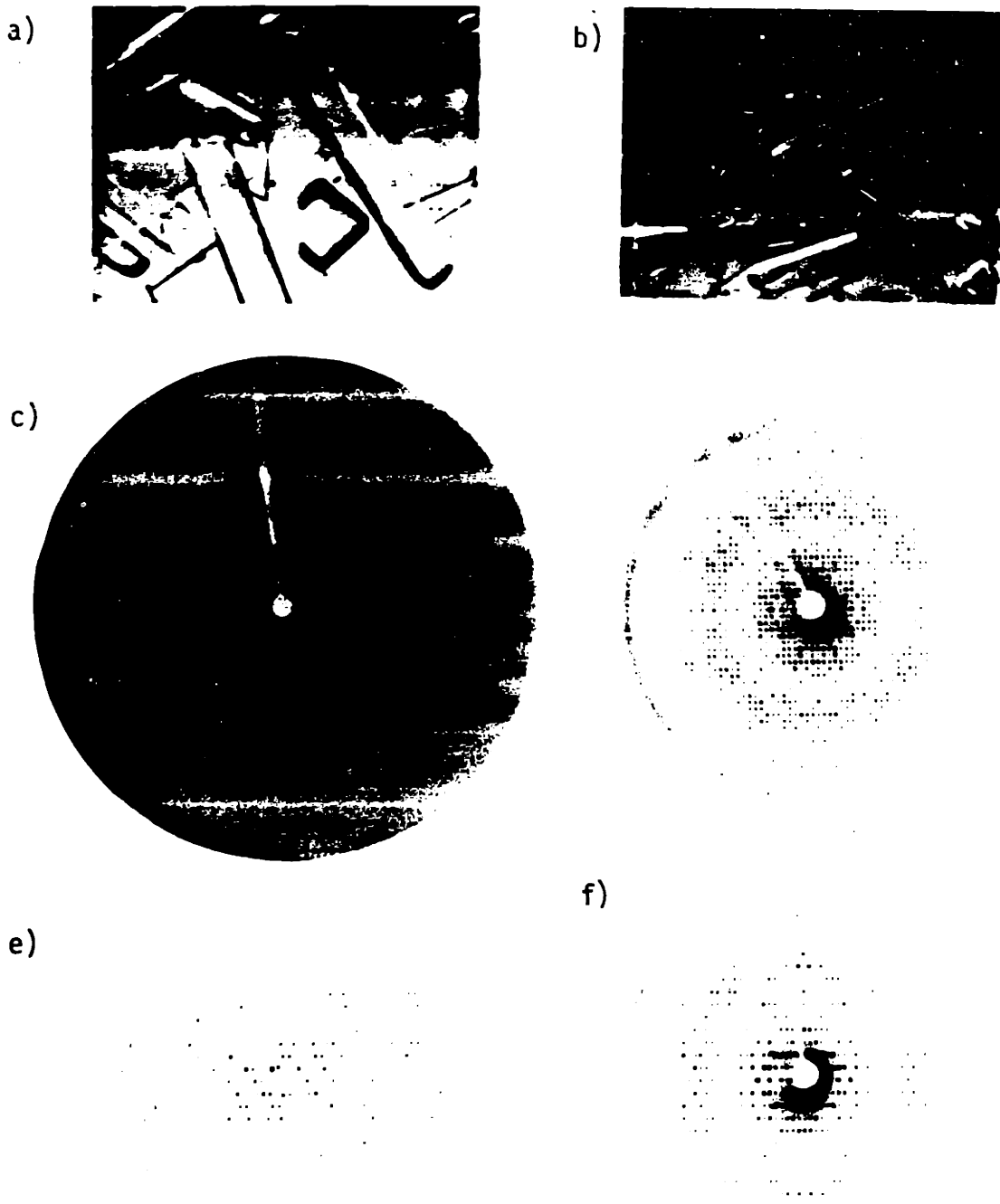


Figure 3. Cocrytals of yeast TIM and phosphoglycolate. a-b) Batch crystallizations. The chunky crystal in (a) is 0.5 mm across. c) Still photograph with spots at 2.0 Å resolution. d) hk0 precession photograph, $\mu = 18^\circ$. e) h0l precession photograph, $\mu = 15^\circ$. f) 0kl precession photograph, $\mu = 18^\circ$.

B. The Structure of the Yeast TIM-DHAP Complex

Despite the apparent reduced stability of the TIM-DHAP crystals, an attempt was made to collect 3-dimensional diffraction data from them. In order to reduce any possible deleterious effects of the enzymatic dephosphorylation of the substrate -- including reduced occupancy of the substrates at the active sites and damage to the crystals which might result from covalent or noncovalent binding of methyl-glyoxal to the protein -- fresh 3 mM DHAP solution was flowed over the crystal at -5° during data collection. This procedure also allowed the crystal to be exposed to substrate gradually in the hope of reducing any damage to the integrity of the lattice caused by the rapid ligand binding which occurred in the soaking experiments. The reduced temperature served to protect the crystal from radiation damage (10) and to slow the decomposition of the substrate (11).

To monitor substrate binding, several reflections were measured at 5 to 10 minute intervals after the flow of substrate solution was begun (12). These reflections were chosen on the basis of their sensitivity to heavy atom substitution in the active site region. After approximately 80 minutes of substrate flow, a large and rapid change in the quality of the diffraction pattern was observed. The diffraction pattern became weaker and the mosaic spread of the reflections tripled. The mosaicity increased anisotropically from an average full peak width of 0.5° in ω for reflections at 4 \AA to 1.5 to 1.8° in ω . Because of this change in the shape of the peaks, intensity measurements were made by scanning over the entire peak and subtracting the individual

background counts measured on each side of it. Data collection statistics are summarized in Table 1.

Data were collected to 3.5 Å resolution and a difference Fourier was calculated using the phases from the multiple isomorphous replacement analysis of the free enzyme. Several sections of this map are shown in Figure 4. The peaks and holes in the map fall predominantly within the boundaries of the molecule, and five features dominate the difference density. A large hole corresponding to the average position of residues 168-177 in the subunits of adjacent molecules in the crystal is shown in Figure 5a. These residues were disordered in the free enzyme. Strong peaks that are too large to be accounted for by the binding of substrate alone are apparent in both active sites of the dimer (Figure 5b). These positive features are somewhat different in shape, but in the main, they are consistently related by the non-crystallographic two-fold rotation axis of the dimer. In each peak, thin fingers of difference density connect with the boundaries of the disordered loop.

A strong positive peak which has no apparent negative mate also occurs between residues 100 and 101 (Figure 5c) in one subunit only. To provide further information about the origin of this feature, the crystal was washed with substrate-free mother liquor at -5° C for 24 hours. At the end of this period, 3-dimensional data with a minimum interplanar spacing of 6.0 Å was collected. An electron density map reflecting the difference between the washed crystal and the free enzyme crystal was calculated. The features in the active site were reduced in

Table 1. Statistics for the data on the yeast TIM-DHAP complex from 60 to 3.0 Å resolution.

Number of measurements -- 5004 (some data missing due to interference by the yoke of the flow cell)

Number accepted -- 3627

Mean F -- 37.0

Mean isomorphous difference -- 4.33

Overall scale factor to the native data -- 1.38

Mean Fractional Isomorphous Change

<u>$4\sin^2\theta/\lambda^2$</u>	<u>Fractional Δ_{iso}</u>
0-0.003	9 %
0.009	8 %
0.016	9 %
0.022	12 %
0.028	15 %
0.034	15 %
0.041	13 %
0.047	11 %
0.053	13 %
0.059	14 %
Overall 60-3.5 Å	11.7%

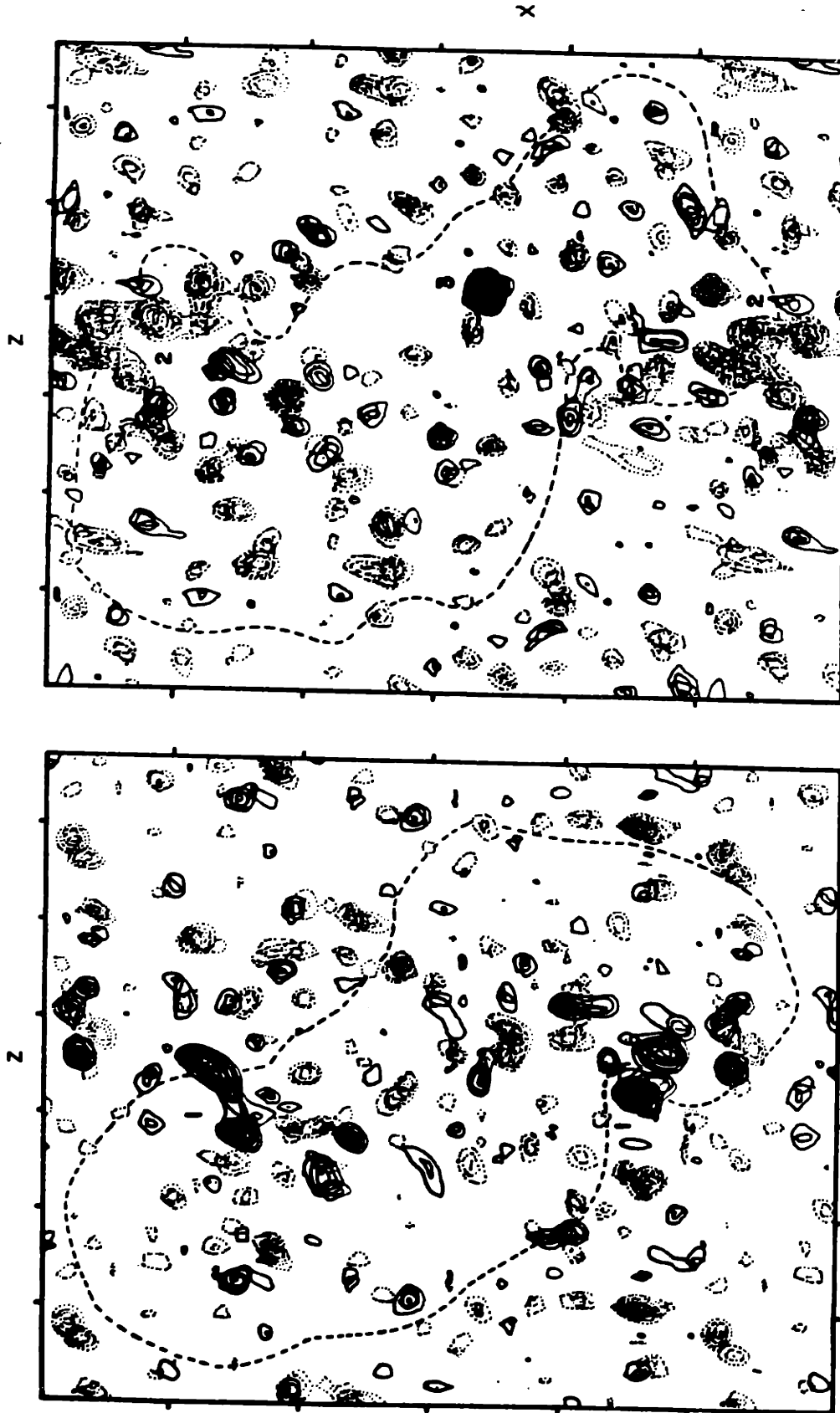
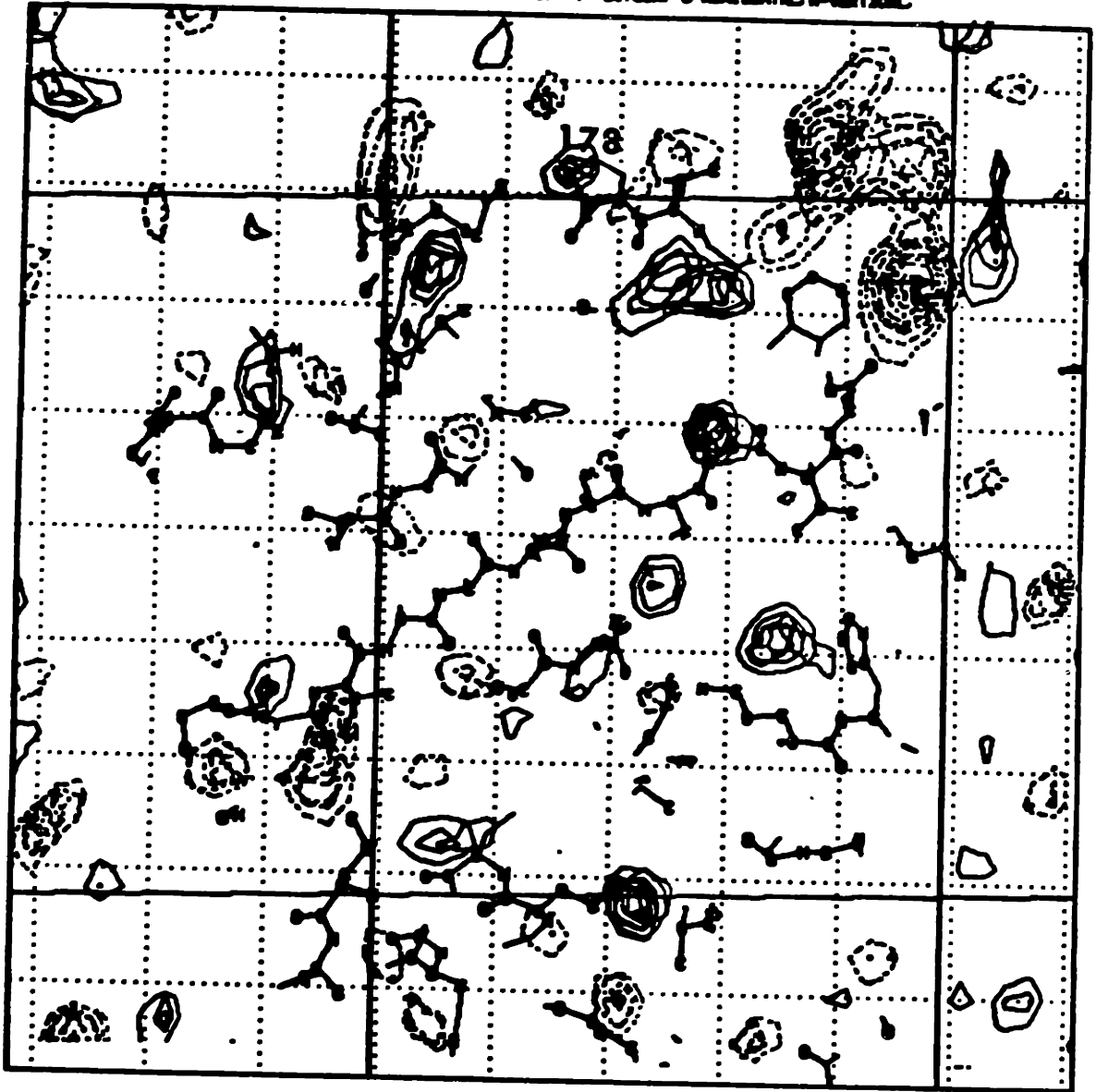


Figure 4. a) Difference map showing the changes in the TIM dimer on substrate binding. The molecular boundary is indicated by the dashed line. Both panels represent a 6 Å slice through the unit cell. a) Sections $y=0.38 - 0.44$. The prominent positive features marked "1" include density for the substrate and the loop of residues 168 to 177. b) Sections $y=0.45-0.51$. The negative features (dotted contours) marked "2" show holes in the average position of residues 168 to 177 in the free enzyme. The loops from adjacent molecules make contact in the free enzyme crystals. The positive feature marked "3" appears in only the lower subunit at the C-terminus of the D_1 -helix. b)



5a)

Figure 5. Difference map showing the binding of substrate in the active site of yeast TIM. The density added on formation of the enzyme-substrate complex is shown in solid contours; electron density present in the native enzyme but absent in the enzyme-substrate complex is shown in broken contours. The coordinates of the chicken enzyme were rotated into the yeast unit cell and are shown superimposed on the difference map.

5a) The large negative feature (top right) corresponds to the position of two loops of chain (residues 168 to 177) in adjacent molecules. The new position of one of these loop is shown in 5b. $y = 51.1 \text{ \AA}$.

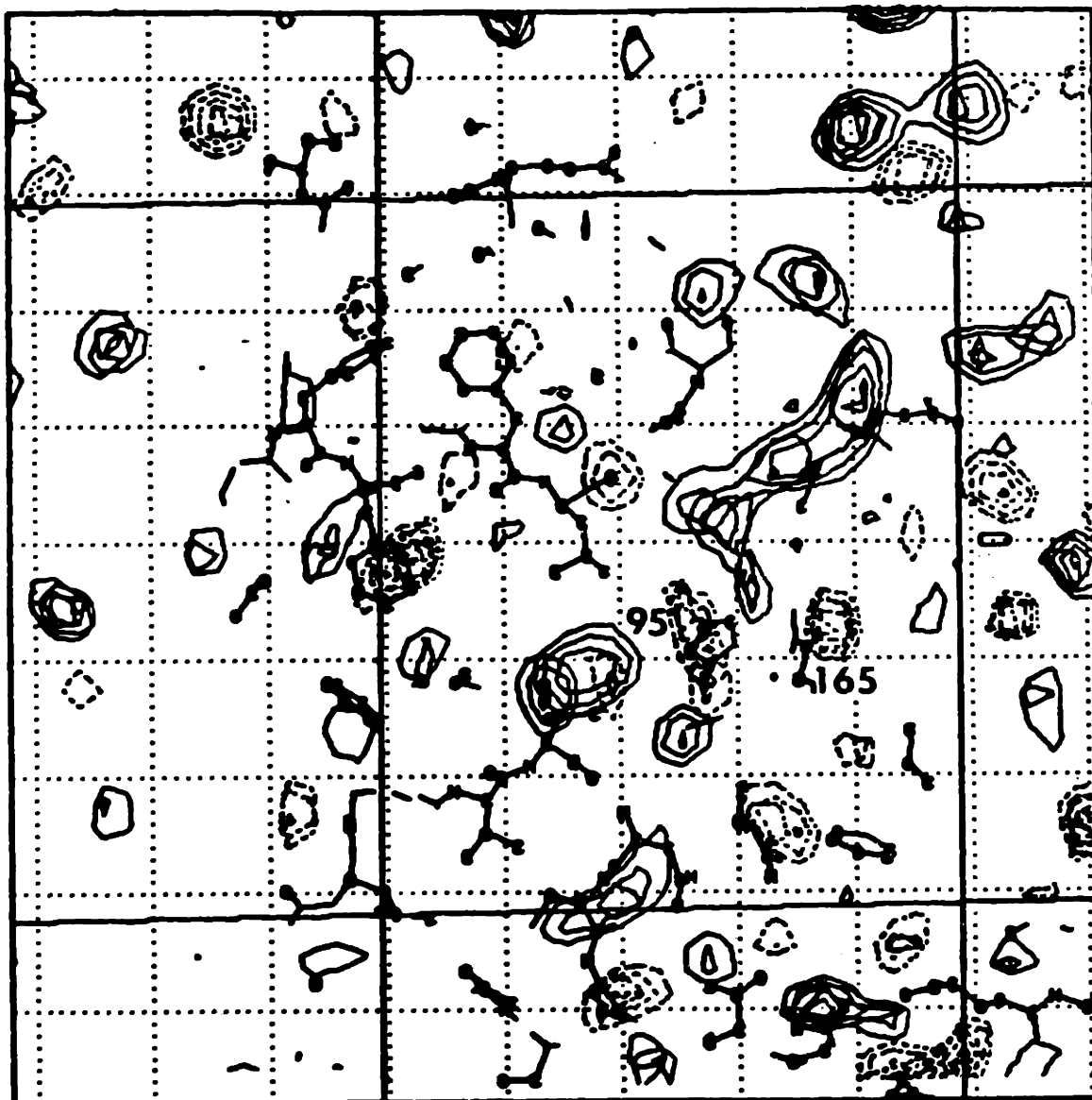


Figure 5b

The peak on the left-hand side of the largest positive feature can be interpreted in terms of substrate, which extends into the plane of the sections. The substrate is bound by the loop of residues 168-177 represented by the density on the right side of the feature. Parts of the side chains of His 95 and Glu 165 are shown below the substrate electron density, to the left and right, respectively. $\gamma = 44.2 \text{ \AA}$.

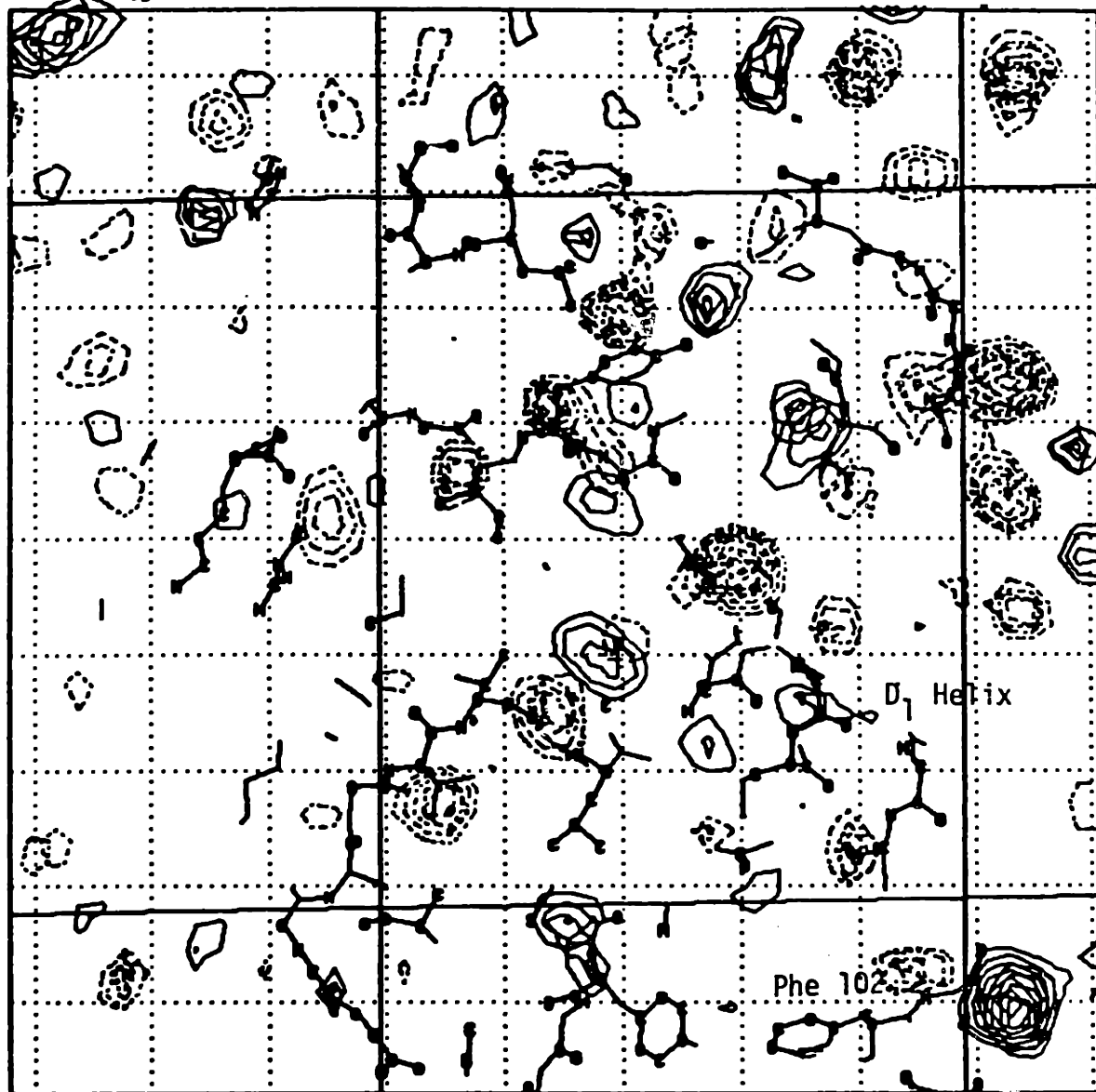


Figure 5c)

The positive feature at the C-terminus of the D₁ helix is shown at the lower right. Atoms in the helix appear above and to the left of the peak. Phe 102 is to its immediate left. Y = 47.1 Å.

intensity by the substrate-free wash, but the intensity of the peak near residues 100 and 101 remained strong.

There are many less dramatic features in the 3.5 Å difference map showing that subtle conformational changes occur throughout the molecule on substrate binding. Several of these features occur at the positions of the active site residues. However, without higher resolution data, better phases and a more complete interpretation of the structure of the free enzyme, these features are difficult to interpret.

IV. Discussion

A. The TIM-DHAP Complex

The studies of ligand binding to crystals of yeast TIM are consistent with previously reported results indicating that the enzyme undergoes significant conformational fluctuations during the catalytic cycle (1-4). The difference Fourier between the enzyme-substrate complex and the free enzyme shows that, in contrast to the results obtained with the crystalline chicken muscle enzyme (4), substrate binds to both active sites of the dimer in the crystals of yeast TIM. However, the changes seen in yeast TIM at 3.5 Å resolution are analogous to the features found at 6 Å resolution in the single subunit of chicken TIM where DHAP bound. These results suggest that the failure of substrate to bind in one subunit of the crystalline chicken enzyme is an

artifact imposed by the crystal lattice and that binding in the other subunit is normal.

The difference density is consistent with the movement of residues 168-177 to interact with the phosphate moiety of the substrate, apparently through contacts made by Gly 171, Thr 172 and Gly 173. Two similar tripeptides (gly-ser-gly) have been implicated in the interaction of yeast hexokinase with the phosphate moieties of ATP. (13). The C_{α} of Thr 172 changes its average position by about 10.4 Å on substrate binding (Figure 6).

Both the electron density in the Fourier map of the free enzyme and the holes in the difference map corresponding to the positions of residues 168 to 177 are weak and discontinuous. These results suggest that this loop of chain adopts a number of different conformations in the crystals of the free enzyme. In contrast, the positive density for the loop is stronger in the difference map, indicating that the number of conformations it can adopt is restricted in the enzyme-substrate complex.

Thus, substrate binding results in a change in the average position of the loop and in a reduction of the number of conformations accessible to it. These properties of the loop were also observed at 6 Å resolution on substrate binding to the chicken muscle enzyme, despite differences in the intermolecular contacts in the crystals (14). The importance of this behavior is hinted at by the high level of conservation of residues 168 to 177 among the sequences of triose

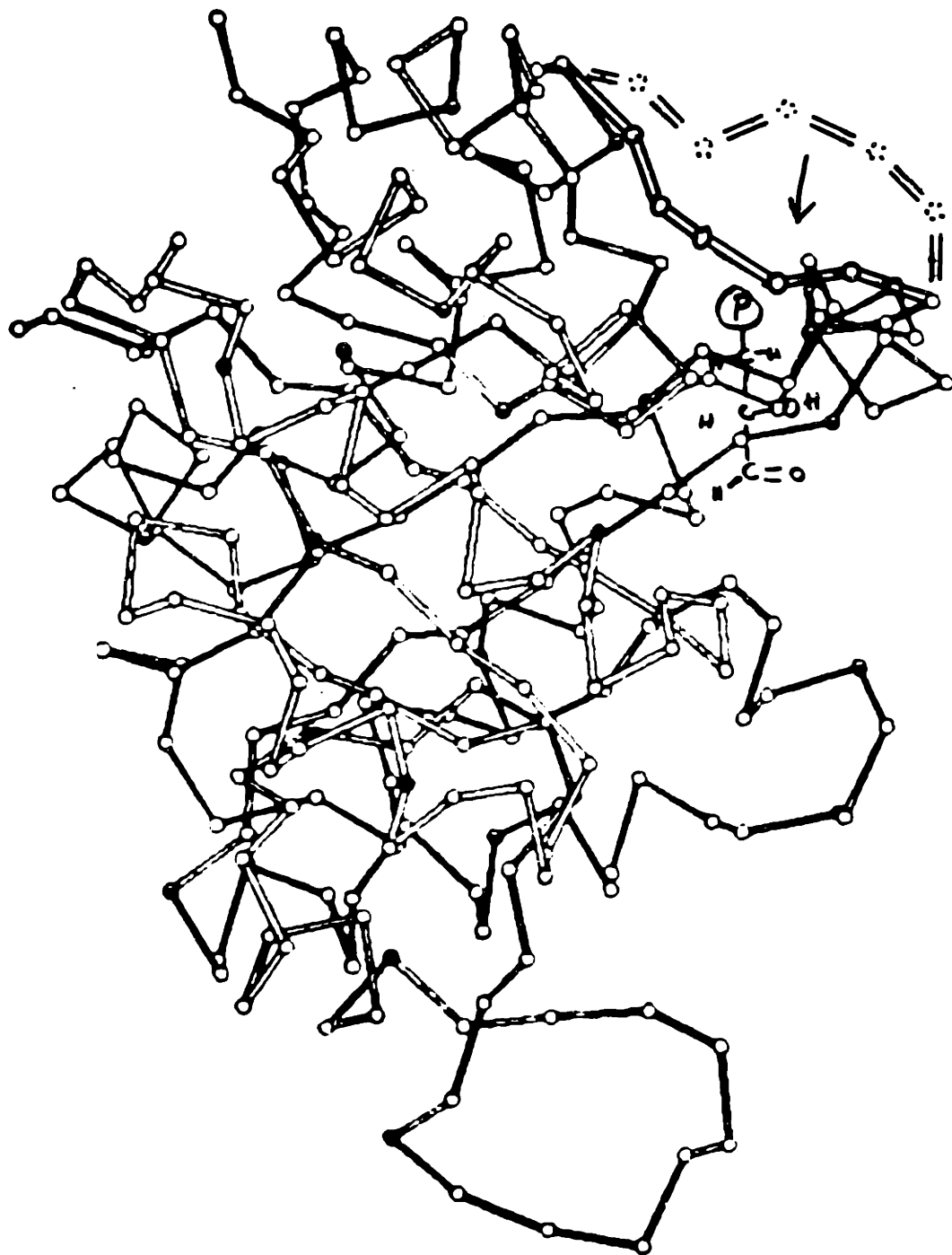


Figure 6. The change in the average position and degree of order of residues 168-177 on substrate binding. A drawing of the α -carbons of one monomer is shown, oriented in roughly the view depicted in the difference electron density maps in figures 4 and 5. The disordered loop is at the upper right. C_{α} of Thr 172 moves approximately 10.4 Å on substrate binding.

phosphate isomerases from yeast, rabbit, chicken, coelacanth and B. stearothermophilus (see Chapter II). The mechanistic implications of the observed changes in the position and degree of order of the loop are discussed in detail in Chapter V and Appendix I.

It is tempting to speculate that the change in the conformation of the loop on substrate binding is responsible for the disordering of the crystals. However, given the disorder of the loop in the free enzyme, this view may be somewhat simplistic. It may be possible to stabilize the crystals with a higher concentration of PEG or by light crosslinking with glutaraldehyde. Reduction of the increase in mosaic spread that accompanies DHAP binding would make it possible to collect high resolution data on the enzyme-substrate complex using the free enzyme crystal system.

The density corresponding to the substrate in both active sites of yeast TIM was generally consistent with both the position and shape of the added density observed at 6 Å resolution on the binding of substrate to one subunit of chicken TIM (14). Figure 7 shows DHAP built into the active site of chicken TIM and illustrates the principle charges and dipoles which are present (D.C. Phillips, personal communication). In addition to its interaction with the disordered loop, the phosphate moiety is bound by the sequence Gly-Gly-Ser at residues 209 to 211 and by the amide nitrogens of Gly's 232 and 233. The latter two glycines are at the positive end of the dipole moment of the C-terminal helix. The importance of the interaction between the phosphate and the helix dipole was first suggested by Hol et al (15). Indirect support for this

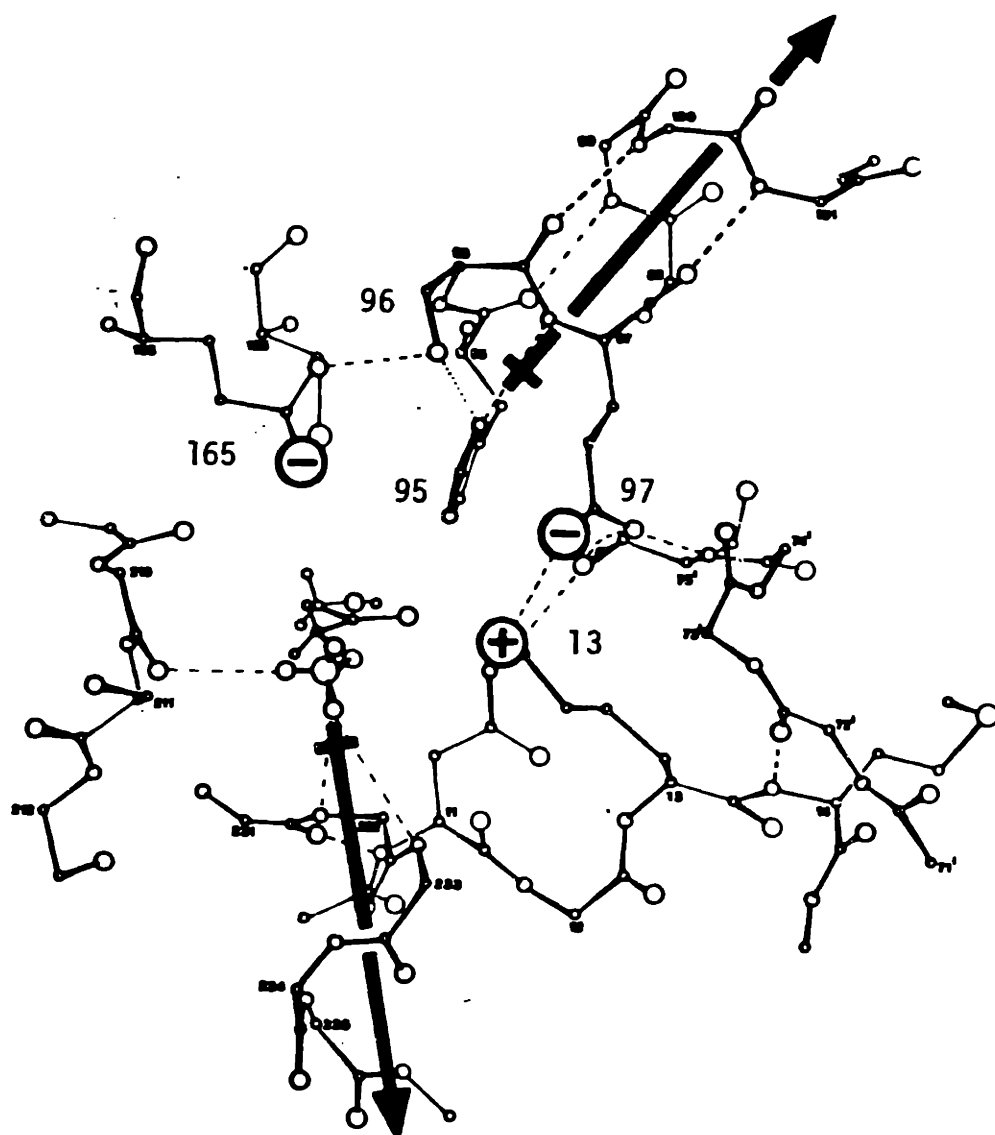


Figure 7. Model of the TIM-DHAP complex showing the principle charges and dipoles in the active site (D.C. Phillips and I.A. Wilson, unpublished results). The phosphate moiety is at the front, and residues 168 to 177 are omitted for clarity.

idea comes from the finding that the C-terminus of the B. stearothermophilus enzyme is two residues longer, perhaps to provide a stronger dipole moment to allow the binding to be sufficiently tight at higher temperatures (16).

A cylinder of density corresponding to the triose moiety extends into the active site pocket from the phosphate position in the difference maps obtained from both the yeast and chicken enzyme-substrate complexes (14). Though the shape of the density did not reveal the details of the conformation of the substrate, it is consistent with the substrate being fully extended. The assignment of the ends of the substrate is supported by the fact that the putative phosphate density is stronger than the density of the triose moiety in the yeast enzyme difference maps.

In the model of the enzyme-substrate complex shown in Figure 7, Glu 165 is in a position to shuttle protons between C-1 and C-2 and thereby act as the catalytic base. Its position is also consistent with its reactivity toward glycidol phosphate and haloacetyl phosphates (17,18). The carbonyl and hydroxyl groups of the bound DHAP are in a cisoid conformation and are oriented in the active site to permit the observed stereospecific attack of sodium borohydride from the si face of C-2 (19).

In this conformation, a number of interesting interactions between the enzyme and the oxygen atoms of the substrate are apparent. The contact distances in the yeast enzyme-substrate complex were measured

from a Fourier calculated at 3.5 Å resolution with the coefficients $3F_{\text{DHAP}}(\text{hkl}) - 2F_{\text{P}}(\text{hkl})$ and α_{P} . A Vector General graphics system was used to display the map and the atomic positions. These measurements are summarized in Table 2.

Though differences are apparent between the sets of contact distances derived from studies of the substrate complexes of the yeast and chicken enzyme crystals, the differences are not significant compared to the magnitude of the errors in the two models. In the chicken enzyme model, errors arise from the low resolution of the data, the absence of the phosphate moiety in the difference density (phosphate replaces the sulfate ion bound to the free enzyme) and the fact that the unrefined coordinates of the free enzyme were used to evaluate the distances. The yeast enzyme model, despite much higher resolution, suffers from a paucity of data due to crystal disorder and from poor phases which do not allow an unambiguous interpretation of the density in the active site.

Nonetheless, the following picture emerges. His 95 is in a position to interact with the carbonyl oxygens of the substrates and to mediate an interaction between these groups and the dipole moment of the helix containing residues 95-102. This dipole is shown in Figure 7. A precedent for such an interaction comes from the binding of DHAP to yeast aldolase. Nuclear magnetic resonance measurements suggest that the catalytic zinc atom is too far from bound DHAP to exert a direct catalytic effect (20). In aldolase, a histidine is in line between the zinc atom and the substrate and may polarize the carbonyl group of DHAP

	O-3		C-2 O		C-1 O	
	Chicken	Yeast	Chicken	Yeast	Chicken	Yeast
Lys 13, N _ε	3.3 Å	4.4 Å	2.9 Å	4.9 Å		
His 95, N _ε			3.5 Å	3.3 Å	3.2 Å	2.2 Å

Yeast

Glu 165, O_{ε2} to C-1 = 2.1 Å

Glu 165, O_{ε2} to C-2 = 2.2 Å.

Table 2. Contact distances in the models of the E-DHAP complex of yeast and chicken TIMs.

by hydrogen bonding (21). The D_1 helix dipole in yeast TIM may be a metal atom analogue, and the interaction of the substrate with the dipole may be facilitated by His 95. His 95 may be situated to interact in an analogous manner with a carbonyl group at C-1 (GAP), and, with a slight movement, C-2 (DHAP).

Lys 13 appears to interact with the C-2 oxygen and/or the bridging oxygen of the phosphate ester. It is held in position by participating in an ion pair with Glu 97 and also by the interaction of the amide nitrogen of Leu 14 with residues from the other monomer. The lysine side chain is in a fully extended conformation, stretching into a cleft formed by residues 72' to 74' in the intersubunit loop.

The binding of the substrate in the active site and the resulting movement of residues 168-177 appears to be slowly reversible in the yeast TIM crystals. In contrast, reduction of the intensity of the peak in the difference map near residues 100 and 101 was not apparent at 6 Å resolution after washing the crystal with substrate-free mother liquor for 24 hours. A similar feature was also observed in one subunit in the difference map of the chicken TIM-DHAP complex. Since it is not paired with a comparable hole in the difference map, this feature is probably not due to a simple conformational change of the protein. There are, however, a number of possible explanations of its origin. 1) The peak may be due to covalent modification of the protein by methyl glyoxal, a decomposition product of the enzyme-bound intermediate. The stoichiometry of its binding may reflect the fact that residues 100 and 101 are at the C-terminus of the D_1 helix of each subunit and are

adjacent in the dimer. 2) The positive feature could result from an increase in the degree of order of residues 100 and 101. The D₁ helix is involved in polarizing His 95, and substrate binding could induce changes far from the active site. This possibility is made less likely by the apparent irreversibility of the observed change. 3) Formally, the peak near residues 100 and 101 could also be a tightly bound molecule of GAP, although this is unlikely because of the known stoichiometry of binding of DHAP and substrate analogues (17,18,22). 4) Since the same feature was observed in both the yeast and chicken enzyme crystals, it is unlikely that it is an artifact of poor phasing.

The studies on the E-S complex at 6.0 Å resolution with chicken TIM and 3.5 Å resolution with yeast TIM are not accurate enough to distinguish between several hypotheses about the enzyme mechanism on structural grounds. A number of hypotheses which are consistent with the structural and solution studies are discussed in detail in the next chapter.

B. The TIM-PGA Complex

Phillips et al have reported that PGA also induces a conformational change involving the movement of residues 168-177 on binding to TIM (4). Consistent with this result, I have found that crystals of the free yeast enzyme dissolve in sulfate-free mother liquor containing even low concentrations of this ligand. Cocrystals of the yeast TIM-PGA complex have different unit cell dimensions than the free enzyme crystals and probably represent a different conformation of the enzyme. Since the

free enzyme crystals crack and break (instead of dissolving) when exposed to sufficiently high concentrations of DHAP, there may also be differences in the protein conformation in the two complexes.

Based on the pH dependence of the ^{31}P nmr spectrum of the complex of chicken TIM and PGA and the low resolution difference map of the complex obtained by Phillips and coworkers, Waley et al have suggested a model for the binding of this ligand (1,4,23). They proposed that PGA is bound in an extended conformation with the phosphate moiety in the phosphate binding site in the active center. They suggested that the carboxyl group shares a proton with O_e of Glu 165. This conformation is not consistent with the model for substrate binding discussed above. In particular, the proposed interaction of the two carboxylate oxygens through a shared proton does not parallel the interaction of the C-2 carbonyl group of the substrate with electrophilic groups on the opposite side of the active site. Consequently, if both models prove to be correct, PGA would not be a transition state analogue. Instead, its binding may mimick a nonproductive enzyme-substrate interaction.

The crystals of the yeast TIM-PGA complex will provide a direct test of the hypothesis of Waley and coworkers. In addition, it may be possible to replace the PGA with DHAP in a flow cell without disordering the crystals. Such an experiment may provide a way to obtain high resolution data on the enzyme-substrate complex.

References

1. Jones, R.B. and Waley, S.G. (1979), *Biochem. J.*, 179, 623-630.
2. Johnson, L.N. and Wolfenden, R.G. (1970), *J. Mol. Biol.*, 47, 93-
3. Banner, D.W., Bloomer, A.C., Petsko, G.A., and Phillips, D.C. (1971), *Cold Spring Harbor Symp. Quant. Biol.*, 36, 151-155.
4. Phillips, D.C., Rivers, P.S., Sternberg, M.J.E., Thornton, J.M. and Wilson, I.A. (1977), *Biochem. Soc. Trans.*, 5, 642-647.
5. Wyckoff, H.W., Doscher, M., Tsernoglou, D., Inagami, T., Johnson, L.N., Hardman, K.D., Allewell, N.M., Kelly, D.M. and Richards, F.M., (1967), *J. Mol. Biol.*, 27, 563.
6. Bragg, W.L., James, R.W. and Bosanquet, C.M. (1921), *Phil. Mag.*, 42, 1.
7. Matthews, B.W. and Czerwinski, E.W. (1975), *Acta Cryst. A*, 31, 480-487.
8. Jones, T.A. (1978), *J. Appl. Cryst.*, 11, 268-272.
9. Marsh, D.J. and Petsko, G.A. (1973), *J. Appl. Cryst.*, 6, 76-80.
10. Petsko, G.A. (1975), *J. Mol. Biol.*, 96, 381-392.
11. Iyengar and Rose (1981), I.A., *Biochemistry*, in press.
12. Alber, T., Petkso, G.A., and Tsneroglou, D. (1976), *Nature*, 263, 297-300.
13. Shoham, M. and Steitz, T.A. (1980), *J. Mol. Biol*, 140, 1-14.
14. Alber, T., Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., Rivers, P.S. and Wilson, I.A. (1981), *Phil. Trans. Roy. Soc. B*, 293, 159-171.
15. Hol, W.G.J., van Diujnen, P.T. and Berendsen, H.J.C. (1978), *Nature*, 273, 443-446.
16. Artavanis-Tsakonas, S. and Harris, J.I. (1980), *Eur. J. Biochem.*, 108, 599-611.
17. Hartman, F.C. and Gracey, R.W. (1973), *Biochem. Biophys. Res. Comm.*, 52, 388.
18. Schray, K.J., O'Connell, G.L. and Rose, I.A. (1973), *J. Biol. Chem.*, 248, 2214.

19. Webb, M.R., Strandring, D.N. and Knowles, J.R. (1977), *Biochemistry*, 16, 2738-2741.
20. Smith, G.M., Mildvan, A.S. and Harper, E.T. (1980), *Biochemistry*, 19, 1248.
21. Smith, G.M. and Mildvan, A.S. (1981), *Biochemistry*, 20, 4340-4346.
22. Noltmann, E.A., in The Enzymes, Vol. VI, ed. P.D. Boyer, Academic Press, New York, 1972, 271.
23. Campbell, I.D., Jones, R.B., Kiener, P.A. and Waley, S.G. (1979), *Biochem. J.*, 179, 607-621.

CHAPTER V.

Conclusions

I. Introduction

The interconversion of dihydroxyacetone phosphate and D-glyceraldehyde-3-phosphate by triose phosphate isomerase is one of the simplest and fastest metabolic reactions. Kinetic and thermodynamic studies are consistent with the existence of three enzyme-bound species on the reaction pathway -- the two substrate complexes and a bound intermediate whose structure is controversial (1,2). The intermediate may be a cis-enediol or symmetrical enediolate (3). It has a marked tendency to decompose in water to form methyl glyoxal and inorganic phosphate, and triose phosphate isomerase reduces the rate of this fragmentation reaction by over four orders of magnitude (4).

The reaction proceeds via a reversible proton transfer between C-2 of GAP and pro-R position of C-1 of DHAP (3,5,6). Isotope exchange studies suggest the presence of a single base on the enzyme which facilitates this step (7). Affinity labelling studies suggested the presence of an essential carboxylate in the active site, Glu 165, which is probably the catalytic base (8-13). Both affinity labelling studies and measurements of the susceptibility of DHAP to reduction by sodium

borohydride suggest the presence of at least one electrophilic group that enhances the lability of the proton by polarizing the carbonyl group of the substrate (11-14). Fourier transform infrared studies of the enzyme-substrate complex provide direct evidence for substrate distortion by the electrophile(s) on the enzyme (15).

The second order rate constant, k_{cat}/K_m , is $3.6 \times 10^8 \text{ s}^{-1}\text{M}^{-1}$ for the formation of DHAP and $1.0 \times 10^7 \text{ s}^{-1}\text{M}^{-1}$ for the formation of GAP. The rate constant for the formation of DHAP is comparable to that expected for a bimolecular diffusion-controlled process (16). Analysis of the microscopic rate constants for the formation and breakdown of the enzyme-bound species on the reaction pathway indicates that the rate limiting transition state is encountered in the binding and release of GAP (1). Thus, the chemical steps of the reaction are so fast that a physical step is fully or partially rate limiting (17). The enzyme enhances the reaction rate by more than nine orders of magnitude over the uncatalyzed reaction in water (18).

The x-ray crystal structure of triose phosphate isomerase from chicken muscle has been determined at 2.5 Å resolution (19). In addition, the conformations of the complexes of chicken TIM with a number of ligands have been studied crystallographically at 6.0 Å resolution (20,21). This thesis has described studies on the nucleotide sequence of the yeast TIM gene, the structure of the protein encoded by that gene at 3.0 Å resolution and the structure of the enzyme-substrate complex at 3.5 Å resolution.

These studies have yielded the following major results. The amino acid sequences of the yeast and chicken muscle enzymes are 50% homologous, and the two polypeptides fold into very similar three dimensional structures. Substrates bind to both subunits of crystalline yeast TIM in the same orientation as they bound to one subunit in the chicken TIM crystals. Consequently, the failure of substrates to occupy both active sites of the chicken enzyme dimer is probably an artifact caused by the intermolecular contacts in the crystal and not a previously undiscovered property of the free protein. The conformational changes that occur in the two proteins on substrate binding are also analogous, though, again, rearrangements can be seen in both subunits of yeast enzyme. In particular, there is a large change in the average position of a loop composed of residues 168 to 177, and the number of conformations accessible to the loop is decreased on substrate binding. There are numerous conformational adjustments throughout both subunits, and, in one subunit, a strong positive feature whose origin is not understood is seen in the difference maps. The feasibility of high resolution studies of enzyme-ligand complexes of yeast TIM has been established.

In this chapter, the results of the structural studies on the enzyme-substrate complex are analyzed in light of the properties of the enzyme catalyzed reaction. Features in the active site region which may account for the specificity and rate enhancing ability of TIM are discussed.

In the absence of high resolution data on enzyme-ligand complexes, the implications of conformation fluctuations away from the active site will not be treated. In addition, since the average error expected in the atomic positions derived from the present structural studies is on the order of an angstrom or more, no arguments can be made that hinge on more precise knowledge of the atomic coordinates.

There are several hypotheses about the structural basis of the properties of the enzyme that are consistent with the work described in this thesis. Some of these ideas are probably not distinguishable by crystallographic methods. Several avenues of experimentation are proposed which may shed light on a number of the central issues which remain unresolved.

II. Rate Enhancement

A productive mode of DHAP binding in the active site of TIM which is consistent with solution and structural studies is illustrated in Figure 1. The substrate is in an extended conformation with the C-1 and C-2 oxygens in a cisoid conformation. The bridging oxygen of the phosphate ester is in the plane of the putative enediol(ate) intermediate.

A. The Catalytic Base

The pK_a of the modification of Glu 165 by haloacetyl sulfate is 3.9 (22). This pK_a is low enough to allow Glu 165 to be catalytically

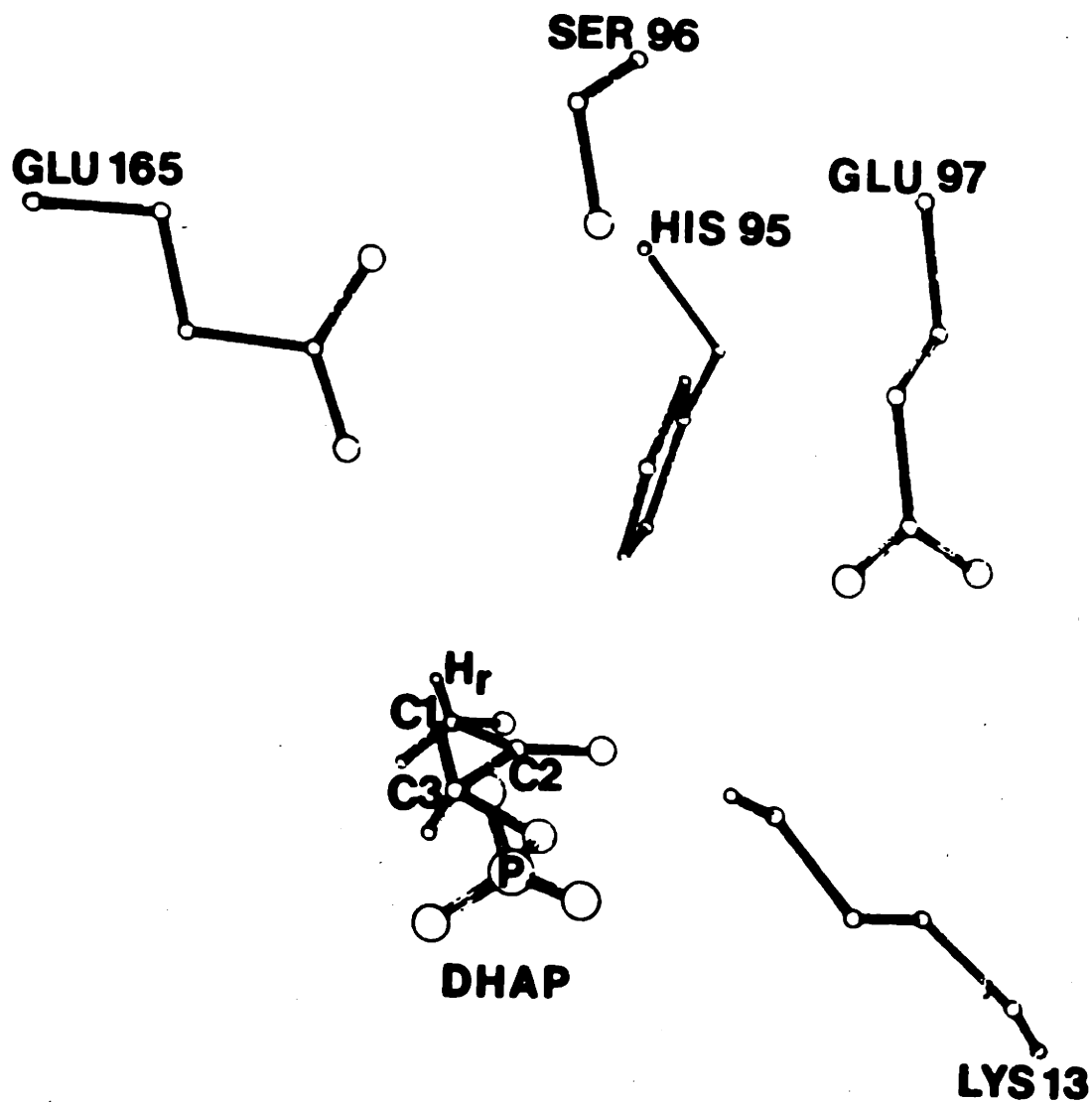


Figure 1. Relative orientations of the substrate and the side chains in the active site in the model of the TIM-DHAP complex.

competent as the active site base. The substrate can be oriented such that this residue can interact with either C-1 or C-2 from the re-re face of the subsequent cis-enediol(ate) intermediate, and this is sufficient to account for the observed stereospecificity of proton transfer (23).

Knowles and coworkers have shown that the catalytic base in the enzyme-intermediate complex is in rapid protonic equilibrium with either the bulk solvent or a pool of water molecules in the active site (24). Though the enzyme-substrate complex may not provide an adequate model of the exchanging species it is clear that the movement of residues 168 to 177 on substrate binding restricts the accessibility of the carboxylate of Glu 165 to the bulk solvent. However, measuring the extent of this restriction and locating the water molecules in the active site will require higher resolution structural studies.

B. The Catalytic Electrophiles

Though electrophilic groups on the enzyme have been implicated in the rate enhancement provided by TIM, it is not clear whether acidic or electrostatic effects are involved. This uncertainty is linked to the uncertainty about the structure of the enzyme-bound intermediate, although, as discussed below, the questions are formally separable.

Experiments which bear on this problem are indirect. For example, Hegarty and Jencks have shown that carboxylic acids catalyze the enolization of acetone in a bifunctional manner (25). The reaction is

apparently third order, involving two concerted proton transfers. However, not all acid/base pairs demonstrate such bifunctional catalysis.

Rose and coworkers have attempted to show that the enzyme-bound intermediate is an enediol, but, as described in Chapter I, their results are inconclusive (4,26). This question is relevant, because either the enzyme or the solvent would have to donate a proton to the transition state to form an enediol. An obligate enediol on the reaction pathway would strengthen the case for general-acid/general-base catalysis by the enzyme electrophile(s).

Kollman has evaluated classical potential functions which describe the active site of chicken TIM (P. Kollman and D.C. Phillips, unpublished results). He has found that the only regions of positive potential center on the positions of C-1 and C-2 in the model of the TIM-DHAP complex. These calculations support the idea that the electrostatic potentials in the active site can destabilize the ground state of the substrate.

An examination of the structure of the enzyme-substrate complex allows the problem to be formulated in exquisite detail. Its resolution, however, will require further studies.

The most likely candidates for the active site electrophiles are His 95 and Lys 13 (Figure 1). In a general-acid/general base mechanism, these groups may act as proton donors and acceptors, while in an

electrostatic mechanism, they would stabilize negative charge in the transition states through charge-charge interactions. A third possibility is that electrostatic polarization allows solvent water to act as an acid/base to facilitate the formation and breakdown of the putative enediol intermediate (20).

His 95 is in a position to interact with the C-1 oxygen and possibly the C-2 oxygen of bound substrates. It is hydrogen bonded to the main chain NH of Glu 97 at the positive end of the dipole moment of a helix containing residues 95-102 (D_1). The N_δ of the imidazole ring is the proton acceptor in this hydrogen bond (Figure 2). The N_ϵ proton of His 95 might be expected to carry an enhanced partial positive charge due to the polarization of the ring by the helix dipole. This could, in turn, polarize the carbonyl group of the substrate. The acidity of the proton would also be enhanced, but it is likely that protonation of the substrate would result in the formation of an imidazolate anion. Compensating protonation of N_δ of the imidazolate is unlikely, because it would require a reorientation of the ring and concomitant disruption of the hydrogen bond between N_δ and the amide nitrogen of Glu 97. On the basis of indirect assignments of the chemical shifts in the pH dependent p.m.r. spectrum of the free protein, Waley et al have suggested that this hydrogen bond is unusually strong (27). They argued that N_δ of His 95 is not protonated even at pH 4.0 in the free enzyme.

The pKa for the formation of the imidazolate form of histidine is 12.3 (28), though clearly, this may be reduced by the electrostatic environment of His 95. Developing negative charge on the substrate C-1

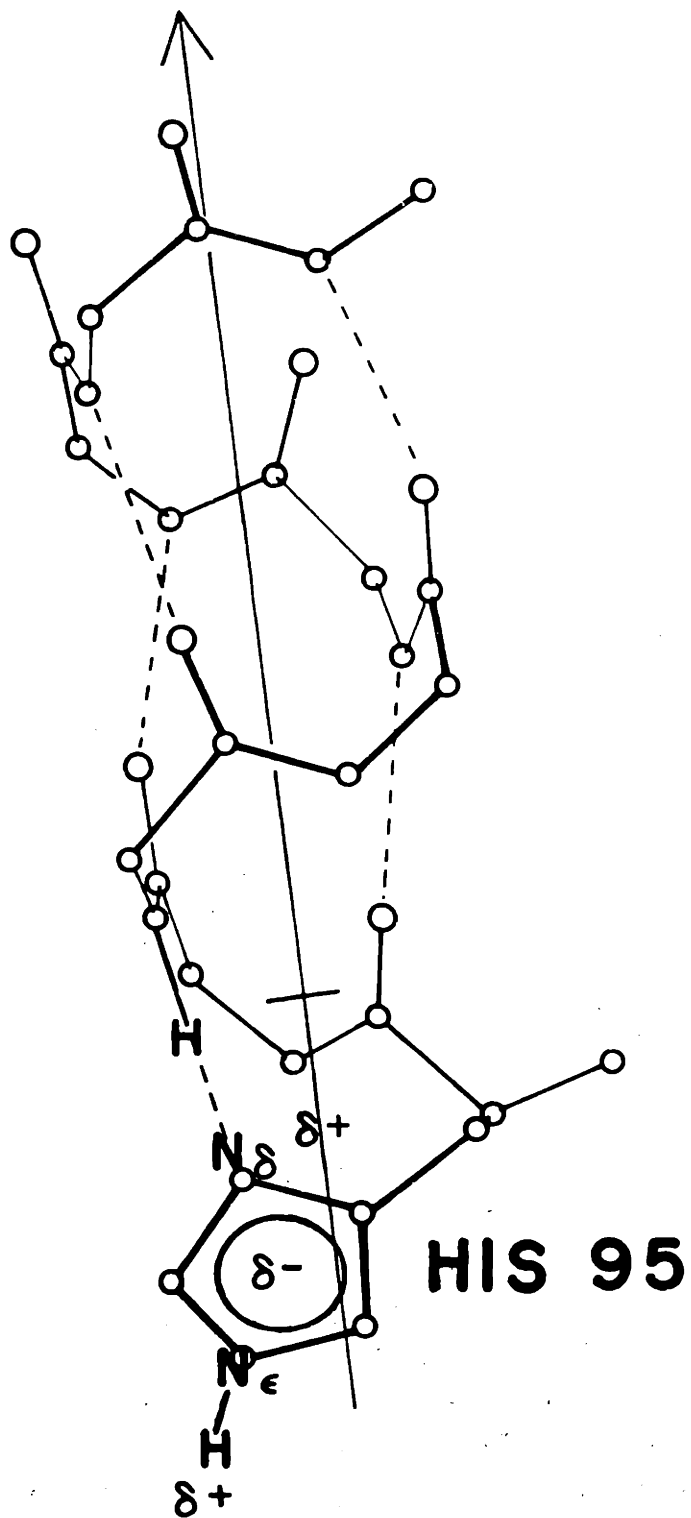


Figure 2. Histidine 95 may be polarized by the dipole moment of the D_1 helix.

oxygen, then, could be stabilized by a charge-charge interaction with the N_{ϵ} proton. Because of the partial electrostatic character of the N_{ϵ} -H bond, it is likely that this interaction would result in at least a partial donation of the proton to the substrate and a concomitant formation of an imidazolate anion (29). This species could be stabilized by the dipole moment of the D_1 helix. The extent of proton donation would be a function of the pKa's of the imidazolate and the oxygen of the triose phosphate in the transition state complex. The problem of whether His 95 acts as an acid or as an electrostatic group collapses into a question of the extent to which the N_{ϵ} proton is transferred to the substrate.

In the preceding argument, the assumption is made that the helix dipole can and does function as an electrophilic center. Hol et al have presented circumstantial evidence supporting this view for the reactions of papain, subtilisin, rhodanese and glyceraldehyde phosphate dehydrogenase (30). If this is the case, the binding of substrate should induce a change in the electrostatic potential of the helix that may be observable as a change in the positions or temperature factors of its atoms. An extreme case of such an induced change has been observed in the transition from the two-zinc to the four-zinc form of crystalline insulin (31). When the sodium chloride concentration of the mother liquor of insulin crystals is raised above 6%, an additional metal atom binds to each molecule. The zinc ion binds to the C-terminus of a surface helix and stabilizes the addition of several residues to the helix, presumably through an interaction with the negative end of its dipole moment. Though the effects of substrate binding to TIM are less

dramatic, they may be nonetheless observable through crystallographic studies at sufficiently high resolution ($\leq 2.0 \text{ \AA}$).

As discussed in the previous chapter, the dipole moment of the D_1 helix in TIM may function as an analogue of the catalytic zinc ion in yeast aldolase (32). This surprising finding suggests that, by virtue of their structural arrangements, organic molecules can mimic the properties of molecules containing more complex atoms. Thus, helices may substitute for the properties inorganic ions; and, indeed, the different dependences on distance and angular direction of ionic and dipolar potentials may allow each to be exploited to different ends in enzymatic reactions.

Theoretical studies help to define a possible role for the dipole moment of the D_1 helix. Using ab initio molecular orbital calculations, Valentine et al have shown that the strength of a hydrogen bond between an imidazole N-H and the carbonyl group of formamide is dramatically increased when the imidazole is liganded to a positively charged atom (33). Assuming a single charge on the positive center, hydrogen bond free energies of 10 to 20 kcal/mole were found. This would be more than sufficient to polarize a carbonyl group and stabilize an enolate intermediate or transition state. It is interesting to note that the measured free energy difference for the transition states of the uncatalyzed and TIM-catalyzed interconversion of dihydroxyacetone phosphate and glyceraldehyde-3-phosphate is on the order of 10 kcal/mole (34).

As expected, the hydrogen bond strengths calculated by Valentine and coworkers are very sensitive to the geometry of the interaction (33). Consequently, if His 95 is functioning as an electrophile in the TIM-catalyzed reaction, small differences in its orientation relative to the substrate oxygens might result in large changes in the interaction free energies. This sensitivity to direction would provide a mechanism for maximizing the free energy of interaction between the enzyme and the planar center of the substrate (C-1 O of GAP and C-2 O of DHAP).

It is not clear from the present model of the TIM-substrate complex that the D₁ helix can polarize both the C-1 and C-2 oxygens. A conformational change in either the entire helix or in the position of His 95 relative to the substrate may be required for the helix-dipole to act as the electrophile in the reaction of DHAP. The difference maps of both the chicken and yeast TIM-substrate complexes provide evidence for some movement of His 95 on substrate binding (Figure 3) (20). However, the present resolution of the structural studies does not allow an unambiguous interpretation of the difference density.

Alternatively, Lys 13 may act to polarize the C-2 carbonyl of DHAP and facilitate the abstraction of the pro-R proton of C-1. In theory, it could act as either an acid or an electrostatic center, but the structural studies do not favor the former possibility. The amino group of Lys 13 is unlikely to lose a proton to the transition state, because this would leave Glu 97 buried with an uncompensated negative charge. Consequently, the energetic cost of breaking the ionic bond between Lys

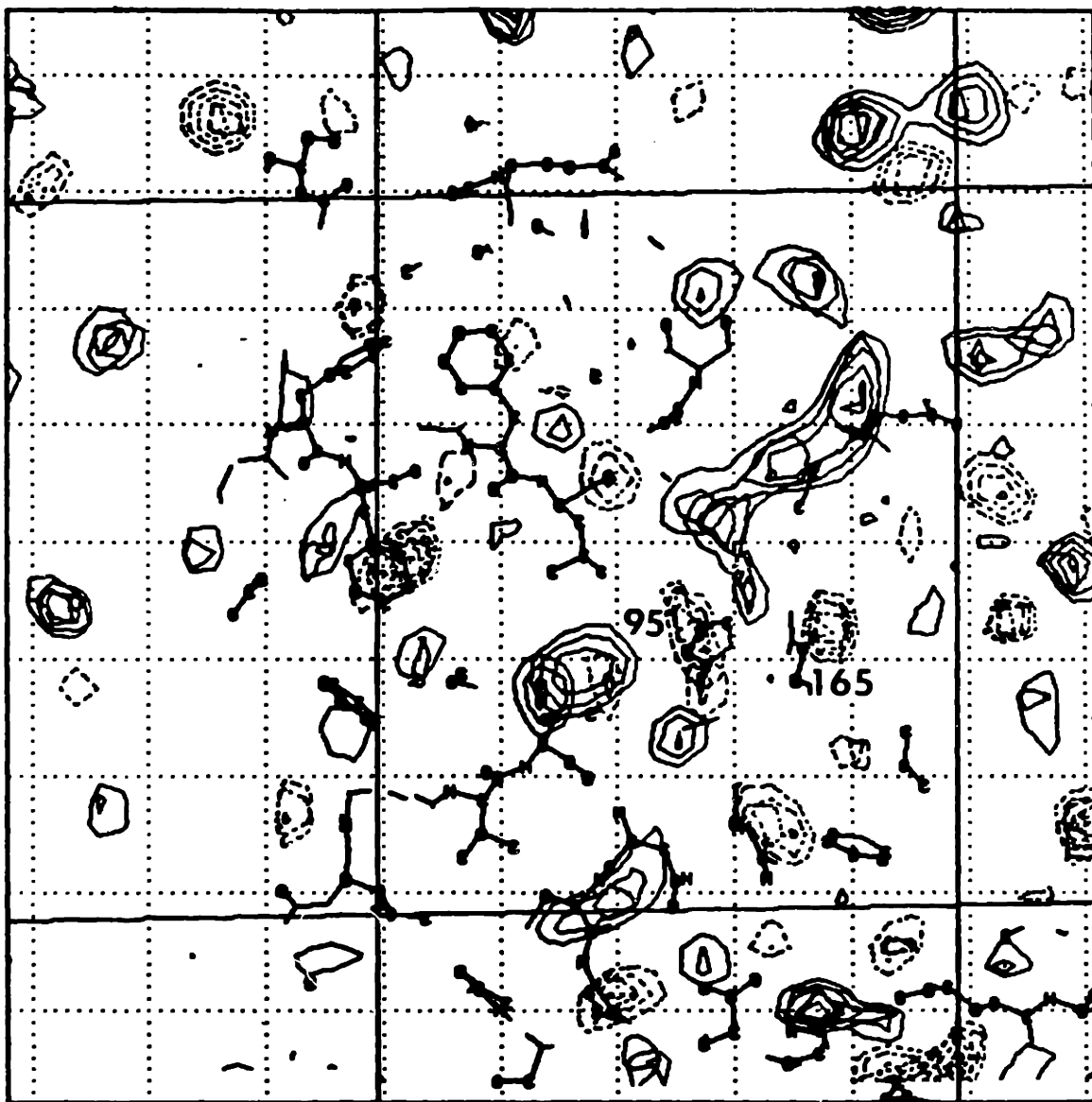


Figure 3. Difference map showing the binding of substrate in the active site of yeast TIM. As in Figure 5b in Chapter 4, the y section = 43.2. The coordinates of His 95 fall in a region of negative density whose presumed positive mate is lower in the map and closer to the C-2 oxygen of the substrate model.

13 and Glu 97 is high, and if Lys 13 acts as a catalytic electrophile, it is likely that its effects are primarily electrostatic.

In summary, the crystal structure of the enzyme-substrate complex is consistent with either His 95 or Lys 13, or both, acting as electrophiles to promote proton abstraction. They may function as electrostatic centers and/or proton donors.

C. Substrate Binding and Release

As described above, the chemical steps in the TIM reaction are so efficient that the binding and release of GAP is at least partially rate limiting (1). Substrate binding is accompanied by a change in the position and degree of order in a surface loop composed of residues 168-177. The change in the number of conformations accessible to this loop has at least two consequences for the energetics, and consequently the rate, of the formation and breakdown of the Michaelis complex.

First, the activation energy for binding is minimized, because the interactions between the enzyme and substrate do not require that a stable structure in the protein be disrupted. The substrate can stabilize one of a number of unstable conformations which are in equilibrium in the free enzyme.

Second, if the reduction in the number of conformational states of the loop on substrate binding is sufficiently large, some of the binding energy will be dissipated by the resulting decrease in the

conformational entropy of the polypeptide chain. The quantitative limits of this effect are derived in Appendix I. For example, a three-fold reduction in the average number of conformational states accessible to each residue in the loop on substrate binding would require 6 kcal/mole of binding energy. Thus, the disorder-order transition raises the free energy of the E-S complex relative to the free energy of the transition state for the formation of this complex and allows rapid release of the products. Were substrate bound by a fully ordered domain, it would be bound more tightly, and its release from the surface of the enzyme would be slower. As shown in Appendix I, the interaction of ligands with disordered domains may be a general mechanism by which a number of proteins -- including tobacco mosaic virus coat protein, acid repressors, peptide hormones, calmodulin, ribosomal proteins, histones and other nucleic acid binding proteins -- bind ligands both specifically and weakly.

In its simplest form, entropic destabilization would result in a variation of the free energy of protein-ligand complexes with temperature. In the reaction of triose phosphate isomerase this implies that the activation energy will be temperature dependent, and the logarithm of the overall rate constant describing the steady state conversion of glyceraldehyde-3-phosphate to dihydroxyacetone phosphate will not be a linear function of $1/T$. In addition, as the temperature is lowered, the kinetic isotope effect observed in the reverse reaction should decrease as the release of dihydroxyacetone phosphate becomes rate limiting (35). When the reaction is carried out in tritiated water, the extent of isotopic labelling in dihydroxyacetone phosphate

should increase at lower temperatures (35). These experiments would allow the role of the loop movement to be investigated.

D. Specificity

TIM is as specific a catalyst as it is efficient. The phosphonate analogues of the triose phosphates are extremely poor substrates (37), and the enzyme will catalyze α -proton exchange between monohydroxyacetone phosphate and the solvent (I. A. Rose, personal communication). The substrates bind productively as unhydrated dianions (38,39). Binding of dihydroxyacetone sulfate cannot be detected (37).

This specificity apparently reflects the small size of the active site cleft and the requirement that phosphate be present to establish the relative orientation of the catalytic groups on the enzyme and the labile proton of the substrates. The basis of the ability of TIM to distinguish between phosphate and sulfate esters is unknown. However, it is not unique to this protein or to proteins in general. The sulfate binding protein of Salmonella typhimurium does not bind phosphate, and, of course, protons have extremely different affinities for the two groups.

Part of the specificity of the reaction is kinetic. It arises from the ability of the enzyme to inhibit side reactions, including the dephosphorylation of the enzyme-bound intermediate. This fragmentation would be facilitated by an out-of-plane conformation of the bridging oxygen (40), and puts a premium on the ability of the enzyme to fix the

position of the phosphate moiety. The phosphate is bound by at least four sets of groups on the enzyme - residues 210 to 211, residues 232 to 233 at the positive end of the dipole moment of the C-terminal helix, Lys 13 (which may interact with the bridging oxygen) and residues 171 to 173 in the disordered loop. The need for the enzyme to enclose and release the phosphate as rapidly as possible may lie at the root of the presence and properties of the disordered loop. Complete solvation of the phosphate by the enzyme rather than by water may be required to establish proper binding geometry and to provide a suitable electrostatic environment for the isomerization.

E. The Future

In general, the crystallographic studies on triose phosphate isomerase have allowed questions about the structural origins of its catalytic power to be formulated in great detail. Higher resolution structural studies of the free enzyme and its complexes with substrates and inhibitors may shed light on many unsolved problems. These include the identification of the catalytic electrophile(s), the determination of the ratio of enzyme-bound species in the steady state, the determination of the conformation and orientation of each of the three enzyme-substrate species in the active site, and the quantitative assessment of the accessibility of solvent to the active site during the reaction. In addition, little is understood about the conformational fluctuations which occur away from the active site during catalysis. The cocrystals of TIM and phosphoglycolate which were described in

Chapter IV may provide the best crystal system for obtaining high resolution data on TIM-ligand complexes.

However, problems such as the extent to which a proton is donated to the central transition states by the catalytic electrophile(s) and the role of the disordered loop in the reaction are probably not directly accessible by x-ray crystallographic studies of the wild type enzyme. Instead, it may be possible to construct isostructural mutants which are systematically altered to test the roles of specific chemical groups. Using the cloned gene, TPI1, two classes of mutants could be made using site specific methods. The first class involves changes in the phosphate binding site, while the second includes proteins with altered catalytic residues. In this manner, both the binding and chemical steps could be probed.

For example, glycine residues could be added to the disordered loop to test the role of chain entropy in substrate binding. Unique sequences which are recognized by the restriction enzymes Kpn I, Asu I, and Hgi CI, and are cut in frame are found in the loop. These sites are obvious targets for mutagenesis.

Mutants which may yield information about the catalytic electrophile(s) would include any of the following changes: His 95 to tyr, asn, or gln; Lys 13 to gln or leu; and Glu 97 and Glu 165 to gln. The codons specifying His 95 and Glu 97 are accessible from the nearby Bgl II restriction site, but Lys 13 will probably have to be altered by using DNA polymerase to elongate a mismatched primer. Eliminating the

active site base may make it possible to study the roles of the electrophile(s) in the absence of turnover. The alteration of Lys 13 would test its ability to act as an electrophile or a hydrogen bond donor to the bridging oxygen of the phosphate ester.

Since 10 to 20% of the soluble protein in yeast strains transformed with a high-copy-number plasmid carrying the TPI1 gene is TIM, the purification of the mutant proteins should not be too difficult. Iyengar and Rose have suggested that it is possible to rapidly determine the ratio of enzyme bound species in the steady state (3). Thus, it should be possible to determine which step(s) in the catalytic reaction are affected by the mutations. Structural studies on the mutant proteins using x-ray crystallography or difference n.m.r. methods will be required to show that changes in the catalytic properties are due to differences in the chemistry of the altered residues rather than changes in the structure of the protein.

The cloned gene may also facilitate studies on the influence of any residue or group of residues on the folding pathway of the protein.

The studies described in this thesis have laid the groundwork for both genetic and crystallographic experiments which may further clarify the structural origins of the catalytic power of triose phosphate isomerase.

References

1. Albery, W.J. and Knowles, J.R. (1976), *Biochemistry*, 15, 5627-5631.
2. Iyengar, R. and Rose, I.A. (1981a) *Biochemistry*, in press.
3. Reider, S.V. and Rose, I.A. (1959), *J. Biol. Chem.*, 234, 1007-1010.
4. Iyengar, R. and Rose, I.A. (1981b), *Biochemistry*, in press.
5. Reider, S.V. and Rose, I.A. (1956), *Fed. Proc.*, 15, 337.
6. Bloom, B. and Topper, Y.J. (1956), *Science*, 124, 982.
7. Rose, I.A. (1962), *Brookhave Symp. Biol.*, 15, 293-309.
8. Hartman, F.C., (1971), *Biochemistry*, 10, 146-154.
9. Coulson, A.F.W. and Knowles, J.R. (1970), *Chem. Commun.*, 7-8.
10. De la Mare, S., Coulson, A.F.W., Knowles, J.R., Priddle, J.D. and Offord, R.E. (1972), *Biochem. J.*, 129, 321-331.
11. Waley, S.G., Miller, J.C., Rose, I.A. and O'Connell, E.L. (1970), *Nature*, 227, 181.
12. Hartman, F.C., LaMuraglia, G.M., Tomozawa, Y. and Wolfenden, R. (1975), *Biochemistry*, 14, 5274-5279.
13. Hartman, F.C. and Raric, H. (1977), *Biochem. Biophys. Res. Comm.*, 77, 746-752.
14. Webb, M.R. and Knowles, J.R. (1975), *Biochemistry*, 14, 4692-4698.
15. Belasco, J.G. and Knowles, J.R. (1980), *Biochemistry*, 19, 472-477.
16. Hammes, G.G. and Schiessel, P.R. (1970), *Enzymes*, 3rd edition, 67-114.
17. Albery, W.J. and Knowles, J.R. (1976), *Biochemistry*, 15, 5631-5640.
18. Hall, A. and Knowles, J.R. (1975), *Biochemistry*, 14, 4348-4352.
19. Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., Pogson, C.I., Wilson, I.A., Corran, P.H., Furth, A.J., Milman, J.D., Offord, R.E., Priddle, J.D. and Waley, S.G. (1975), *Nature*, 255, 609-614.
20. Alber, T., Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., Rivers, P.S. and Wilson, I.A. (1981), *Phil. Trans. Roy. Soc. B*, 293, 159-171.

21. Phillips, D.C., Rivers, P.S., Sternberg, M.J.E., Thornton, J.M. and Wilson, I.A. (1977), *Biochem. Soc. Trans*, 5, 642-647.
22. Hartman, F.C., LaMuraglia, G.M., Tomozawa, Y. and Wolfenden, R. (1975), *Biochemistry*, 14, 4692-4698.
23. Rose, I.A. (1975), *Adv. Enzymology*, 43, 491-517.
24. Fisher, L.M., Albery, W.J. and Knowles, J.R. (1976), *Biochemistry*, 15, 5621-5626.
25. Hegarty, A.F. and Jencks, W.P. (1975), *J. Amer. Chem. Soc.*, 97, 7188-7189.
26. Rose, I.A. (1981), *Phil. Trans. Roy. Soc. B.*, 293, 131-143.
27. Browne, C.A., Campbell, I.D., Kiener, P.A., Phillips, D.C., Waley, S.G. and Wilson, I.A. (1976), *J. Mol. Biol.*, 100, 319-343.
28. Handbook of Chemistry and Physics, 53rd edition, ed. R.C. Weast, The Chemical Rubber Co., Cleveland, 1972, C741.
29. Pauling, L., The Nature of the Chemical Bond and the Structure of Molecules and Crystals, 3rd edition, Cornell University Press, Ithaca, 1960.
30. Hol, W.G.J., van Duijnen, P.T. and Berendsen, H.J.C. (1978), *Nature*, 273, 443-446.
31. Bentley, G., Dodson, E., Dodson, G., Hodgkin, D. and Mercola, D. (1976) *Nature*, 261, 166-168.
32. Smith, G.M. and Mildvan, A.S. (1981), *Biochemistry*, 20, 4340-4346.
33. Valentine, J.S., Sheridan, R.P., Allen, L.C. and Kahn, P.C. (1979), *Proc. Nat. Acad. Sci. USA*, 76, 1009-1013.
34. Hall, A. and Knowles, J.R. (1975), *Biochemistry*, 14, 4348-4352.
35. Leadlay, P.F., Albery, W.J. and Knowles, J.R. (1976), *Biochemistry*, 15, 5607-5620.
36. Fletcher, S.J., Herlihy, J.M., Albery, W.J. and Knowles, J.R. (1976), *Biochemistry*, 15, 5612-5617.
37. Belasco, J.G., Herlihy, J.M. and Knowles, J.R. (1978), *Biochemistry*, 17, 2971-2978.
38. Trentham, D.R., McMurray, C.H. and Pogson, C.I. (1969), *Biochem. J.*, 114, 19-24.
39. Reynolds, S.J., Yakes, D.W. and Pogson, C.I. (1971), *Biochem. J.*, 122, 285-297.

40. Deslongchamps, P., Atlanti, P., Frebel, D., and Malaval, A. (1972),
Can. J. Chem., 50, 3405-3408.

Appendix I.

Energetic Consequences of Unfolded Regions in the Binding Sites of Proteins

ABSTRACT: A number of proteins contain flexible binding domains which become ordered during their association with ligands. A portion of the intrinsic binding energy must be dissipated to restrict the motions of the protein. This will lower the stability of protein-ligand complexes if the free energy needed to order the polypeptide chain is of the same order of magnitude as the binding energy. An examination of the statistical mechanics of chain folding suggests that the disorder-order transition can modulate binding energy.

Several proteins whose functions and properties may depend on the loss of flexibility in their binding sites are described. Ligand binding through conformationally unstable domains has at least two energetic consequences. First, it can allow specific interactions between macromolecules which have large intrinsic binding energies to be reversible. In addition, the loss of motional freedom in enzyme active sites can destabilize the complexes of enzymes with their substrates and products. In triose phosphate isomerase, this may contribute directly to the efficiency of the enzyme.

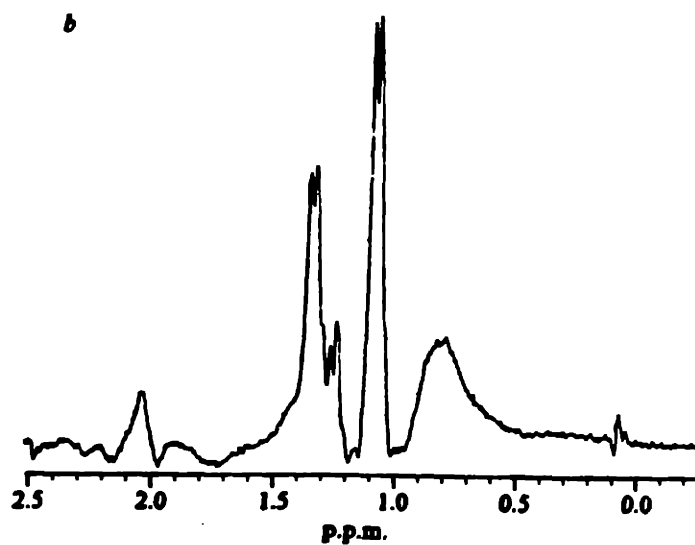
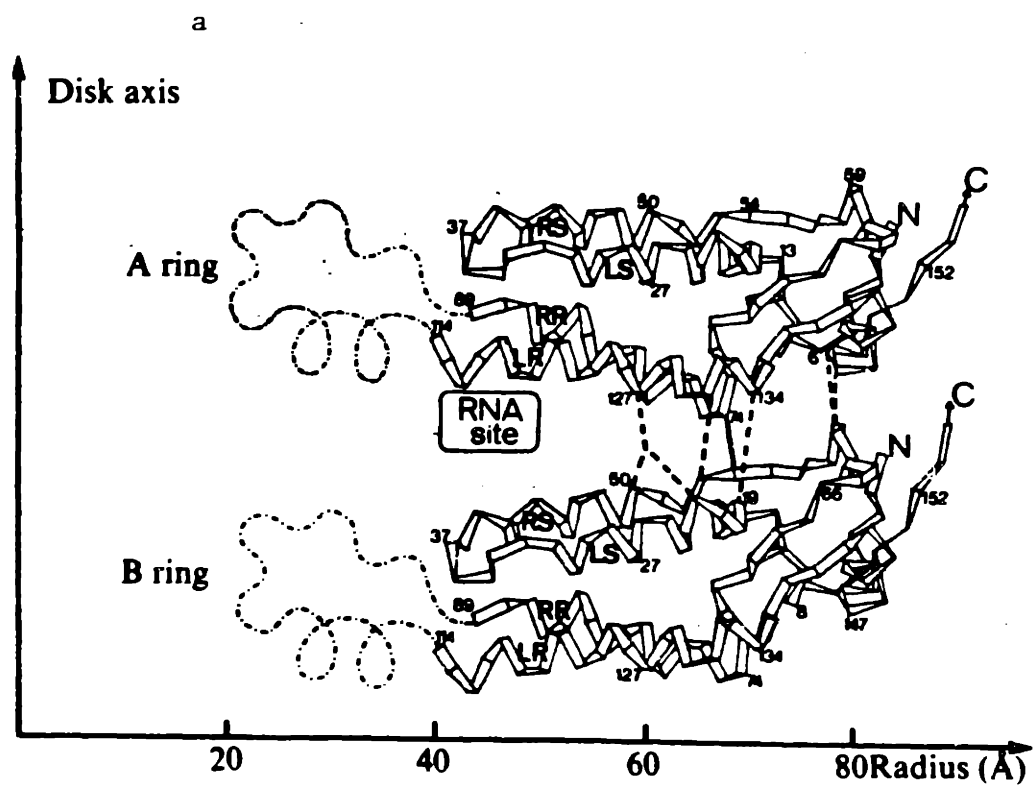
Introduction

There is growing evidence that a number of proteins contain binding sites which are conformationally unstable in the absence of ligands. Such "unfolded" regions provide examples of extreme flexibility in which the angular range of rotations and the spatial amplitudes of motions are quite large. Little is known about whether this flexibility is a feature of such regions which is central to the properties and functions of the proteins which contain them.

For example, in each of the 34 coat protein molecules in the 20S disk of tobacco mosaic virus, 29 of the 158 amino acids cannot be seen in the electron density map obtained at 2.8 Å resolution (Figure 1a (1)). Twenty five of these disordered residues make up part of the RNA binding site and are visible as a stable structure in the electron density map of the intact virus obtained from fiber diffraction studies at 4.0 Å resolution (2). However, the diffraction studies do not, by themselves, distinguish between static and dynamic disordering of the RNA binding loop. Nuclear magnetic resonance measurements on the ¹³C and proton nuclei in mutant and wild type virus particles and coat protein aggregates have established that the loop undergoes rapid motion in the absence of nucleic acid (Figure 1b (3-5)). RNA packaging is accompanied by the folding of these residues, amino acids 88 to 112. A model for RNA-protein interactions in tobacco mosaic virus which involves residues in the unfolded loop as well as several in the ordered part of the coat protein has been proposed (6,7).

Figure 1a. Side view of two subunits in the TMV disk with the ordered region of the polypeptide chain drawn as a folded ribbon (1). Between amino acids 89 and 114, the protein is disordered in the free disk. A proposal for the conformation which is stabilized when oligonucleotides bind is shown (-.-). 1b. Resolution enhanced proton nmr spectrum of disks from TMV Vulgare Ni 725 (5). Sharp lines in the 1-2.2 p.p.m. region suggest that all the residues between Asp 88 and Arg 112 are mobile.

Figure 1.



Nmr studies have suggested that several other proteins contain unfolded regions including histone H1 (8), calmodulin (9,10), fd phage coat protein (11) and several ribosomal proteins (4,12). The hormone glucagon has also been shown to exist as a number of interconverting conformers in solution, only one of which is stabilized on binding to its receptor (13,14). The amount of secondary structure in bacterial flagellin apparently increases during the assembly of flagella (15).

X-ray diffraction studies indicate that part of the ligand binding sites of the cro repressor of phage lambda (16) and triose phosphate isomerase (17,18) are disordered. Crystallographic evidence for flexibility in the activation domain of trypsinogen has also been reported, but the extent of this flexibility has been disputed (19-21). Indeed, recent work on the temperature dependence of the disorder (22) and nmr studies on the protein in solution (23) indicate that trypsinogen exists in several discreet conformations rather than being highly flexible.

In recent years, the thermodynamics of processes involving proteins have begun to be probed in general terms. Experimental and theoretical studies on protein folding have led to estimates of the energies required to stabilize a particular conformation of the polypeptide chain (24,25). Sturtevant has identified six possible structural sources of the large changes in heat capacity, enthalpy and entropy which characterize the reactions of proteins (26). These include hydrophobic effects, formation and breakage of hydrogen bonds and ion pairs, changes in the number of accessible isoenergetic conformations, changes in

intramolecular vibrational modes, and temperature-dependent shifts in conformational equilibria. He suggested a method for estimating hydrophobic and vibrational contributions which is consistent with measurements of heat capacity and entropy changes associated with several binding reactions. Jencks has suggested that a reduction in the number of ligand conformations on binding can result in a substantial entropic destabilization of the protein-ligand complex (27,28). If this conformational trapping is on the pathway to the formation of an enzyme-transition state complex, it can, in theory, greatly enhance the catalytic reaction rate.

In general, however, few correlations have been made between thermodynamic measurements and the known structural features of particular proteins. In this communication, the thermodynamic consequences of changes in the mobility of protein atoms resulting from ligand binding are examined and are used to illuminate the mechanisms by which particular proteins carry out their biological functions.

Unfolded Regions Destabilize Protein-ligand Complexes

The folding of proteins is known to involve the compensation of large entropic and enthalpic terms. However, the net free energy of stabilization of the folded state is modest, usually on the order of 10 kcal/mole (24). The existence of unfolded regions in native proteins suggests that the compensation can be incomplete. In such regions, interactions which favor the folded state, including the desolvation of

the exposed peptide chain, are not of sufficient magnitude to overcome destabilizing factors such as chain entropy and possible high energy charge-charge or charge-dipole interactions.

When ligand binding is accompanied by the ordering of an unfolded part of the protein, the reduction in the number of iso-energetic conformational states is accompanied by the acquisition of new normal modes of vibration. Some of the intrinsic binding energy must be dissipated to compensate for reduction in entropy associated with this process. For example, if the intrinsic binding energy for a protein-ligand interaction is 20 kcal/mole and 10 kcal/mole is required to order the binding site around the ligand, then the observed binding energy will be only 10 kcal/mole. This modulation of binding energy will be significant only if the interaction energy and the energy required to order the unfolded binding site are similar.

It has not been possible to directly measure the conformational entropy of the polypeptide chain, because the large number of contributions to the measured entropy change on folding are not experimentally separable (27,29). However, employing a number of different assumptions, several workers have attempted to calculate this quantity. Treating the polypeptide chain as a polymer coiling statistically about virtual bonds connecting the alpha-carbon atoms, Brant, Miller, and Flory have estimated thermodynamic parameters for chains of alpha-L-polyalanine and polyglycine in which the end-to-end distance is unconstrained (30). Assuming that a 10 degree rotation about the bonds represents a distinguishable state of the chain, the

average entropy per residue was found to be about 10 cal/degree mole. The loss of this conformational entropy upon folding at 37° C would entail a free energy increase of T times this value, or approximately 3 kcal/mole/residue.

The magnitude of the conformational entropy change on folding can also be evaluated by the expression:

$$\Delta S = R \ln \frac{(W_u)}{(W_f)}, \quad (1)$$

where W_u and W_f are the number of ways of arranging the unfolded and folded chain (and the surrounding solvent), respectively, and R is the gas constant (31). If, on average, there are m times more conformational states accessible to each residue in a peptide when it is unfolded as compared to when it is folded, Equation 1 reduces to:

$$\Delta S = R \ln \frac{1}{(m^n)}, \quad (2)$$

where n is the number of residues in the chain. Since it is not possible to determine the sign of the contribution of the solvent from first principles, the change in the number of solvent distributions is formally included in the evaluation of the change in the number of conformational states per residue, m. Ignoring the contributions from the side chains and the solvent, and noting that the area accessible to each residue in configuration space (as described by the Ramachandran plot) is decreased ten fold on going from a random coil to a helical conformation, the entropy loss for a polypeptide of 100 residues

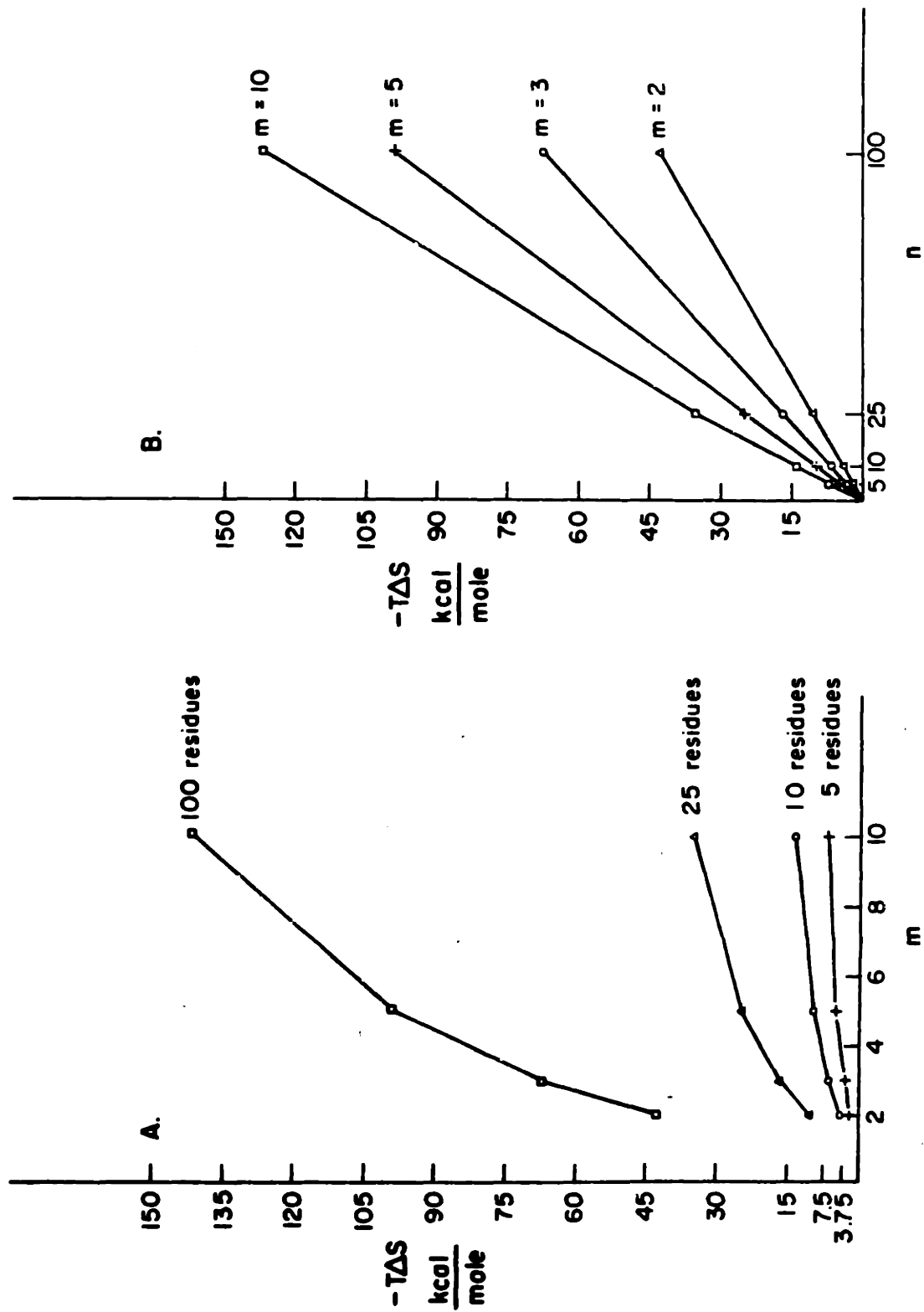
undergoing this transition was found to be -460 cal/degree mole (32). At 37° C, this results in a free energy increase of 142 kcal/mole.

Since the number of states accessible to a protein can be factored into the product of the number of states accessible to individual regions within it, Equation 2 can also be used to evaluate the conformational entropy change associated with the folding of only part of the polypeptide. For example, if the average number of states populated by each residue in the ten amino acid disordered loop of triose phosphate isomerase is reduced threefold when the substrate is bound, the free energy increase due to the loss of conformational entropy on folding the loop is about 6 kcal/mole at 37° C. Using the same assumptions, the free energy cost of folding a 25 amino acid peptide--for example, the RNA binding loop of the tobacco mosaic virus coat protein--is 17 kcal/mole at 37° C. Calculations showing the relationship between m , n and $T\Delta S$ in the absence of compensating low frequency vibrations in the bound complexes are summarized in Figure 2.

Though these calculations are extremely crude, they suggest that the conformational entropy of disordered residues may be of the same order of magnitude as experimentally determined entropies of internal rotations in small molecules in the gas phase. For the cyclization of saturated hydrocarbons up to 8 carbons in length, the entropy loss per internal rotation was found to range between 3.7 and 4.9 cal/degree mole (29).

Figure 2. Free energy cost of folding a polypeptide chain calculated by evaluating Equation 2 as a function of (A) the ratio of the number of rotational states accessible to each residue in the unfolded and folded conformations, m , and (B) the number of residues in the chain, n . For long chains, even a modest decrease in the number of states accessible to each residue results in a significant destabilization energy. Short unfolded regions can modulate ligand binding energies if the restriction in the number of rotational states in the folded peptide is sufficiently severe.

Figure 2.



The calculations presented above suggest that the conformational entropy of even a short stretch of disordered polypeptide chain can be significant compared to the energies of protein-ligand interactions. Following a similar argument, it can be seen that the restriction of a sufficiently large number of side chain rotations on ligand binding may also contribute to the dissipation of binding energy. In contrast, folded domains which move relative to each other do not contribute appreciably to the conformational entropy of a protein, because their breathing motions do not define a sufficiently large number of conformational states (27). Thus, unfolded regions of protein binding sites which become ordered on ligand binding can act as sources of negative entropy which must be compensated by the protein-ligand interactions. In doing so, they serve to decrease the stability of protein-ligand complexes. As discussed below, this can have important consequences for the activities of some enzymes and binding proteins.

Biological Roles of Unfolded Binding Sites

Though it is unusual to consider mechanisms by which proteins bind ligands weakly, there are several functions which require the dissipation of binding energy if they are to proceed at a rapid rate. These include enzymatic catalysis and reversible binding of macromolecules. In enzymatic catalysis, rate enhancement occurs to the degree that the free energy difference between the E-S complex and the E-transition state complex is less than the free energy difference between the substrate and transition state in the absence of the enzyme.

This can arise from a destabilization of the E-S complex due to the loss of rotational and translational entropy by the substrate on binding (27,28). It does not formally require a difference in the intrinsic binding energy between the enzyme and substrate as compared to that between the enzyme and the transition state. These concepts are treated in depth by Jencks (27,28).

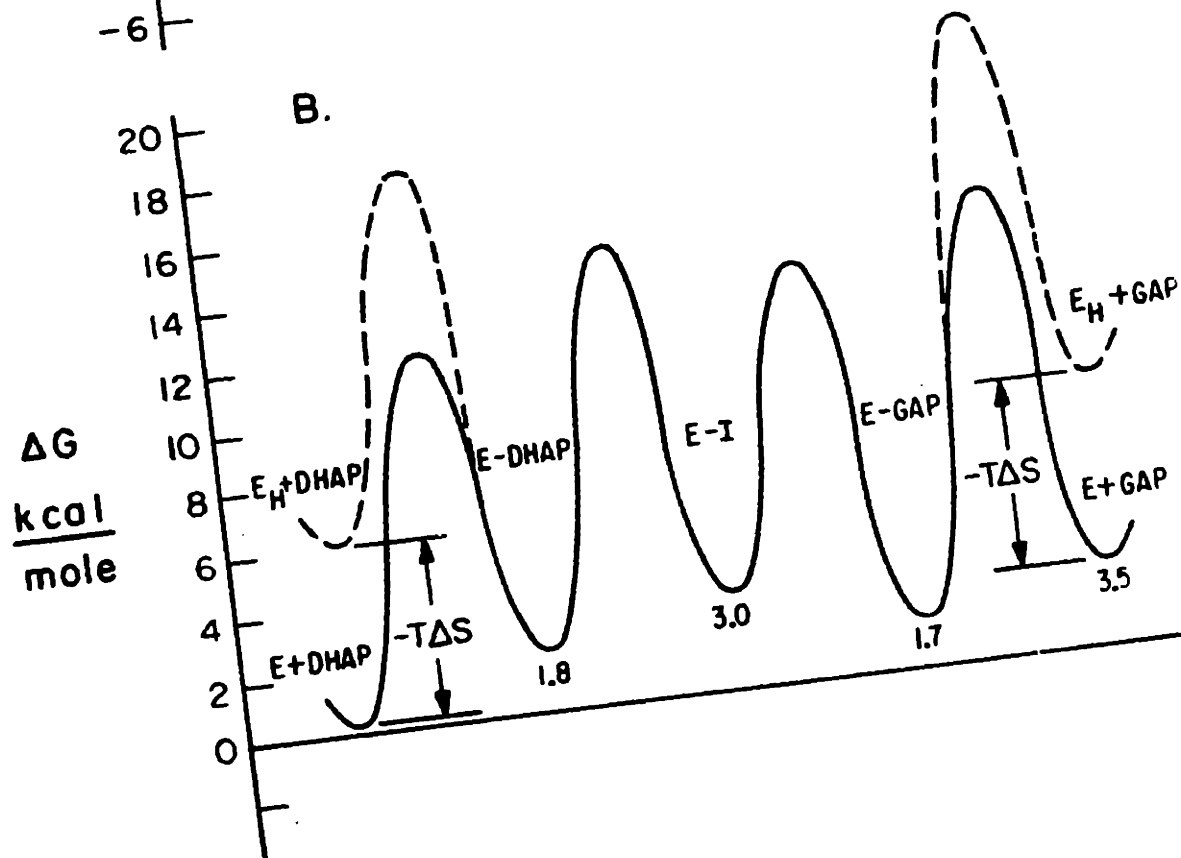
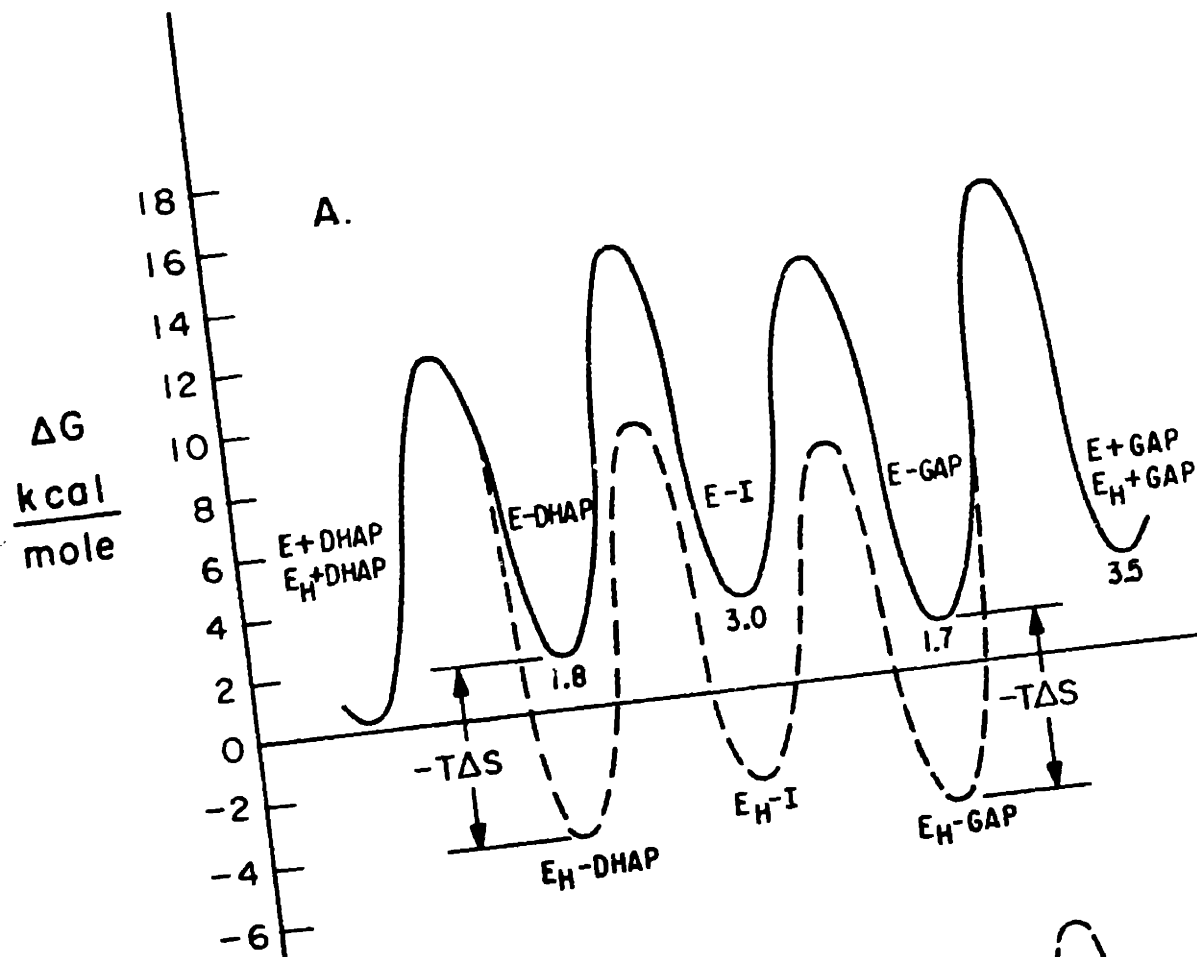
The ordering of a portion of an enzyme on substrate binding may also result in destabilization of the E-S complex. This mechanism may facilitate relatively tight binding of the transition state when its geometry is similar to that of the substrate itself, because, in such cases, high energy contacts between the enzyme and substrate would not be alleviated in the enzyme-transition state complex.

Triose phosphate isomerase catalyzes the interconversion of dihydroxyacetone phosphate and glyceraldehyde-3-phosphate (GAP) in glycolysis. Under optimum conditions, the reaction rate is at least partially diffusion controlled (33). The free energy profile for the reaction, with the intracellular concentration of triose phosphates defining the standard state, is shown in Figure 3 (33,34). The rate limiting transition state in the reaction of the enzyme purified from a number of organisms is encountered during the binding and release of GAP. As discussed above, the binding of substrates is accompanied by the folding of a disordered loop of ten residues. The sequence gly-thr-gly in the loop interacts with the phosphate moiety of the substrate. This interaction encloses the substrate in the active site, holding it in position to be acted on by the catalytic residues, and

Figure 3. Free energy profile for the reaction catalyzed by triose phosphate isomerase. The solid line shows the free energy profile derived from the kinetic and thermodynamic results of Albery and Knowles (27) and Rose (28). The dotted line shows the reaction coordinate of a hypothetical enzyme (E_H) which is identical to the natural one except that the substrate binding loop is stably folded prior to its interaction with the substrate. The value of $-T\Delta S$ describes the extent to which the E-S complex is destabilized by the ordering of the loop. The catalytic rates of the natural and hypothetical enzymes can be compared using the expression:

$$\ln k_f = \ln (k T \kappa / h) - (\Delta G^\ddagger / RT)$$

where k_f is the rate constant, k is the Boltzmann constant, h is Planck's constant, T is the absolute temperature, R is the gas constant, κ is the partition coefficient (assumed to be 1) and ΔG^\ddagger is the activation energy of product release for the natural and hypothetical enzymes. Assuming, as described in the text, that 6 kcal/mole of binding energy is dissipated on folding the loop, then at 37° C, the natural enzyme is 1.7×10^4 times faster than its hypothetical homologue. A. Free energy profiles assuming no change in the standard state. In this case, the amino acid sequence in the loop of the hypothetical enzyme is assumed to be different than its real counterpart. B. Free energy profiles assuming that the amino acid sequences of E and E_H are the same and that the energy of E_H is raised by the folding of the loop prior to substrate binding.



inhibits the nonenzymatic dephosphorylation of the enediol intermediate (18). This helps make the reaction specific without hindering the access of the substrates to the active site. The enthalpy of activation for substrate binding is minimized, and, in addition, the dissipation of some of the binding energy by the folding of the loop raises the free energy of the E-S complex relative to what it would have been had the phosphate been bound only by well ordered domains. As shown in Figure 3, this directly affects the activation energy, and consequently the rate, of the slowest step of the reaction. If, as calculated above, 6 kcal/mole is dissipated in folding the loop, then the triose phosphate isomerase found in nature is 1.7×10^4 times faster at 37°C than a hypothetical enzyme which is identical except that the loop is folded throughout the reaction.

This disorder-order transition has been observed in crystallographic studies of both the yeast and chicken triose phosphate isomerases (18). These proteins are similar in structure yet crystallize in different space groups with different intermolecular contacts, supporting the idea that the disorder may reflect an intrinsic flexibility of this part of the protein and is not an artifact of crystallization (35). Evidence consistent with the dissipation of binding energy by the loop disorder comes from crystallographic studies of the binding of sulfate and phosphate to chicken triose phosphate isomerase (17,18). Sulfate and phosphate are competitive inhibitors which bind to the enzyme in the same position in the active site, and both have dissociation constants of about 10^{-3} M (36). When sulfate is bound, the disordered loop remains disordered. However, phosphate

binding is accompanied by the ordering of the loop in a manner reminiscent of its interaction with the phosphate moiety of the substrate. The similarity in the dissociation constants occurs despite these additional interactions between the protein and the phosphate ion, suggesting that some of binding energy may be expended to order the loop.

In binding to the active site, the substrates change the equilibrium distribution of loop conformations. An examination of the rates of protein folding in vitro (as fast as several msec) and the measured rates of helix propagation (10^{-5} to 10^{-7} sec) indicates the this process can be at least as fast as the diffusion-controlled limit for a bimolecular reaction (37).

Bode et al (20,38) have proposed that a disorder-order transition acts through a similar mechanism to produce weak binding of substrates and inhibitors to trypsinogen. Trypsin is produced from trypsinogen by cleavage of the flexible N-terminal hexapeptide (23). After cleavage, the new N-terminus forms an ion pair with Asp 194, and statically disordered residues in the activation domain -- Gly 142 to Pro 152, Gly A184 to Gly 193, and Gly 216 to Asn 223 -- assume the fold found in trypsin. Binding of bovine pancreatic trypsin inhibitor and a dipeptide which mimicks the N-terminus of the active enzyme (Ile-Val) stabilizes trypsinogen in a trypsin-like conformation (20,23). The flexibility and/or the concentration of negative charge in the trypsinogen N-terminus may help destabilize a trypsin-like conformation and give rise to the requirement for a positively charged group which can

interact specifically with Asp 194 if the active conformation is to be formed (20,39).

However, nmr studies have shown that the N-terminus is the only part of the activation domain which is highly flexible (23). In this light, the direct role of conformational entropy in producing weak binding of substrates and inhibitors to trypsinogen remains unclear. Working with a different crystal form of the zymogen than Bode et al., Stroud and coworkers have found considerably less disorder in the protein structure (21). They have suggested that differences in the active site conformation leading to differences in the mode of ligand binding are sufficient to account for the reduced binding and catalytic activity of trypsinogen as compared to trypsin (21,39). Consequently, there are at least two models for the structural basis for the lack of activity of trypsinogen which have not been invalidated by experiments.

- 1) The zymogen may exist in a small number of conformations which are inactive and considerably more stable than the active conformers.
- 2) Alternatively, the active conformers may form only very rarely and be unstable because of the existence of a large number of iso-energetic inactive conformations.

Binding proteins are another class of molecule which may interact with ligands through flexible domains at the expense of binding energy. Interactions between macromolecules--such as the binding of peptides by enzymes and hormone receptors, protein oligomerization and protein-nucleic acid interactions--are often extremely energetically favorable (40). The large surface areas over which the interactions can

occur allows an appreciable entropic stabilization from the release of many bound water molecules and a large favorable contribution from the sheer number of groups which can form hydrogen bonds, ion pairs, or participate in induced interactions. Consequently, the intrinsic binding energy between macromolecules can be, in principle, very high, particularly when the specificity of recognition requires that many groups interact. If the binding is to be both specific and reversible, some of the intrinsic binding energy must be dissipated. Apparently a number of proteins do this by coupling specific binding with the ordering of unfolded regions of their binding sites. Several examples illustrate these points.

The assembly of tobacco mosaic virus is initiated by the binding of a 20S disk to a specific site in the viral RNA. The disk binds 102 nucleotides, and though the structural origin of the specificity of initiation is not well understood, the RNA sequence around the initiation site is known (41). A loop of 25 amino acids in each of the 34 coat protein molecules of the disk becomes ordered as the RNA binds, and the disk undergoes a conformational change. The disorder of the loop suggests the presence of high energy interactions which are compensated by the interaction with RNA. This feature of the coat protein may contribute to the specificity of initiation by requiring an optimum interaction between the RNA and the first disk to be bound, an interaction which may occur only at a specific RNA sequence or structure. The elongation of the virus could then be facilitated by protein-protein interactions which are not possible during the initiation step (41,42).

Proteolytic cleavage of the RNA binding loop results in nonphysiological and irreversible polymerization of 20S disks (43). Thus, the disorder of the loop may also help determine the free energy of the assembled virus and contribute to the reversibility of the assembly process. The freezing out of mobility in the coat protein backbone may also be important for the reversibility of assembly of fd phage (11).

Disordered regions have been found in a number of repressor proteins. Genetic and physical studies of the E. coli lac repressor suggest that it is a tetramer of subunits which contain N- and C-terminal domains in rapid independent relative motion (44-47). The two domains are connected by a flexible peptide, residues 51 to 59. Mutant repressors which bind DNA more tightly than the wild type protein can arise from lesions in the both domains and in a residue adjacent to the flexible connecting peptide (48). The presence of inducers changes the protease susceptibility of the disordered peptide which connects the two domains (49). Ogata and Gilbert compared the operator affinities and methylation protection patterns of wild type and mutant repressors and N-terminal fragments (50). They found that the N-terminal headpieces protected purines in the operator DNA in a manner qualitatively similar to protection by intact repressor. At room temperature, repressor, induced repressor and mutant repressors elicited the same pattern of methylation protection and enhancement. They concluded that two N-terminal headpieces per tetramer make all the contacts with the operator DNA and argued that binding of the inducer

reduces the strength of the many contacts between repressor and operator rather than totally disrupting one or a few bonds (50). They also found that the dissociation constant of the headpiece-operator complex was low enough to account for all of the binding energy of the repressor-operator complex.

More recently, Matthews found that "core tetramers", composed solely of C-terminal domains, specifically bind operator DNA in a manner which is sensitive to the presence of inducer (51). If this represents a physiological interaction between operator and intact repressor, it would contribute 40 to 50% of the observed free energy of operator binding and 60% of the change in operator affinity brought about by the inducer. Matthews and coworkers have proposed a model for the lac repressor and its interactions with ligands in which both N- and C-terminal domains of all four subunits bind to the operator and effectors bind only to the C-terminal domain (52). Taken together, these data require the dissipation of some of the intrinsic binding energy during interactions of the repressor with DNA and as a mechanism of induction. The presence of a flexible region in lac repressor is consistent with the idea that binding may be weakened by an entropic mechanism involving a reduction in the number of conformational states accessible to the protein.

The fluorescence depolarization experiments of Bandyopadhyay, Wu and Wu provide evidence which is directly consistent with this hypothesis (53). By measuring the rate of decay of the intrinsic fluorescence emission anisotropy, these workers found that the

tryptophans in the free lac repressor undergo rapid independent motion. They concluded that one or both of the tryptophans, located at positions 190 and 209 in the C-terminal domain, are in a flexible segment of the molecule. In the presence of the inducer, isopropyl- β -D-galactoside, the flexibility was enhanced. In contrast, binding of an anti-inducer caused a restriction of the mobility of the tryptophan residues, as did the binding of calf thymus DNA. Similar results were obtained by monitoring motional changes of a fluorescent probe covalently linked to the protein. Thus the lac repressor has the characteristics of a protein in which binding energy is modulated by chain flexibility. Its weak binding state (induced) is more flexible than the free protein, and its tight binding state is less flexible. In addition, motions in the protein are damped on binding to DNA.

The cro repressor of phage λ recognizes three approximately symmetrical 17 base pair sequences in the right operator of the phage DNA (54). The x-ray structure of this protein has recently been solved, and a model for repressor-operator interactions has been proposed (16). The residues which are thought to interact with the DNA include several in a helix containing amino acids 27 to 36, amino acids 54 to 56 and the C-terminal amino acids, four of which are disordered in the crystal. If the disordered amino acids are mobile in solution and occupy well defined positions when the protein is bound to the operator, the flexibility may decrease the net binding energy. This would prevent all of the intrinsic binding energy of the repressor-operator interaction from being expressed in the complex. Thus, the binding can have the extreme specificity of interactions with 17 base pairs and still be

reversible in a biologically useful amount of time. In addition, if interactions with non-operator DNA sequences are too weak to order the C-terminus of the repressor, then these interactions will not contribute to the binding of repressor to non-operator DNA. This may increase the free energy difference between interactions with specific and nonspecific sequences.

If interactions with unfolded binding sites decrease the binding energy between ligands and proteins, then tight binding may be facilitated by interactions with rigid regions of proteins. This conclusion is consistent with the finding that peptide inhibitors and toxins which bind extremely tightly to their target structures have been found to be relatively inflexible molecules extensively crosslinked by disulfide bonds (11,40,55-57).

Discussion

In its simplest form, the model for entropic destabilization of protein-ligand complexes presented here predicts that the free energy of the complexes will vary with temperature. This may be the easiest way to detect such an effect. In the reaction of triose phosphate isomerase, for example, it implies that the activation energy will be temperature dependent, and the logarithm of the overall rate constant describing the steady state conversion of DHAP to GAP will not be a linear function of $1/T$ (58). In addition, as the temperature is lowered, the kinetic isotope effect observed in the reverse reaction

should decrease as the release of DHAP becomes rate limiting (Figure 3)(59). When the reaction is carried out in tritiated water, the extent of isotopic labelling of DHAP should increase at lower temperatures (60).

Motions in proteins are thought to play a number of functional roles. These include providing ligands access to binding sites, facilitating conformational changes, and allowing proteins to entrap ligands in order to maximize binding energy or exclude solvent (61-64). These functions can be carried out by ordered regions and do not necessarily involve a large change in the number of conformations which the protein can adopt.

In contrast, extremely flexible regions of protein binding sites can also act to decrease the observed free energy of protein-ligand interactions if binding is accompanied by a sufficiently severe restriction in the number of protein (and solvent) conformations. This has at least two consequences. First, it allows binding between macromolecules to be both extremely specific and sufficiently reversible. Binding through unfolded regions may be a general mechanism by which a number of proteins--including TMV coat protein, repressors, peptide hormones, calmodulin, ribosomal proteins, histones and other nucleic acid binding proteins--interact with ligands through a large number of contacts without the expression of all of the intrinsic binding energy. Second, it can raise the free energy of complexes of enzymes with their substrates and products. In triose phosphate isomerase, this may contribute directly to the efficiency of the enzyme.

ACKNOWLEDGEMENTS

I thank James Kinsey, Greg Petskc, Ed Loechler and Bob Sauer for helpful discussions. This work was supported by a Danforth Foundation Graduate Fellowship and a National Institutes of Health Graduate Training Grant to the Biology Department at MIT.

REFERENCES

1. A.C. Bloomer, J.N. Champness, G. Bricogne, R. Staden and A. Klug (1978), *Nature*, 276, 362-368.
2. G. Stubbs, S. Warren and K. Holmes (1977), *Nature*, 267, 216-221.
3. J.L. deWit, N.C.M. Alma-Zeestraten, M.A. Hemminga and T.J. Schaafsma (1979), *Biochemistry*, 18, 3973-3976.
4. J.L. deWit, L.C.J. Dorssers, and T.J. Schaafsma (1979), *Bioch. Biophys. Res. Comm.*, 89, 435-440.
5. O. Jardetzky, K. Akasaka, D. Vogel, S. Morris and K.C. Holmes (1978), *Nature*, 273, 564-566.
6. K.C. Holmes (1979), *J. Supramolecular Structure*, 12, 305-320.
7. G. Stubbs and C. Stauffacher (1980), *Biophys. J.*, 32, 244-246.
8. G.E. Chapman, P.G. Hartman and E.M. Bradbury (1976), *Biochemistry*, 61, 69-75.
9. K. Seaman (1979), *Biochem. Biophys. Res. Comm.*, 86, 1256-1265.
10. J.S. Evans and B.A. Levine (1980), *J. Inorg. Biochem.*, 12, 227-239.
11. S.J. Opella, T.A. Cross, J.A. DiVerdi and C.F. Sturm (1980), *Biophys. J.*, 32, 531-548.
12. C.A. Morrison, E.M. Bradbury and R.A. Garrett (1977), *FEBS Letters*, 81, 435-439.
13. W.B. Gatzler, G.H. Beavan, H.W.E. Rattle and E.M. Bradbury (1968), *Eur. J. Biochem.*, 3, 276-283.
14. E.C. Moran, P.Y. Chou and G.D. Fasman (1977), *Biochem. Biophys. Res. Comm.*, 77, 1300-1306.
15. Y. Uratani, S. Asakura and K. Imahori, *J. Molec. Biol.*, 67, 85-98.
16. W.F. Anderson, D.H. Ohlendorf, Y. Takeda, and B.W. Matthews (1981), *Nature*, 290, 754-758.
17. D.W. Banner, A.C. Bloomer, G.A. Petsko, D.C. Phillips, C.I. Pogson, I.A. Wilson, P.H. Corran, A.J. Furth, J.D. Milman, R.E. Offord, J.D. Priddle and S.G. Waley (1975), *Nature*, 255, 609-614.
18. T. Alber, D.W. Banner, A.C. Bloomer, G.A. Petsko, D.C. Phillips, P.S. Rivers, and I.A. Wilson (1981), *Phil. Trans. Roy. Soc., London, B*, 293, 159-171.

19. H. Felhammer, W. Bode and R. Huber (1977), *J. Mol. Biol.*, 111, 415-438.
20. W. Bode, P. Schwager and R. Huber (1978), *J. Mol. Biol.*, 118, 99-112.
21. A.A. Kossiakoff, J.L. Chambers, L.M. Kay and R.M. Stroud (1977), *Biochemistry*, 16, 654-664.
22. J. Walter, W. Steigemann, T.P. Singh, H. Bartunik, W. Bode and R. Huber (1981), Submitted to *Acta Cryst. A*.
23. S.J. Perkins and K. Wuthrich (1980), *J. Mol. Biol.*, 138, 43-64.
24. C.N. Pace (1975), *CRC Crit. Rev. Biochem.*, 3, 1-43.
25. N. Go (1976), *Adv. Biophys.*, 9, 65-113.
26. J.M. Sturtevant (1977), *Proc. Nat. Acad. Sci. USA*, 74, 2236-2240.
27. W.P. Jencks (1980), in Chemical Recognition in Biology, F. Chapeville and A.-L. Haenni, eds., Springer-Verlag, New York, pp 3-25.
28. W.P. Jencks (1975), *Advances in Enzymology*, 43, 219-410.
29. M.I. Page and W.P. Jencks (1971), *Proc. Nat. Acad. Sci., USA*, 68, 1678-1683.
30. D.A. Brant, W.G. Miller and P.J. Flory (1967), *J. Mol. Biol.*, 23, 47-65.
31. F.C. Andrews (1975), Equilibrium Statistical Mechanics, John Wiley and Sons, New York.
32. G.E. Schulz and R.H. Schirmer (1979), Principles of Protein Structure, Springer-Verlag, New York, page 150.
33. W.J. Albery and J.R. Knowles (1976), *Biochemistry*, 15, 5627-5631.
34. I.A. Rose (1981), *Phil. Trans. Roy. Soc. London, B*, 293, 131-143.
35. The level of noise in an electron density map can be used to establish the minimum number of alternate conformations. In yeast triose phosphate isomerase, the analysis of the structure at 3.0 Å resolution shows that the disordered loop must exist in a minimum of four conformations in the free enzyme (T.A. and G.A. Petsko, unpublished results).
36. E.A. Noltmann, in The Enzymes, Vol. VI, P.D. Boyer, ed., Academic Press, New York, 1972, pg. 271.
37. P.S. Kim and R.L. Baldwin (1982), *Ann. Rev. Biochem.*, 51, in press.

38. W. Bode (1979), *J. Mol. Biol.*, 127, 357-374.
39. R.M. Stroud, A.A. Kossiakoff and J.L. Chambers (1977), *Ann. Rev. Biophys. Bioeng.*, 6, 177-193.
40. D. Tsernoglou and G.A. Petsko (1976), *FEBS Letters*, 68, 1-4.
41. D. Zimmern (1977), *Cell*, 11, 463-482.
42. G. Lebeurier, A. Nicolaieff and K.E. Richards (1977), *Proc. Nat. Acad. Sci., USA*, 74, 149-153.
43. A.C.H. Durham (1972), *FEBS Letters*, 25, 147-152.
44. F. Buck, H. Ruterjahn and K. Beyreuther (1978), *FEBS Letters*, 96, 335-338.
45. N. Wade-Jardetzky, R.P. Brady, W.W. Conover, O. Jardetzky, N. Geisler and K. Weber (1979), *J. Mol. Biol.*, 128, 259-264.
46. M.A.C. Jarema, P. Lu and J.H. Miller (1980), *Biophysical J.*, 32, 450-452.
47. K. Adler, K. Beyreuther, E. Fanning, N. Geisler, B. Gronenborn, A. Klemm, B. Muller-Hill, M. Pfahl and A. Schmitz (1972), *Nature*, 237, 322-327.
48. J.H. Miller, C. Coulondre, M. Hofer, U. Schmeissner, H. Sommer, A. Schmitz and P. Lu (1979), *J. Mol. Biol.*, 131, 191-222.
49. K. Beyreuther, in *The Operon*, J.H. Miller and W.S. Resnikoff, eds., Cold Spring Harbor Lab. Quant. Biol., Cold Spring Harbor, NY, 1978, 123-154.
50. R.T. Ogata and W. Gilbert (1979), *J. Mol. Biol.*, 132, 709-728.
51. K.S. Matthews, *J. Biol. Chem.*, 254, 3348-3353.
52. M. Dunaway, S.P. Manly and K.S. Matthews (1980), *Proc. Nat. Acad. Sci., USA*, 77, 7181-7185.
53. P.K. Bandyopadhyey, F.Y-H. Wu and C-W. Wu (1981), *J. Mol. Biol.*, 145, 363-373.
54. T. Maniatis, M. Ptashne, K. Backman, D. Kleid, S. Flashmar, A. Jeffrey and A. Maurer (1975), *Cell*, 5, 109-113.
55. W.A. Hendrickson and M.M. Teeter (1981), *Nature*, 290, 107-113.
56. R. Huber, D. Kukla, W. Bode, P. Schwager, K. Bartels, J. Deisenhofer and W. Steigemann (1974), *J. Mol. Biol.*, 89, 73-101.
57. R.M. Sweet, H.T. Wright, J. Janin, C.H. Chothia and D.M. Blow (1974), *Biochemistry*, 13, 4212-4228.

58. S.G. Maister, C.P. Pett, W.J. Albery and J.R. Knowles, (1976), *Biochemistry*, 15, 5607-5612.
59. P.F. Leadlay, W.J. Albery, and J.R. Knowles (1976), *Biochemistry*, 15, 5617-5620.
60. S.J. Fletcher, J.M. Herlihy, J.M. Albery, J.R. and Knowles (1976), *Biochemistry*, 15, 5612-5617.
61. H. Frauenfelder, G. A. Petsko and D. Tsernoglou (1979), *Nature*, 280, 558-563.
62. R. Huber (1979), *Trends in Bioch. Sci.*, 4, 271-276.
63. D.E. Koshland (1964), *Brookhaven Symposium*, 473-482.
64. R. Wolfenden (1974), *Molec. Cell. Bioch.*, 3, 207-211.

RECOGNITION OF EUKARYOTIC SIGNAL SEQUENCES

Tom Alber
Department of Biology
Massachusetts Institute of Technology
Cambridge, Mass. 02139

During their synthesis, proteins begin to be sorted to function in different parts of cells.¹ While cytoplasmic proteins are made by soluble polyribosomes, proteins which are secreted or integrated into membranes are synthesized on ribosomes bound to the rough endoplasmic reticulum (ER). It has been found that the N-terminal amino acids of a nascent polypeptide can contain the signal which mediates its binding to and synthesis through the ER membrane.²⁻⁵ Several groups have demonstrated that dog pancreas membranes added to a cell free protein translating system can sequester and specifically modify secreted as well as integral membrane proteins from a wide variety of eukaryotic organisms and tissues.⁵⁻⁷ Competition experiments suggest that the sequestration of these proteins proceeds through at least one common early step.⁸

Warren and Dobberstein have accomplished a dissection and reconstitution of the protein transporting system in dog pancreas membranes.⁹ Their results suggest the presence of protein components tightly bound to the cytoplasmic side of the rough ER which may interact with proteins which span the ER membrane. In the present study, attention is focused on

the signal sequences from a variety of eukaryotic proteins in order to shed more light on the recognition of the nascent polypeptide chain and its subsequent transport and modification.

The proteins whose signal sequences are compared are shown in figure 1.^{8,10-21} The signal peptides have been defined as the sequences cleaved from the nascent polypeptide chain by the microsomal signal peptidase, though it is not formally required that membrane recognition activity reside totally in this fragment. (For the VSV G protein, membrane recognition activity has been shown to be localized in no more than the first 40 amino acids.²²) The signal sequences are between 16 and 30 amino acids long, and they can be grouped into sets with homologous regions. The signal peptides of albumin, MOPC 41 IgG, trypsinogen 2, and rat insulin all contain stretches of six to eight amino acids in which 62-100% of the residues are identical. MOPC 104E pre-IgG, prelysozyme, prepro-PTH, and prealbumin form another set in which the homologies are less dramatic (up to four of six residues). The leu-leu-cys-leu-leu sequence of pre-RGH is intermediate between the two. Preovomucoid contains two overlapping hexapeptides which are 50% homologous to the set containing prealbumin.

Since the probability of finding identical hexapeptides in randomly generated sequences of 20 hydrophobic amino acids is on the order of 10^{-6} and that of finding hydrophobic octapeptides in which six of the residues are identical is about

5×10^{-6} , it can be seen that the observed homologies are decidedly non-random. However, there is no basic sequence which is common to all of the signal peptides; some pairs of sequences have only dipeptides in common.

It is interesting to note that pre-VSV G contains an octapeptide, leu-leu-tyr-leu-ala-phe-leu-phe, which is the mirror image of the region characteristic of the set of sequences containing prealbumin in six of eight positions. The N-terminus of ovalbumin contains a hexapeptide, phe-cys-phe-asp-val-phe, which is very nearly the mirror image of the phe-val-leu-phe-ser-phe sequence of preovomuroid. The sequences also contain a high degree of perfect or approximate internal mirror symmetry. The importance of this is not apparent, though it has also been found in glucagon, a 29 amino acid hormone involved in sugar metabolism.²³

On average, the fraction of hydrophobic residues in signal sequences is over twice as large as the average for soluble proteins. By aligning the sequences near their C-terminal ends, it can be seen that there is some conservation in the positions of uncharged polar, charged, and hydrophobic amino acids in the primary structure. Most importantly, there is a stretch of 11 amino acids which occurs in the middle of the sequences in which charged residues are never found. This is shown in figure 1.

Since the basis of the common function of the signal peptides was not apparent from a comparison of primary structures, secondary structural probabilities were investigated by a predictive method based on that of Chou and Fasman.²⁴ In this method, the aggregate relative probabilities of finding each sequential run of four to six amino acids in a given secondary structure-- α -helix, β -sheet, or β -turn--are calculated and compared numerically with average probabilities for observed helices, sheets, and turns. An updated set of individual relative probabilities for each of the 20 common amino acids was kindly provided by Mike Levitt.²⁵

Though the Chou and Fasman method has been shown to be one of the best available, no method of secondary structure prediction can currently be used to describe the real fold of a polypeptide chain with a known degree of accuracy.²⁶ For example, Matthews has shown that the Chou and Fasman method can sometimes generate a prediction which correlates with the observed structure hardly better than a random assignment of secondary structural features. In most cases, the prediction may be much better, but there is no way to gauge its accuracy short of determining the structure of the protein in question. Consequently, the predictions described below are in no way meant to represent the actual structural features of the peptides. Rather, they have been done to compare the folding tendencies of the signal sequences and to determine whether

these sequences fold into similar or dissimilar structures. The same approach has been used by Dufton and Hider to compare the structure-activity relationships in 57 neurotoxins and cardiotoxins from snake venoms. 27

Predictions of secondary structure for several signal sequences are shown in figure 2, and these are summarized graphically in figure 3. The predictions fall into two classes, one of which is a structural subset of the other. The N-terminal six to ten amino acids of the longer signal sequences (22 to 30 amino acids) generally include residues which are often found in β -turns. This is followed by a stretch of 8 to 14 amino acids in which predictions of α -helix and β -sheet overlap and may even alternate as the amino acids are sequentially considered as the first of the predicted set.

This feature corresponds to the region in which charged residues are never found. The region of helix-sheet overlap is generally followed by a β -turn prediction near the C-terminal end of the sequences. As shown in figure 3, this predicted turn is in register with the site of signal peptidase cleavage.

Predictions for the shorter signal peptides contain the second two features described above and lack the N-terminal β -turn. The absence of a predicted β -turn in the first ten amino acids of the signal sequence of the G protein of vesicular stomatitis virus (VSV) resulted in the correct assignment of this sequence to the short class of signal peptides before its length (16 amino acids) was experimentally determined (Flora

Katz, individual communication).

The N-terminal sequences of the primary translation products of four proteins found in different locations on the outside of *E. coli* -- β -lactamase, lipoprotein, and the coat proteins of fd phage -- were found to have predicted secondary structures somewhat similar to the pattern described above for eukaryotic signal peptides. However, predictions of the secondary structures of the N-termini of a number of intracellular proteins from prokaryotes and eukaryotes show no consistent regions of homology with each other or with the signal sequence pattern.^{24,25}

As shown in figure 2, the predicted secondary structure of the first 30 amino acids of ovalbumin is similar to the pattern observed in the longer signal sequences. In particular, β -turn probabilities are relatively high in the first eight amino acids and for amino acids 24 to 27. A six amino acid stretch where helix and sheet predictions overlap is found in residues 11 to 16. This suggests that ovalbumin contains a signal sequence which is not cleaved by the signal peptidase and that it is transported across the membrane of the rough ER by a mechanism similar to that for other secreted proteins. This implies that the transport and cleavage functions are mediated by different features of the signal peptide structure. Such a separation of functions has been observed in a signal sequence mutant of the *E. coli* outer membrane lipoprotein by Lin *et al.*²⁸ This mutant protein is inserted into the outer membrane, but its signal peptide is not

cleaved.

These findings with regard to ovalbumin are not in agreement with the conclusions of Palmiter *et al* that the absence of an extended sequence which is cleaved from the primary translation product implies a different mechanism of secretion for this protein.²⁹ They are, however, compatible with the finding that nascent ovalbumin chains compete with other signal peptides for sequestration in dog pancreas membranes.⁸ They also suggest that amino acids 11 to 16 (phe-cys-phe-asp-val-phe) comprise an important part of the ovalbumin signal sequence. Hydrophobic residues are localized in this region of the N-terminus, and overlapping helix and sheet predictions are made in this sequence. The sequence is also roughly the mirror image of the phe-val-leu-phe-ser-phe sequence in the signal peptide of chicken ovomucoid. The differences in the processing of ovalbumin *in vitro* may be based on primary structural differences in its "signal peptide". For example, asp 14, lys 17 and lys 20 all fall into the region where charged residues are not found in other signal sequences, the region of helix-sheet overlap is smaller, and the sequence is generally less hydrophobic.

The results of the secondary structure prediction suggest that signal peptides tend to fold into similar conformations *in vivo* despite the absence of consistent sequence homology. This, in turn, implies that transmembrane transport is mediated by proteins which specifically act as receptors for the nascent polypeptide chain and that only one or two types of such

receptors may recognize all the secreted and integral membrane proteins synthesized on the rough ER. This supports the conclusions of Lingappa *et al* that ovalbumin, prolactin and VSV G are recognized by common receptors.⁸ That different sequences have common receptors suggests that they share other steps in transport, as well.

However, the real process of signal peptide recognition must be more complex than described above. The assertion that secondary structural homologies among signal sequences account for their common activity assumes that they do, in fact, have a single stable structure in solution. Given that the basic units of protein structure are thought to be made up of combinations of secondary structural features whose tertiary fold is stabilized by hydrophobic interactions, it is unlikely that the short 16 to 30 amino acid stretch of the nascent chain is large enough to fold into a single stable conformation.^{30,31} It has been found, for example, that the 29 amino acid hormone, glucagon, exists as a mixture of conformers in solution, only one of which is thought to be stabilized by binding to the receptor.^{20,32,33} (It is interesting to note that multiple conformations for glucagon were predicted by the Chou and Fasman method.³⁴)

In a similar fashion, the signal sequence receptor may serve to stabilize one of the many conformations of the nascent signal peptide. The predominance and conservation of hydrophobic residues in the signal sequences suggest that the recognition

process is heavily dependent on hydrophobic interactions. This is supported by the observation that the hydrophobic regions are not in register with the site of cleavage (figure 1). The localization of charged and polar residues at the ends of the sequences implies that these types of amino acids have specific roles as well. However, the lack of conservation of sign in the charged residues of several signal sequences and the absence of charged residues in MOPC 104E pre-IgG light chain and in the signal peptides of two pancreatic secreted proteins point away from the obligatory involvement of salt bridges in the recognition process.

In summary, the prediction of secondary structures provides evidence for a similar fold among signal peptides in the environment of a protein receptor, where peptide-receptor interactions may be equivalent to the tertiary interactions within a protein. The specificity of signal peptide recognition may depend on the conservation of hydrophobicity in definite regions of the peptides and on a sequence of residues which favors a fold in which the hydrophobic residues form similar domains. Thus, one or two types of receptor may recognize signal peptides with dissimilar sequences from a variety of proteins. The region recognized by the signal peptidase is near a predicted β -turn. In the case of ovalbumin, it is likely that receptor recognition is separated from the process of cleavage. The proteins which can be washed from dog pancreas membranes with high salt may include

the signal sequence receptors. Further purification of the components of the transport system will allow the number of receptors to be determined. Physical studies of the characteristics of the signal peptides in solution have been initiated to investigate their properties and compare their conformations.

ACKNOWLEDGEMENTS

I appreciate the special interest and critical guidance of Alex Rich, Flora Katz, Greg Petsko, Joe Majzoub, Hank Kronenberg, Joel Habener, Harvey Lodish, Mike Levitt, and Gerald Fasman. Special thanks to Milt Anisman for getting me started and to Carol Jean Papineau for secretarial assistance. I wish to thank the Danforth Foundation and the Biology Department at M.I.T. for financial support during the course of this work.

REFERENCES

1. Reviewed by Palade, G. G., *Science*, 189 (1975), 347-358.
2. Blobel, G., "Synthesis and Segregation of Secretory Proteins: The Signal Hypothesis" in International Cell Biology 1976-1977: Papers Presented at the First International Congress on Cell Biology. Brinkley, B. R. & Porter, K. R., eds., Rockefeller University Press, 1977, 318-325.
3. Milstein, C., Brownlee, G., Harrison, T., and Mathews, M. B., *Nature New Biology*, 239 (1972), 117-120.
4. Blobel, G. and Dobberstein, B., *J. Cell Biol.*, 67 (1975), 835-851.
5. Blobel, G. and Dobberstein, B., *J. Cell Biol.*, 67 (1975), 852-862.
6. Dobberstein, B. and Blobel, G., *Biochem. Biophys. Res. Commun.*, 74 (1977), 1675-1682.
7. Katz, F. N., Rothman, J. E., Lingrippa, V. R., Blobel, G. and Lodish, H. F., *Proc. Natl. Acad. Sci. USA*, 74 (1977), 3278-3282.
8. Lingappa, V. R., Katz, F. N., Lodish, H. F. and Blobel, G., *J. Biol. Chem.*, 253 (1978), 8667-8670.
9. Warren, G. and Dobberstein, B., *Nature*, 273 (1978), 569-571.
10. Seeburg, P. H., Shine, J., Martial, J. A., Baxter, J. D. and Goodman, H. M., *Nature*, 270 (1977), 486-494.
11. Strauss, A. W., Bennett, C. D., Donohue, A. M., Rodkey,

- J. A. and Alberts, A. W., *J. Biol. Chem.*, 252 (1977), 6846-6855.
12. Habener, J., individual communication.
 13. Sutcliffe, J. G., *Proc. Natl. Acad. Sci. USA*, 75 (1978), 3737-3741.
 14. Suchanek, G., Kreil, G., and Hermodson, M. A., *Proc. Natl. Acad. Sci., USA*, 701-704 (1978).
 15. Thibodeau, S. N., Palmiter, R. D., and Walsh, K. A., *J. Biol. Chem.*, 253 (1978), 9108-9023.
 16. Caterall, J. F., O'Malley, B. W., Robertson, M. A., Staden, R., Tanakay, Y., and Brownlee, G. G., *Nature*, 275 (1978), 510-513.
 17. Devillers-Thiery, A., Kindt, T., Scheele, G. and Blobel, G., *Proc. Natl. Acad. Sci. USA*, 72 (1975), 5016-5020.
 18. Shields, D. and Blobel, G., *Proc. Natl. Acad. Sci. USA*, 74 (1977), 716-720.
 20. Strauss, A. W., Donohue, A. M., Bennett, C. D., Rodkey, J. A. and Alberts, A. W., *Proc. Natl. Acad. Sci. USA*, 74 (1977), 1358-1362.
 21. Inouye, S., Wang, S., Sekizawa, J., Haleboua, S. and Inouye, M., *Proc. Natl. Acad. Sci. USA*, 74 (1977), 1004-1008.
 22. Rothman, J. E. and Lodish, H. F., *Nature*, 269 (1977), 775-780.
 23. Sasaki, K., Dockerill, S., Adamiak, D. A., Tickle, I. J. and Blundell, T., *Nature*, 257 (1975), 751-757.

24. Chou, P. Y. and Fasman, G. D., *Biochemistry*, 13 (1974), 211-222; and Chou, P. Y. and Fasman, G. D., *Biochemistry*, 13 (1974), 222-245; also, Chou, P. Y. and Fasman, G. D., *J. Mol. Biol.*, 115 (1977), 135-173.
25. Levitt, M., *Biochemistry*, 17 (1978), 4277-4284.
26. Matthews, B., *Biochim. Biophys. Acta.*, 405 (1975), 442-451.
27. Dufton, M. J. and Hider, R. C., *J. Mol. Biol.*, 115 (1977), 177-194.
28. Lin, J. J. C., Kanazawa, H., Ozols, J. and Wu, H. C., *Proc. Natl. Acad. Sci. USA*, 75 (1978), 4891-4895.
29. Palmiter, R. D., Gagnon, J. and Walsh, K. A., *Proc. Natl. Acad. Sci. USA*, 75 (1978), 94-98.
30. Kauzmann, W., *Advances in Protein Chemistry*, 14 (1959), 1-63.
31. Levitt, M. and Chothia, C., *Nature*, 261 (1976), 552-558.
32. Gratzer, W. B., Beavan, G. H., Rattle, H. W. E., and Bradbury, E. M., *Eur. J. Biochem.*, 3 (1968), 276-283.
33. Moran, E. C., Chou, P. Y. and Fasman, G. D., *Biochem. Biophys. Res. Commun.*, 77 (1977), 1300-1306.
34. Chou, P. Y. and Fasman, G. D., *Biochemistry*, 14 (1975), 2536-2541.

FIGURE LEGENDS

Figure 1. The distribution of amino acids in signal sequences. Sequences of the signal peptides from the G protein of vesicular stomatitis virus (VSV); trypsinogen 2 and other canine pancreatic proteins; albumin, insulin, and growth hormone (RGH) from rat; IgG's from mouse; bovine parathyroid hormone (PTH); lysozyme, conalbumin, ovomucoid, and ovalbumin from chicken; honey bee melittin; bacterial lipoprotein and exopenicillinase; and the coat protein of bacteriophage fd are shown with their C-termini aligned.^{8,10-21} The distribution of amino acid types -- hydrophobic, uncharged polar, positively charged, and negatively charged--in the eukaryotic sequences is displayed schematically and the percentage of each type of amino acid is tabulated for each sequence. A largely hydrophobic region, from which charged residues are excluded, can be seen roughly in register with the N-termini of the sequences. The sequence of the N-terminus of ovalbumin is presented for comparison. Unidentified residues are indicated by (x). Leucine was not distinguished from isoleucine in the canine pancreatic proteins. Prokaryotic signal sequences are marked with (*).

Figure 2. Signal sequence secondary structure predictions. The predicted secondary structures of the eukaryotic signal peptides are shown with their C-termini aligned. H, S, and T indicate amino acids predicted in helix, sheet, and turn structures, respectively. Residues in turns which have a greater than average probability which is lower than the probability of

overlapping helices, sheets, or adjacent turns are indicated by t. The location of residues which are often found in bends-- glycine, proline, and aspartate -- are indicated by a squiggle when they fall outside of predicted β -turns, because β -turns are generally underpredicted by the method.²⁴ Overlapping predictions within a sequence resulted when different sets of four to six amino acids were considered sequentially. Ambiguities were not resolved by comparing the total aggregate probabilities calculated using all of the amino acids assigned to the overlapping structures. These aggregate probabilities are listed on the right. The predictions need not literally represent actual structures; they do, however, suggest that the signal sequences adopt a common fold.

Figure 3. Summary of secondary structure predictions. Helix, sheet, and turn predictions are tabulated for each position in the sequences and shown with the C-termini aligned and the N-termini aligned. The ordinate represents the number of sequences in which the indicated feature occurs, and sequence position is on the abscissa. Only the first residue of each tetrapeptide β -turn is plotted. Turn sequences with greater than average probability are distinguished from predicted turns. The overlap of helix and sheet predictions is evident. The C-terminal turn distribution is narrower with the cuts matched than with the ends matched, suggesting that this region of the signal peptides may be involved in recognition.

by the signal peptidase. The predicted secondary structure of the N-terminus of ovalbumin is represented by the blocks above each of the distributions.

Figure 1.

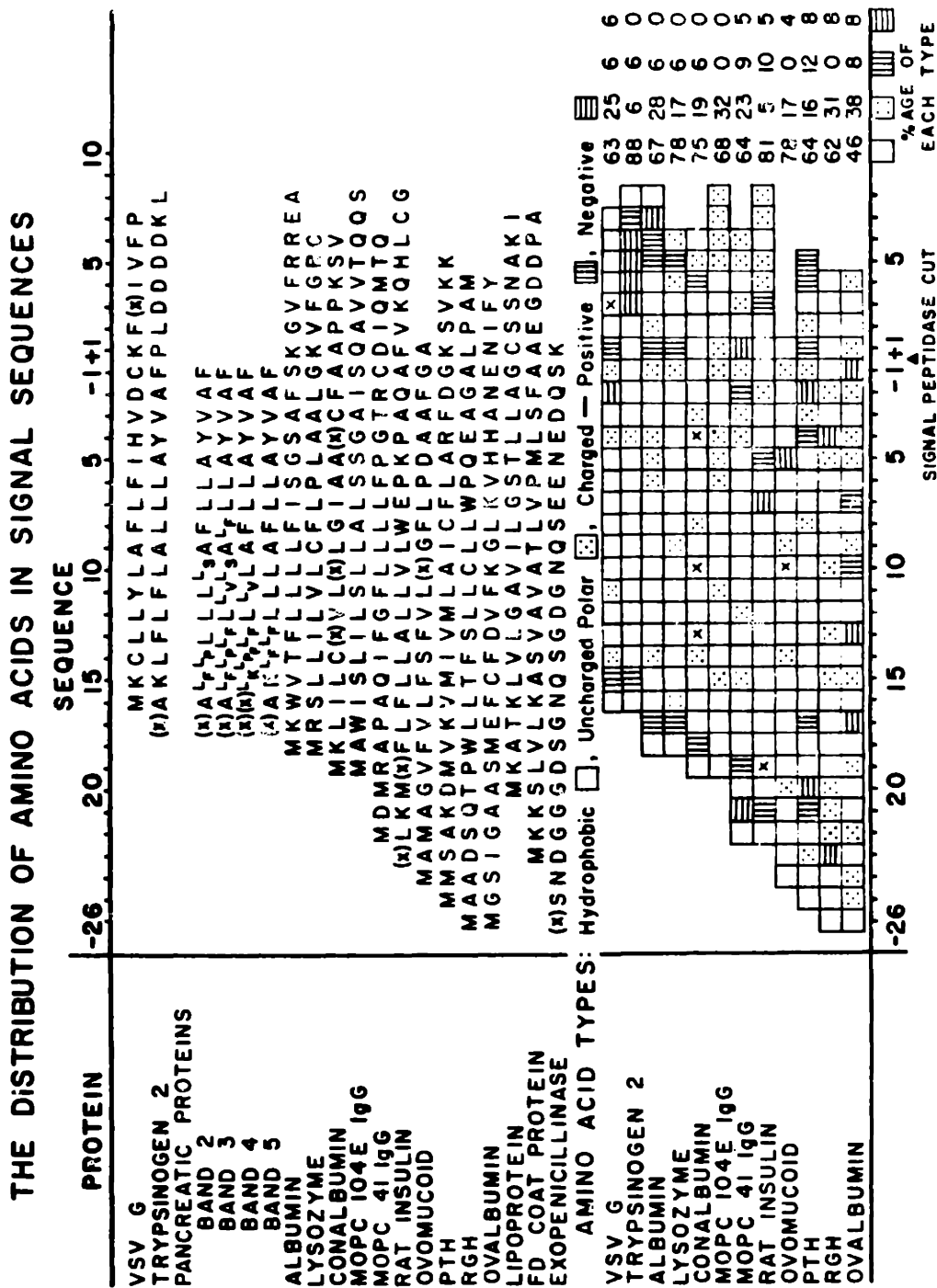


Figure 2.

SIGNAL SEQUENCE SECONDARY STRUCTURE PREDICTIONS

H = HELIX S = SHEET T = TURN † = > AVERAGE TURN
 ~ = GLY, PRO, OR ASP PROBABILITY

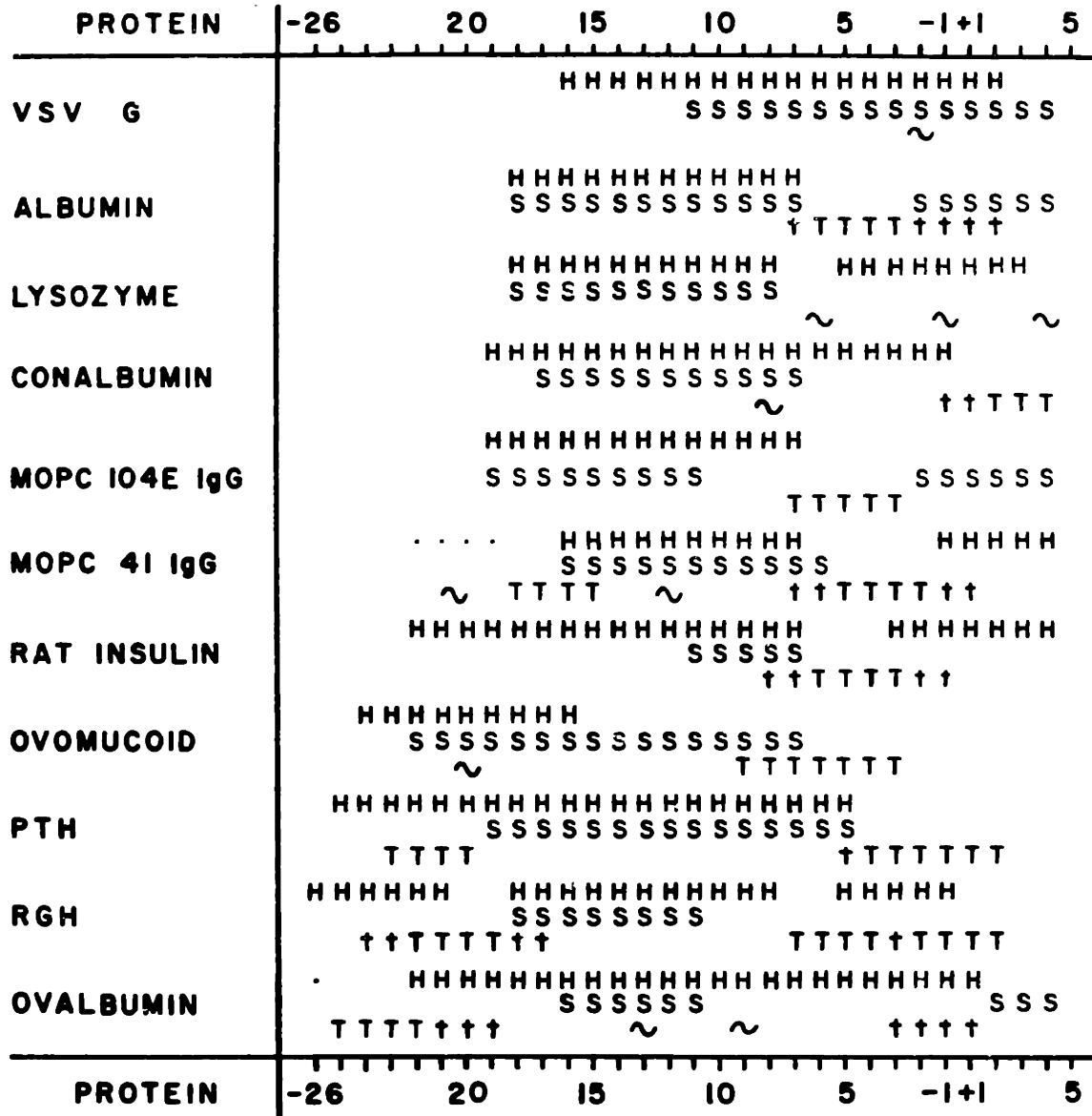
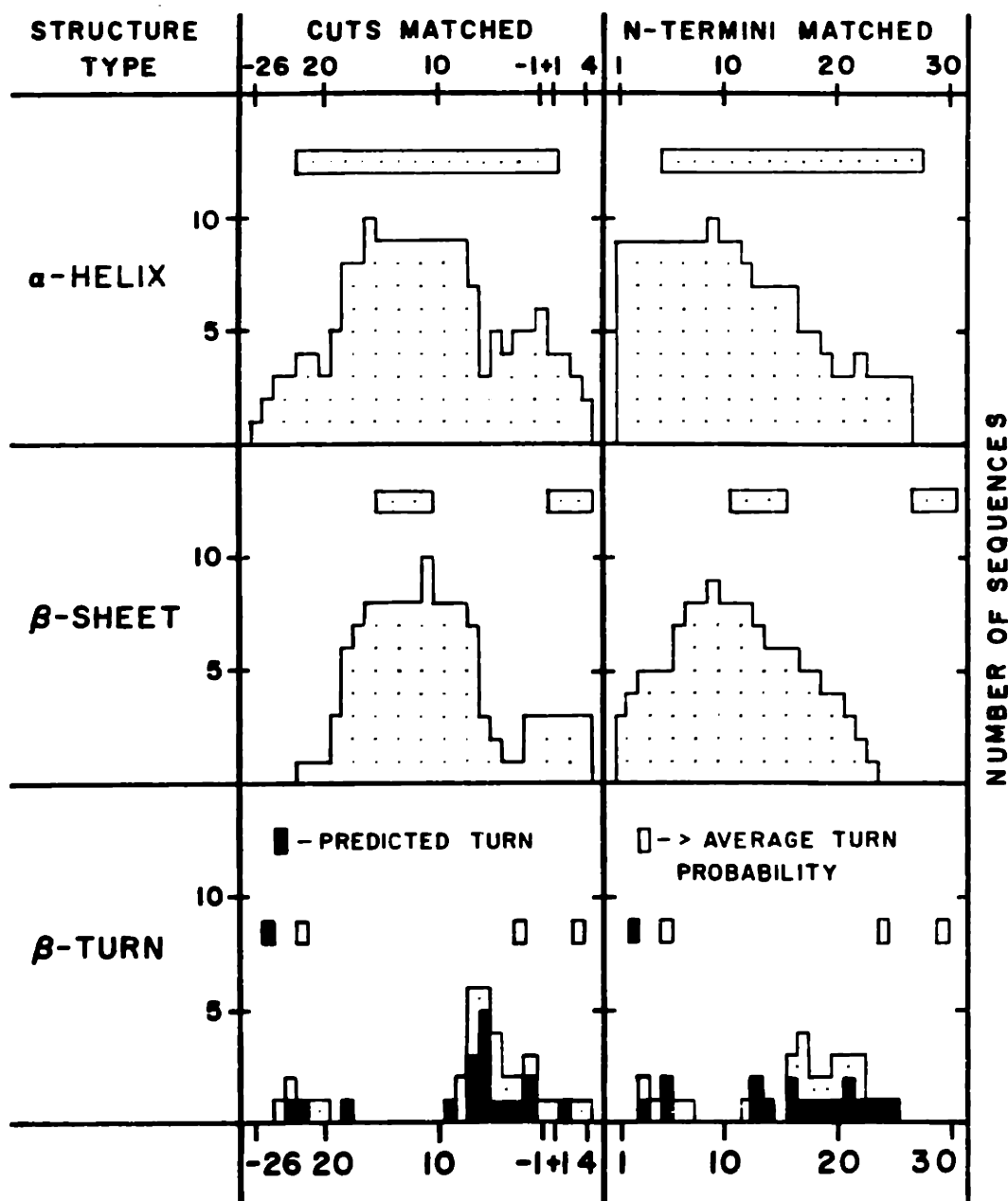


Figure 3.

SUMMARY OF SECONDARY STRUCTURE PREDICTIONS



Acknowledgements and Testimonials

"Art is . . . a longest road through life, and when I think how slight and beginnerish what I have done till now is, I am not surprised that this production (which resembles a strip of half-tilled field a foot wide) does not sustain me. For plans bear no fruit, and seed prematurely sown does not sprout. But patience and work are real and can at any moment be transformed into bread. 'Il faut toujours travailler', said Rodin whenever I attempted to complain to him about the schism in daily life; he knew no other solution, and this course had been his . . . To stick to my work and have every confidence in it, this I am learning from his great and greatly given example . . ."

--- R.M. Rilke
Letters to a Young Poet

Greg Petsko has inspired and watched and helped in a planting these last three years. Watched a process that is hidden, even to me, worked and watched patiently without imposing an end (perhaps knowing that death is at the end). What thanks could be enough? We have been lucky.

And far from alone, for other friends and colleagues have stood shoulder-to-shoulder and sung the songs of working and clapped the rhythms of the planting. Alex Rich and David Phillips have provided tools, honed sharp with intuition and intellect, and each in their own way, they have woven an image of the world which is to be fed. They have seen many seasons. They have told the tales and the histories (—"The rivers were sired by the oceans in the womb of the mountains and run between their parents like children for all time. Come, fill the buckets as others before have done. It is for you to do now.")

Glenn Kawasaki, Geoff Hendy, Hank Kronenberg, Carl Cork and Xuong Nguyen-hu have helped to drag the buckets to the fields and shouldered

our primitive plow. Their hearts are in the work of this thesis. Ned Seeman and Mary Roberts have nursed blistered hands and spirits.

Lab mates, masters of different fields, have grinned encouragement and admonition -- Will Gilbert, Sherry Mowbray, Barbara Seaton, Gary Quigley, Andy Wang, Jon King, David Goldenberg, Dan Kemp, Mitch Lewis, Jeremy Knowles, and other friends and strangers.

Thanks to Minx Fuller, Connie Cepko, Matt Scott, Gary Zieve, Bill Earnshaw, Helene Langevin, Janet Corpus, Louis Slesin, Eisa King, Michael Johnson, Jean Weinberg, Aaron Bernstein, Pam Fink and the Alber family.

The production of this thesis was done by Mary Burke, Sherry Mowbray, Dagmar Ponzi and Greg Petsko. Thanks to the Danforth Foundation and the Biology Department for financial support.

Something hidden. Go and find it.
Go and look behind the Ranges --
Something lost behind the Ranges.
Lost and Waiting for you! Go.

-- Rudyard Kipling

Never let success hide its emptiness from you, achievement its nothingness, toil its desolation. And so keep alive the incentive to push on further, that pain in the soul which drives us beyond ourselves. Whither? That I don't know. That I don't ask to know.

-- Dag Hammarskjold

Biographical Note

THOMAS CLIFFORD ALBER

Born: January 5, 1954 in Tokyo, Japan.

Age: 27.

Marital status: Single.

EDUCATION

Massachusetts Institute of Technology, 9-76 to present

University of California at Santa Cruz, 1972 to 1976, A.B. in Chemistry, 1976.

University of California at Los Angeles, 1971 to 1972, Coursework in mathematics.

HONORS and FELLOWSHIPS

National Institutes of Health Post-Doctoral Fellowship, 8-81.

Sigma Psi Grant-in-Aid of Research, 1-81.

Danforth Foundation Graduate Fellowship, 9-76 to 6-80.

Graduated UC Santa Cruz with Honors in Chemistry and College Honors, 6-76.

University of California President's Undergraduate Fellowship, 9-75 to 6-76.

UC Los Angeles, Dean's List, College of Letters and Science, 1972.

Bausch and Laum Science Award, 6-72.

PUBLICATIONS

1. Alber, T., Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, G.A., Rivers, P.S. and Wilson, I.A. (1981), *Phil Trans. Roy. Soc. B*, 293, 159-171. "On the Three-dimensional Structure and Catalytic Mechanism of Triose Phosphate Isomerase".
2. -----, Fahnestock, M., Mowbray, S. and Petsko, G.A. (1981), *J. Mol. Biol.*, 147, 471-474. "Preliminary X-ray Data for the Galactose Binding Protein from Salmonella typhimurium".
3. -----, Hartman, F.C., Johnson, R.M., Petsko, G.A. and Tsernoglou, D. (1981), *J. Biol. Chem.*, 1356-1361. "Crystallization of Yeast Triose Phosphate Isomerase from Polyethylene Glycol: Protein Crystal Formation Following Phase Separation".
4. -----, Petsko, G.A. and Tsernoglou, D. (1976), *Nature*, 263, 297-300. "Crystal Structure of Elastase-substrate Complex at -55 °C".