CRYSTALLOGRAPHY AND MUTAGENESIS OF
TRIOSEPHOSPHATE ISOMERASE

by

Elias Lolis

B.A., Chemistry
Columbia College, 1984

Submitted to the Department of Chemistry in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in Biological
Chemistry at the Massachusetts Institute of Technology.

February 1990
Copyright (c) Massachusetts Institute of Technology 1990
All rights reserved

Signature of Author ___________________________
Department of Chemistry
2 November 1989

Certified by ___________________________
Professor Gregory A. Petsko
Thesis Supervisor

Accepted by ___________________________
Glenn A. Berchtold
Chairman, Departmental Committee on Graduate Students
This doctoral thesis has been examined by a Committee of the Department of Chemistry as follows:

Prof. Joanne Stubbe
Department of Chemistry, MIT
Chairperson

Prof. Gregory A. Petsko
Department of Chemistry, MIT
Thesis Supervisor

Prof. Alexander Klibanov
Department of Chemistry, MIT

Prof. Martin Karplus
Department of Chemistry
Outside Examiner
Harvard University
ACKNOWLEDGEMENTS

It is with great pleasure that I acknowledge all those who have aided me during this five year adventure.

First and foremost, I would like to thank all members of the Petsko lab. I am most grateful for the technical help, scientific discussions, and conversations about life in general. But what I will miss most is their camaraderie and friendship during this five year battle known as graduate school. I toyed with the idea of mentioning everybody specifically with some anecdote or story about an interaction with them but decided against it when many of the "bylines" were either cryptic and understood by very few people, or personal and had no business being in this document. There is only one exception to the "no-name rule for Petsko-lab members": Greg Petsko.

I do not want to lie by stating he was excellent to work for at all times during the past five years. I have felt for Greg Petsko what Isiah Thomas felt for Bobby Knight. There were times - many times - when I wanted to hug him and tell him how much I loved him, and there were times - maybe one or two - when I wanted to put a bullet through his head. To me, that's an excellent record for a five year period.

He is a brilliant, enthusiastic, and committed scientist and teacher who cares about the people in his lab, a quality that I do not take for granted. I have learned much from him, including lessons that he probably does not know he has taught me, and I am sincerely appreciative.

There are many, many people at MIT whose resources (both intellectual and structural - that is, equipment in their lab) I have used. The list is too long to mention. They know who they are, and I thank them.

There are collaborators at Harvard who have made my last year fruitful and enjoyable. Steve Blacklow and Betsy Komives kept me busy with mutant TIM crystals. Patti Lodi and Dave Pompliano provided me with essential technical information regarding TIM. Paul Bash (of Martin Karplus's group), with his knowledge of UNIX and XPLOR, was a necessary catalyst for the refinement of mutant TIM structures that do not appear in this thesis, but make the TIM story more complete. Prof. Jeremy Knowles always made time for me during my unannounced visits. Discussions with him were always exhilarating and motivating. I am particularly grateful. The Knowles group has been an excellent group of people with whom to collaborate and play poker. Without them, both my thesis and my wallet would have been thinner.

There were a number of others, most of whom are non-scientists, who made substantial contributions by providing necessary R&R (rest and relaxation). In particular, I would like to thank one roommate (who cannot be named because he is Petsko-lab member) who has not only been an excellent roommate, but also an excellent colleague and an excellent friend. I believe I am a better person and a better scientist than I would have otherwise been. There really are no words that can express how glad I am I met him. There never can be when you find a true friend like that.

I also want to thank the following people from home who visited me and provided the necessary distraction from my work to keep me sane. I hope that by living in another city, I provided the change of pace to
also keep them sane. They are: my sister Effie and John Drakopoulos, John Palatianos - who had the uncanny ability of turning Boston into one big summer camp no matter what time of year, and whose unique outlook on life was always refreshing to me, Chris Kolefas, Tom and Helen Kolefas, Steve Raftopoulos, and my cousins Elias Lolis and Pete Sotos. I also thank my other cousins, Jerry and Steve Makris, who provided much entertainment when I went home but who never visited me in Boston. This remains a sore point with me, although it should be one for them too because they missed out on a great town. Angela Claudatos also deserves thanks for encouragement and friendship during the critical early years of graduate school. There are many other friends and relatives too numerable to mention who expressed their pride in me. Little did they know that their words provided encouragement during times of frustration.

How can I express my appreciation to the people without whom this would not have been possible? My mother has to take some responsibility for being the way I am. Some people might think she should go to jail for it, but I think she's done a decent job. It is because she always insisted on good grades ("or else...") that I first got into the habit of studying before realizing I actually liked learning. Although he never quite seemed to understand why it was that I was in school for so long, it was partly because of my dad's own hard work and success that I was inspired to continue my education. Without their love and encouragement, this thesis would have been much harder to complete. How can I express my love and appreciation? By dedicating this thesis to them.
To My Parents
Crystallography and Mutagenesis of Triosephosphate Isomerase
by
Elias Lolis

Submitted to the Department of Chemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemistry
October 1989

ABSTRACT

The structure of yeast triosephosphate isomerase has been refined against data to 1.9 Å resolution to a final R-factor of 21.0%. The final model consists of all non-hydrogen atoms in the polypeptide chain and 119 water molecules, a number of which are found in the interior of the protein. The structure of the active site clearly indicates that the carboxylate of the catalytic base, Glu 165, is involved in a hydrogen bonding interaction with the hydroxyl of Ser 96. In addition, the interactions of the other active site residues, Lys 12 and His 95, are also discussed. For the first time in any TIM structure, the 'flexible loop' has well defined density; the conformation of the loop in this structure is stabilized by a crystal contact. Analysis of the subunit interface of this dimeric enzyme hints at the source of the specificity of one subunit for another and allows us to estimate an association constant of $10^{14}$-$10^{16}$ M$^{-1}$ for the two monomers. The analysis also suggests that the interface may be a particularly good target for drug design. The conserved positions (20%) among the sequences from 13 sources ranging on the evolutionary scale from E. coli to humans reveal the intense pressure to maintain the active site structure.

The binding of the transition-state analogue 2-phosphoglycolate to triosephosphate isomerase from yeast has been investigated crystallographically. An atomic model of the enzyme-inhibitor complex has been refined against data to 2.5 Å resolution to a final R-factor of 18%. The interactions between the inhibitor and enzyme have been analyzed. The inhibitor forms strong hydrogen bonds to the side chains of His 95 and Glu 165. The latter hydrogen bond confirms that Glu 165 is protonated upon PGA binding. The structure of the complexed enzyme has been compared to that of the unbound form of the enzyme and conformational changes have been observed: the side chain of Glu 165 moves over 2 Å and a ten residue loop moves over 7 Å to close over the active site. The implications for catalysis of these results are noted.

The structure of the uncomplexed form of a mutant enzyme from chicken TIM has also been refined to an R-factor of 18% against data to 2.5 Å. The structure indicates that the conformation of the active site is slightly changed, and that the changes are not identical in the two subunits. In one subunit, the hydroxyl of Ser 96 rotates toward the shorter Asp side chain in order to maintain the hydrogen bond that is found in the wild type enzyme. In the other subunit, there is no conformational change at Ser 96. The conformation of the Asp side chain in this subunit is maintained by hydrogen bonds with water molecules. In addition, the position of the sulfate is different in this second
subunit. Conditions have also been found to co-crystallize PGH with the pseudorevertant E165D/D96P and wild type enzymes.

Finally, the cloning of the yeast triosephosphate isomerase gene into various plasmids for overexpression is described. The construction of various mutants to test hypotheses of structure and function is also described. Purification and partial kinetic characterization of a loop mutant, T172D, is given.

Thesis Supervisor: Gregory A. Petsko

Title: Professor of Chemistry
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ABSTRACT</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHAPTER 1: THE STRUCTURE, FUNCTION, AND MECHANISM OF TRIOSEPHOSPHATE ISOMERASE.</td>
<td>11</td>
</tr>
<tr>
<td>I. Introduction</td>
<td>12</td>
</tr>
<tr>
<td>II. The Chemical and Kinetic Facts</td>
<td>12</td>
</tr>
<tr>
<td>III. The Structural Facts</td>
<td>17</td>
</tr>
<tr>
<td>IV. The Enzyme Mechanism</td>
<td>22</td>
</tr>
<tr>
<td>A. The Isomerization Reaction</td>
<td>22</td>
</tr>
<tr>
<td>B. The Elimination Reaction</td>
<td>25</td>
</tr>
<tr>
<td>V. Mutagenesis and Computational Experiments to Study the Mechanism</td>
<td>26</td>
</tr>
</tbody>
</table>

| FIGURES | 33 |

<table>
<thead>
<tr>
<th>CHAPTER 2: THE STRUCTURE AND REFINEMENT OF YEAST TRIOSEPHOSPHATE ISOMERASE AT 1.9 Å RESOLUTION.</th>
<th>41</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Introduction</td>
<td>42</td>
</tr>
<tr>
<td>II. Experimental</td>
<td>43</td>
</tr>
<tr>
<td>A. The Initial Model</td>
<td>43</td>
</tr>
<tr>
<td>B. Refinement</td>
<td>43</td>
</tr>
<tr>
<td>III. Results and Discussion</td>
<td>46</td>
</tr>
<tr>
<td>A. Analysis of the Structure</td>
<td>49</td>
</tr>
<tr>
<td>B. Comparison of Subunits and Estimation of Errors</td>
<td>51</td>
</tr>
<tr>
<td>C. The Active Site</td>
<td>51</td>
</tr>
<tr>
<td>D. The Subunit Interface</td>
<td>53</td>
</tr>
<tr>
<td>E. Correlation of Sequence, Structure, and Function</td>
<td>56</td>
</tr>
<tr>
<td>IV. Conclusion</td>
<td>59</td>
</tr>
</tbody>
</table>

| FIGURES | 62 |

<table>
<thead>
<tr>
<th>CHAPTER 3: CRYSTALLOGRAPHIC ANALYSIS OF THE COMPLEX BETWEEN TRIOSEPHOSPHATE ISOMERASE AND 2-PHOSPHOGLYCOLATE AT 2.5 Å RESOLUTION: IMPLICATIONS FOR CATALYSIS.</th>
<th>79</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Introduction</td>
<td>80</td>
</tr>
<tr>
<td>II. Materials and Methods</td>
<td>81</td>
</tr>
<tr>
<td>A. Crystallization, Data Collection, and Data Reduction</td>
<td>81</td>
</tr>
<tr>
<td>B. The Initial Model</td>
<td>82</td>
</tr>
<tr>
<td>C. Refinement</td>
<td>83</td>
</tr>
<tr>
<td>III. Results</td>
<td>84</td>
</tr>
<tr>
<td>A. Comparison of Subunits and Estimation of Errors</td>
<td>84</td>
</tr>
<tr>
<td>B. The PGA-Bound Active Site</td>
<td>86</td>
</tr>
<tr>
<td>C. Conformational Changes Upon Binding of PGA</td>
<td>89</td>
</tr>
<tr>
<td>IV. Discussion</td>
<td>91</td>
</tr>
<tr>
<td>A. The Origin of Tight Binding of PGA to TIM</td>
<td>91</td>
</tr>
<tr>
<td>B. Function of the Conformational Change</td>
<td>92</td>
</tr>
<tr>
<td>C. Dynamics and Mechanism of the Conformational Change</td>
<td>96</td>
</tr>
</tbody>
</table>

| FIGURES | 100 |

<table>
<thead>
<tr>
<th>CHAPTER 4: CRYSTALLOGRAPHY OF MUTANTS OF CHICKEN TIM</th>
<th>108</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Introduction</td>
<td>109</td>
</tr>
<tr>
<td>II. Materials and Methods</td>
<td>111</td>
</tr>
<tr>
<td>A. Chicken E165D</td>
<td>111</td>
</tr>
<tr>
<td>B. In Search of Chicken TIM-PGH Crystals</td>
<td>114</td>
</tr>
<tr>
<td>1. Crosslinked E165D crystals soaked with PGA or PGH.</td>
<td>115</td>
</tr>
</tbody>
</table>
2. Crystallization of PGH complexed to wild type enzyme and the S96P/E165D pseudorevertant.
III. Results
A. Estimation of Errors in the E165D Structure
B. Subunit 1
C. Subunit 2
IV. Discussion
FIGURES

CHAPTER 5: MOLECULAR BIOLOGY OF YEAST TRIOSEPHOSPHATE ISOMERASE
135
I. Introduction
II. Experimental Procedures
A. Materials
B. Methods
2. Preparation of M13 phage and single stranded DNA.
4. Mutagenesis reactions.
   (a) T172D
   (b) C126S
   (c) ΔLoop (170-174)
5. Preparation of double stranded DNA.
III. Results
A. Cloning of yTPI Gene Into M13mp18
B. Construction of Various Expression Vectors
   1. pUCyTPI(T172D)
   2. pBS-Δ
IV. Appendix to Chapter 5: Purification and Preliminary Kinetics of Loop Mutants
A. Purification of T172D
B. Purification of ΔLoop
C. Preliminary Kinetics of T172D
FIGURES
REFERENCES

APPENDIX: TRANSITION-STATE ANALOGUES IN PROTEIN CRYSTALLOGRAPHY: PROBES OF THE STRUCTURAL SOURCE OF ENZYME CATALYSIS
161
I. Introduction
II. Perspective and Summary
III. Transition-State Theory and Enzyme Catalysis
   A. Implications and Paradoxes of Transition-State Theory
IV. The Structural Source of Catalysis for Various Enzymes
   A. Proteases
   B. Triosephosphate Isomerase
   C. Aspartate Transcarbamylase
   D. Ribonuclease
   E. RuBisCo
V. Conclusion
VI. References
FIGURES
<table>
<thead>
<tr>
<th>Abbreviations</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>DHAP</td>
<td>Dihydrc yacetone phosphate</td>
</tr>
<tr>
<td>FTIR</td>
<td>Fourier transform infrared</td>
</tr>
<tr>
<td>GAP</td>
<td>Glyceraldehyde-3-phosphate</td>
</tr>
<tr>
<td>N.M.R.</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>TIM</td>
<td>Triosephosphate isomerase</td>
</tr>
<tr>
<td>PGA</td>
<td>2-Phosphoglycolate</td>
</tr>
<tr>
<td>PGH</td>
<td>Phosphoglycolohydroxamate</td>
</tr>
<tr>
<td>u.v.</td>
<td>ultraviolet</td>
</tr>
</tbody>
</table>
CHAPTER 1

THE STRUCTURE, FUNCTION, AND MECHANISM OF
TRIOSEPHOSPHATE ISOMERASE

A writer doesn't often tell a reader anything the reader doesn't already know or suspect. The best the writer can do is put the idea in words and by doing that make the reader aware that he or she isn't the only one who knows it. This produces the warm bond between reader and writer that they're both after because it feels so good.

—Andy Rooney
I. INTRODUCTION

Triosephosphate isomerase (TIM; E.C.5.3.1.1) is a dimer of identical subunits, each of molecular weight about 26,000. TIM is a glycolytic enzyme that catalyzes the interconversion of the two products of the aldolase reaction, D-glyceraldehyde-3-phosphate (GAP) and dihydroxyacetone phosphate (DHAP) (Figure 1). Since only GAP proceeds along the glycolytic pathway to the Krebs cycle (Figure 2), TIM is absolutely essential for efficient energy production. People who are defective in only one of the alleles for this enzyme are found to suffer from chronic anemia and neuromuscular disorders (Valentine et al., 1983).

There is a long history of biophysical and chemical studies on TIM. This history now includes exhaustive kinetic and chemical investigations, x-ray crystallography, folding studies, computer simulations and random and site-directed mutagenesis. As a result of these powerful and complementary techniques, we are now in a situation where a great deal is known about the structure and function of this enzyme. This chapter is an attempt to review the structure and function of TIM and the mechanism that has been formulated based on the above experimental methods.

II. THE CHEMICAL AND KINETIC FACTS

In 1959 it was determined that the isomerization reaction proceeds through an enediol(ate)\(^1\) intermediate (Figure 3a) (Rieder and Rose,

---

\(^1\) It has not yet been conclusively determined whether the intermediate that is formed is an enediol or an enediolate
1959). It was further pointed out that TIM provides a number of mechanistic advantages for study that other researchers have since exploited.

In a tour de force, the free energy profile of this enzyme from chicken muscle was determined using a combination of isotopic methods (Albery and Knowles, 1976). The profile indicated the reaction proceeded through four steps: substrate binding, proton abstraction to form the intermediate, protonation of the intermediate to form product, and product release. The individual rate constants, as well as the pseudo second-order rate constant, \( k_{cat}/K_m \), indicated that the reaction was catalyzed at the diffusion-controlled rate; the slowest step of the reaction in the thermodynamically favored direction was binding of substrate (GAP) to the enzyme. Subsequent viscosity-dependence experiments confirmed this observation (Blacklow et al., 1988). In 1977, a set of equations was derived that allowed enzyme kinetic data to be used to assess the efficiency of any enzyme. Using this criteria, it was concluded that TIM had reached evolutionary perfection (Albery and Knowles, 1977).

There are two physical events that occur during the catalytic cycle whose rates deserve discussion. There is substantial evidence that a major conformational change occurs during catalysis. A ten residue loop moves about 8 Å to close over the active site after substrate has bound (Phillips et al., 1977a). For product to be released, it must again move 8 Å. This process occurs 8,700 times per second (\( k_{cat} \)) with D-GAP as substrate (Nickbarg and Knowles, 1987). Although an initial look at these numbers might suggest the loop moves at an astounding speed, the average velocity is only about 0.05 meters per hour. The dynamics of
loop movement has been studied using molecular dynamics simulation (Brown and Kollman, 1987; Joseph et al., 1989).

The other physical event relates to the interconversion between two forms of the unliganded enzyme. It has been shown experimentally that the enzyme exists in a form that binds and isomerizes DHAP and a form that binds and isomerizes GAP. These two forms interconvert with a rate constant of about $10^6 \text{s}^{-1}$ (Raines and Knowles, 1987). The structural difference between these two forms is unknown, although it has been postulated the difference is due to the ionization state of the active site.

Triosephosphate isomerase is also of historical importance because it is one of the first enzymes for which a transition-state analogue was proposed (Wolfenden, 1969). The enzyme has been subjected to inhibition studies by a large number of molecules, from which much mechanistic data have been derived. A number of active-site-directed inhibitors have been used to identify Glu 165 as the catalytic base that abstracts and transfers the proton back and forth between the two carbon atoms (Hartman, 1971; de la Mare et al., 1972; Corran and Waley, 1973). $^{31}$P N.M.R. studies (Campbell et al., 1979) of the transition-state analogue, 2-phosphoglycolate (PGA) (Figure 3b), and studies with substrate analogues (where the dianionic phosphate is replaced by the monoanionic phosphonate or sulfate) (Belasco et al., 1978) indicate that substrates bind productively to the enzyme as phosphate dianions. The weak binding of phosphonate analogues of the substrate and of PGA (where the $O_3$ oxygen bridging the phosphorus to the $C_3$ atom has been replaced by a $\text{CH}_2$) has also implicated interactions between the phosphate bridging
oxygen and enzyme groups as the source of substrate specificity (Dixon and Sparkes, 1974; Belasco et al., 1978).

Since the interconversion of DHAP and GAP involves the transfer of a proton from one carbon to another as well as from one oxygen to another (Figure 1), the reaction can be considered in terms of acid-base chemistry. The pKₐ of the substrate carbon acid, either GAP or DHAP, has never been measured but can be estimated from analogy to similar compounds to be at least 15 (Alber et al., 1987b). The pKₐ of the hydroxyl group on the substrate is much less than this. It is known the protons on the hydroxyl rapidly exchange with solvent protons, and from analogy with other compounds can be assumed to have a pKₐ less than 10.

The usual side chain pKₐ of glutamic acid is about 4. An active-site-directed inhibitor, chloroacetol sulfate, does not bind productively to the enzyme (due to the monoanionic sulfate group) but nevertheless derivatizes it in a pH-dependent manner with a pKₐ of about 4 (Hartman et al., 1975), consistent with the pKₐ of a solvent-exposed glutamic acid. The reason the distinction for non-productive binding of this inhibitor must be made is that productive binding by other inhibitors or substrates to this enzyme changes the pKₐ's of various enzyme groups. The assignment of pKₐ's to specific residues based on pH-dependence experiments must always be done tentatively, if at all (Knowles, 1976). The existence of an inhibitor that does not bind to the enzyme to any extent, but nevertheless modifies a group in a pH dependent manner allows the assignment of a pKₐ to the side chain functional group to be made with a little more certainty. For proton transfer to occur, the enzyme must reduce the discrepancy between the
pKₐ's of the reacting chemical groups. It must find a way to lower the pKₐ of the substrate carbon acid and raise the pKₐ of Glu 165.

Various other pH-dependence experiments have revealed a number of other chemical facts. The enzyme operates optimally at around neutral pH. The dependence of $k_{cat}/K_m$ is bell-shaped with pKₐ's of 6 and 9 (for the chicken muscle enzyme) (Plaut and Knowles, 1972). Since the substrates have second ionization constants at pH 6.0 and 6.3 respectively, there had been some question about the chemical group responsible for the lower pKₐ. The pH dependence of $k_{cat}/K_m$ at around pH 6 could have been due to only substrate ionization, to only ionization of an enzyme group, or to a combination of the two. This controversy was resolved when pH-dependence studies of unnatural substrates (which had pKₐ's different from 6.0) showed that only a mechanism where the ionization of both an enzyme group and the substrate are catalytically important is consistent with the available kinetic data (Belasco et al., 1978). Furthermore, the same study showed that the pH-dependence of $k_{cat}$ arises from the liganded enzyme. It was concluded that upon substrate binding the pKₐ of Glu 165 is raised from 4 to 6 (from 4 to 4.6 in the yeast enzyme (Hartman and Ratrie, 1977)).

The pKₐ of the enzyme-bound carbon acid has never been determined but the enzyme obviously lowers it significantly. The presence of a carbonyl α to the target C-H bond could decrease the barrier for removal of a proton through polarization. Various studies have suggested the enzyme does just this. The rate of reduction of enzyme-bound DHAP by sodium borohydride is increased nine-fold relative to that of free DHAP in solution, even though access of borohydride to DHAP must be partially hindered by the active site (Webb and Knowles, 1974).
Interestingly, enzyme-bound GAP is protected from borohydride reduction, indicating a difference in the accessibility to the various substrate atoms in the active site (Webb and Knowles, 1975). Polarization of the carbonyl was also confirmed spectroscopically. FTIR spectroscopy revealed that the carbonyl stretching frequency in enzyme-bound DHAP decreases 19 cm\(^{-1}\) relative to free DHAP (Belasco and Knowles, 1980).

In 1984 the uncatalyzed, hydroxide-catalyzed, and buffer catalyzed isomerization reaction was studied in solution (Richard, J.P., 1984). The most surprising aspect of the study was that the enediol(ate) intermediate formed by proton abstraction from either GAP or DHAP to form the enediol(ate) decomposes to methyl glyoxal and inorganic phosphate (Figure 3c). In the uncatalyzed reaction, the partition ratio of enediolate between protonation and phosphate elimination is 0.0083. Apparently, in solution, the intermediate is more prone to undergo phosphate elimination than reprotonation to complete the isomerization reaction. TIM is a much more perfect catalyst than initially thought. The enzyme catalyzes, at the diffusion-controlled rate, the isomerization reaction through the enediol(ate) intermediate, while it disfavors the phosphate elimination reaction. Evolutionary pressure to maintain a very high catalytic rate has thus been applied from two vantage points. Not only does the quick production of energy for the "flight or fight response" require a high turnover for the isomerization reaction, but neither DHAP nor GAP can be allowed to accumulate because they are readily decomposed into toxic methyl glyoxal by simple bases that are present \textit{in vivo} (Riddle and Lorenz, 1968; Cooper, 1984).

\textbf{III. THE STRUCTURAL FACTS}
The structure of the chicken muscle enzyme was determined to 2.5 Å resolution in 1975 (Banner et al., 1975). It first showed the overall architecture of the prototypical α/β-barrel enzyme. However, the precise interactions between protein and substrate atoms could not be described since the enzyme had only water and a sulfate ion in each active site of the dimer. It was suspected that the sulfate ion might bind near the position occupied by the phosphate group of a substrate, and indeed this sulfate is in the vicinity of the catalytic base Glu 165.

One of the early ambitions for the TIM project was to crystallographically observe the enzyme-substrate complex during the isomerization reaction. The complex formed with DHAP diffused into chicken TIM crystals was studied at 6.0 Å resolution in 1977 (Phillips et al., 1977a). Due to crystal contacts, the substrate had access to only one of the two subunits. Extensive conformational changes around the positions of residues 167-176 were observed, but the precise interactions between substrate and enzyme again could not be discussed due to the absence of high resolution reflections. Because chicken TIM crystals suffer from a combination of poor diffraction and frequent nonisomorphism, crystals of the yeast enzyme more suitable for high resolution diffraction studies were grown. Two monoclinic crystal forms, based on the presence or absence of the transition-state analogue PGA were obtained (Alber et al., 1981a). The native uncomplexed enzyme was solved to 3.0 Å resolution (Alber, 1981) and refined to 1.9 Å resolution (Chapter 2, this thesis). The structure of the complex between the enzyme and phosphoglycolohydroxamate (PGH) (Figure 3b), thought to be an intermediate analogue, was also solved and refined to
1.9 Å resolution (Davenport et al., 1989). The complex of 2-phosphoglycerolate was studied to 2.5 Å resolution (Chapter 3, this thesis). The complexed enzyme structures show in detail the interactions between enzyme functional groups and inhibitors. Comparison of the uncomplexed and complexed forms of this enzyme reveals in greater detail the conformational change of the flexible loop, and also a smaller yet significant conformational change involving Glu 165 (Chapter 3). A hypothesis about the structural basis of catalysis has been formulated from these coordinates and other mechanistic data. These more accurate coordinates also allow the TIM reaction to be more rigorously analyzed by computer simulation. All these facets of the mechanism are described in the chapters of this thesis.

The structure of TIM from a pathogenic organism has also been studied. Trypanosoma brucei is a parasite that causes sleeping sickness (African trypanosomiasis) in humans over a wide range of Africa. This organism has the property of having an extremely rapid rate of glycolysis in an intracellular compartment (the glycosome) where all the glycolytic enzymes are stored (Oppordoes and Borst, 1977). It has been shown that disruption of the glycolytic pathway kills the organism (Clarkson and Brohn, 1976; Fiarlamb et al., 1977). For this reason, a program to structurally characterize glycosomal enzymes as part of a rational drug design process has been initiated. To this end the structure of TIM to 2.4 Å resolution has been determined (Wierenga et al, 1987). The structures of the enzyme with four other inhibitors that have been soaked into the crystals are also under active investigation (Noble and Wierenga, 1989).
Other biophysical studies have also been important in gathering structural data. Various spectroscopic experiments predicted the presence of conformational changes upon substrate and inhibitor binding (Johnson and Wolfenden, 1970). Refolding studies as a function of enzyme concentration indicated that only the dimer is active (Waley, 1973), even though cooperativity between the two active sites has never been observed. In light of the structure, it is now understood why only the dimer is active. Although each active site is composed of residues from its own subunit, the other subunit forms a wall upon which the active site rests. The structural integrity of the active site presumably cannot be maintained upon disruption of this wall.

A site-directed mutagenesis experiment, using the x-ray structure as a guide, also corroborates this conclusion. A residue buried at the hydrophobic interface between the two subunits, Asn 78, was converted to the isosteric aspartic acid. At pH 6.0, $k_{cat}$ of this mutant is two-thirds that of wild-type enzyme. At pH 7.9, where presumably Asp 78 is fully ionized, the mutant enzyme is not stable enough to detect any activity. The uncompensated anion in a hydrophobic environment probably forces the dimer apart. To further investigate whether the monomer is active, the mutant was incubated in the presence of a lyotropic salt, tetrabutylammonium bromide. It was reasoned that the association of the bulky cation with the charged Asp 78 residue at the interface would stabilize the monomer. A chromatographic elution profile in the presence of the salt indicated the mutant eluted as a mixture of dimer and monomer. Under these conditions, the mutant retains just 5% residual activity, whereas the wild-type enzyme exhibits a 10-fold
increase in activity (Casal et al., 1987). The reason for this increase in activity for the wild type enzyme is not known.

The question of monomer/dimer activity is very important from the point of view of drug design. There are now sequences of TIM from 13 organisms ranging from E. coli to humans. Of the roughly 250 amino acids in each subunit, about 40 are strictly conserved among all these organisms. When these residues are highlighted in the context of the three dimensional structure most fall on or around the active site. With the active sites of these organisms so similar, it is unlikely that an active site-directed inhibitor with significant specificity for a pathogenic TIM will ever be found. The subunit interface of TIM from these organisms is a bit more diverse, however, and thereby could provide a potential target of an inhibitor that could both be specific and knock out activity.

The dimer requirement for activity has also been exploited to make TIM more stable to irreversible thermal inactivation. Replacement of asparagine residues at the interface (Asn 14 and 78) with approximately isosteric but non-deamidating amino acids (Thr, Ile) makes the enzyme more resistant to irreversible heat-induced denaturation. To show that asparagine deamidation could dramatically affect the stability of the enzyme, the Asp 78 mutant (the product of deamidation of Asn 78) was characterized. This mutant has a a shorter half life at 100 °C, a decreased temperature for reversible unfolding, and reduced stability against dilution-induced dissociation (Ahern et al., 1987). Double non-deamidating mutants of two asparagine residues at the interface (N14T/N78I) doubled the half life of the enzyme at 100° C, pH 6 from 13 to 25 minutes.
IV. THE ENZYME MECHANISM

From all of the above data, a mechanism for the TIM reaction has been formulated. The description of the mechanism is divided into two parts. First, the features responsible for catalyzing the isomerization reaction are discussed. Then, the reasons why the phosphate elimination reaction is disfavored are considered.

A. The Isomerization Reaction

GAP binds to the enzyme in its unhydrated (Trentham et al., 1969), dianionic form at the diffusion-controlled rate. Based on studies with various inhibitors, there appear to be two steps in productive substrate binding. The simple encounter complex that is formed during the diffusion process is followed by a reorientation of enzyme and/or substrate in the active site. For dianions, the encounter complex is rate-limiting, while for monoanions, the reorientation is rate-limiting (Belasco et al., 1978). Crystallographic and spectroscopic studies indicate that part of this reorientation might be due to the loop and Glu 165 conformational changes. It is not known whether the two changes occur independently, or are linked and occur simultaneously. It is known, however, that upon substrate binding Glu 165 breaks its hydrogen bond to the side chain of Ser 96 moves 2 Å toward the substrate for proton abstraction and transfer. The loop, which in crystals of TIM from other species has been shown to be disordered in the uncomplexed form but which is stabilized by a crystal contact in yeast TIM, moves over 8 Å to close over the substrate, forming van der Waals and hydrogen bonding interactions with the phosphate group. This change also has the consequence of removing bulk solvent from the substrate and active site.
A Connolly surface calculation (which measures the points of contact between water and the surface of the protein) of the PGA-complexed enzyme indicates that all of the substrate except a small portion of the phosphate is inaccessible to bulk solvent. This picture of the enzyme appears inconsistent with isotope exchange experiments that have the labile substrate proton exchanging with bulk solvent, and with reduction of TIM-complexed DHAP by sodium borohydride. This paradox can be reconciled by two explanations. The protein could breathe during the catalytic process, allowing individual water molecules (and borohydride) into the active site. Access into all areas of the active site is limited as evidenced by the inability of borohydride to reduce enzyme-complexed GAP, whose carbonyl group lies deeper into the active site. Alternatively, the labile proton might exchange with a small pool of water molecules that are known to be trapped in the active site after the loop has closed. Nevertheless, binding of substrate and closure of the loop makes the active site practically nonaqueous with the removal of bulk solvent. In a computer modelling experiment, the loop in the PGA-complexed enzyme was replaced with the conformation of the loop in the uncomplexed enzyme and a Connolly surface was calculated. Residues deep into the active site, as well as the inhibitor, were totally accessible to bulk solvent in this "hybrid" protein (Chapter 3).

A nonaqueous (or more hydrophobic) active site increases the strength of the electrostatic interactions within it. Polarization of the substrate carbonyl is known to be important in catalysis (Webb and Knowles, 1974; 1975; Belasco et al., 1980). The structure implicates \( \alpha \)-helix D, (residues 95-102) and the side chains of His 95 and Lys 12 as playing polarizing roles (Figure 4). The axis of the \( \alpha \)-helix extends
from His 95 at the substrate binding site to Phe 102. The helix
macrodipole moment (Hol et al., 1978) is believed to contribute to the
carbonyl polarization by placing an effective partial positive charge at
the amino terminal end, which is precisely where the substrate
(cis-bound) oxygens are found. The imidizole of His 95 might act as a
conduit for this dipole moment as it is found approximately equidistant
to both oxygens in the TIM-PGH structure. In this position, it can
polarize the carbonyls regardless of the direction of the isomerization
reaction.

There has been some controversy regarding the mechanism by which
His 95 polarizes the carbonyl. It can do so electrostatically by
forming a hydrogen bond to the carbonyl or by acting as a general acid
and parting with its proton. This question is equivalent to that of the
existence of an enediol or enediolate intermediate. If the imidazole
acts as an acid, then it is likely that an enediol intermediate is
formed. If it interacts electrostatically (via hydrogen bonding) then
presumably an enediolate is formed. Analysis of the potential hydrogen
bonding pattern of the PGA-complexed enzyme suggests that polarization
occurs through electrostatic interactions. This analysis is discussed
in greater detail in Chapter 3. A site-directed mutagenesis experiment
has also been used to address this question, and this is discussed in a
later section of this chapter.

Lys 12 is also found on the side of the substrate oxygens and is
within hydrogen bonding distance (3.0 Å) of the O2 oxygen. The enzyme,
by placing all these electrophilic groups on the side of the active site
where the substrate oxygens are bound, provides an electropositive
environment that polarizes the carbonyl and weakens the adjacent C-H bond.

On the other side of the active site, next to the position that would be occupied by the labile proton, is where the carboxylate of Glu 165 is found. In the PGH-complexed enzyme, one of the glutamate oxygens forms a hydrogen bond to the N₂ nitrogen of PGH. In fact, this nitrogen is believed to be deprotonated by Glu 165 to form an analogue of the enediol(ate) intermediate (Figure 3b) (Davenport, 1985). The base is not equidistant to both the N₁ and C₂ atoms, as might have been expected for transfer between these two positions (for the normal substrate), but the presence of the PGH nitrogen instead of the substrate carbon might bias the conformation due to the potential hydrogen bonding interaction. Even if the enzyme adopts this side chain conformation when substrate is bound, the carboxylate probably has enough conformational flexibility to transfer the proton between the two carbon atoms.

B. The Elimination Reaction

While the hydrophobic environment increases the strength of the electrostatic interactions responsible for isomerization, it also makes the elimination reaction unfavorable. For phosphate elimination to occur, the bridging oxygen must either become negatively charged or be protonated. However, there are no functional groups that can hydrogen bond to the bridging oxygen. Phosphate elimination would place an uncompensated, and therefore highly unstable, negative charge on this oxygen. Besides this electrostatic effect, the conformation of the bridging oxygen is also important in preventing the phosphate elimination reaction. From stereoelectronic considerations, phosphate elimination is most favored when the overlap between the C₁ sp³ orbital
and the enediol π orbital system is maximized. The conformation of the bridging oxygen (O₃) directly affects the relative orientation of the C₃ sp³ orbital to the π system. When the distance between the O₃ oxygen and enediol plane is at a maximum, the overlap among the orbitals is also at a maximum and phosphate elimination is favored. When the O₃ oxygen is in the enediol plane, the overlap is minimized and elimination is disfavored. These factors are discussed in Chapter 3.

V. MUTAGENESIS AND COMPUTATIONAL EXPERIMENTS TO STUDY THE MECHANISM

The cloning and over-expression of the genes for TIM from chicken and yeast have provided the opportunity to test the various mechanistic hypotheses. One of the mechanistic questions that has yet to be resolved is the nature of the intermediate and the role of His 95. The three possible roles of His 95 are shown in Figure 5. His 95 might initially be an imidazolium and act as an acid to stabilize the enediol intermediate through hydrogen-bond interactions (Figure 5a). Alternatively, the histidine could be uncharged and stabilize the enediolate through hydrogen bonding interactions (Figure 5b). As a last possibility, the uncharged imidazole might act as an acid to form an enediol intermediate and imidazolate (Figure 5c). The last possibility has been considered to be remote due to the high (conventional) pKₐ of the imidazole proton and the lack of evidence for an imidazolate in any other protein. To address this question, His 95 was converted to glutamine with the expectation that the consequences of the mutation would reveal the role of His 95 (Davenport, 1985). If polarization by the imidizole is purely by electrostatic means (i.e., through hydrogen bonding interactions), the mutation should not have drastic effects,
since the glutamine side chain should also be able to make the same hydrogen bond. If the histidine functions as an acid during catalysis, the mutant should have dramatically decreased activity, since it is unlikely that glutamine could part with a proton. The H95Q mutant was found to be 400-fold less active than the wild-type enzyme. More rigorous kinetic studies revealed that the catalytic mechanism for this mutant changes from a one base to two base mechanism (Figure 6), with a substrate oxygen functioning as one of the bases (Nickbarg et al., 1988). Assuming the interactions between Gln 95 and substrate are conserved (i.e., assuming the conformation of Gln 95 resembles that of His 95), this result strongly implies an acidic role for the wild type His, as the glutamine is apparently unable to stabilize the intermediate to the same extent.

The effects of this mutation were also analyzed using combined quantum and molecular dynamics. Assuming its side chain conformation is similar to that of the wild type enzyme, Gln 95 should be as effective as histidine in stabilizing the enediolate intermediate (and presumably the transition state) by hydrogen bonding. However, molecular mechanics simulations indicate that Gln 95 should adopt a different conformation (Alagona et al., 1984). The structure of this mutant is currently under investigation.

In a different quantum mechanics/molecular dynamics simulation, the three possible roles of His 95 were investigated (Bash et al., 1989). It was found that the mechanism depicted by Figure 5a was too stable. A free energy well which the enzyme would have difficulty overcoming was formed from the interaction between an uncharged imidazole and an enediol intermediate. The other two possibilities, the imidazole-
substrate/imidazolate-enediolate (Figure 5b) mechanism and the
imidazole-substrate/imidazolate-enediol mechanism (Figure 5c), produced
more reasonable energies. The more favorable mechanism could not be
determined due to the similarity in the energetics of the two.

Each of the chemical steps in the reaction was also simulated using
quantum mechanics (Alagona et al., 1989; Bash et al., 1989). The
results of one study (Bash et al., 1989) proposes roles for conserved
residues that are not directly in contact with substrates, transition
states, or the intermediate during the reaction. Some of these residues
form long-range, electrostatic interactions that are partly responsible
for setting up the proper environment for catalysis to occur. The other
study (Alagona et al., 1984) confirms that Lys 12 and His 95 contribute
to the stabilization of the intermediate. This stabilization has the
consequence of lowering the barrier between DHAP and intermediate to
10-15 kcal/mole from about 25 kcal/mole in the uncatalyzed reaction.
The distance between the carboxylate oxygen and the substrate carbon was
shown to have a dramatic effect on the energetics of proton abstraction.
As long as the (Glu 165)-O$^-$...C$_1$-(DHAP) distance is less than 3.2 Å,
reasonable energies for the proton abstraction process are obtained
(Alagona et al., 1984).

In an effort to correlate experimentally the distance between the
catalytic base and the substrate with catalytic activity, Glu 165 has
been mutated to an aspartic acid. This change results in a mutant that
is 500-fold less active than wild-type enzyme (Raines, et al., 1986).
Computer simulation of the energetics of the isomerization reaction for
mutant gives similar results (Alagona et al., 1986). But the distance
between the catalytic carboxylate and substrate is not the only
interaction that is affected by this mutation. In the uncomplexed enzyme, the carboxylate of Glu 165 forms a hydrogen-bond interaction with the hydroxyl group of Ser 96. The mutation to an Asp causes some subtle rearrangements in the active site of the uncomplexed enzyme that are discussed in more detail in Chapter 4.

It was previously mentioned that almost 20% of the TIM residues from 13 different sources are strictly conserved, and found to be positioned around the active site. Aside from their possible role in long-range electrostatic interactions to stabilize various species along the reaction pathway (Bash et al., 1989), some of these residues are also responsible for maintaining a precise configuration for those residues that do make contact with the substrate. TIM was probably perfected early in evolution, and the selective pressure to maintain the perfect active site geometry must have been enormous. With that kind of history, it is no wonder that no other active sites of TIM have ever arisen. The gene for the E165D mutant has been used in a series of experiments aimed at doing what nature has not: finding alternative active site geometries with high isomerase activity (Hermes et al., 1987).

Since the relationship among amino acid sequence, structure, and chemical reactivity is not known, it is not possible to predict which other mutations could compensate for the presence of an aspartic acid at position 165. It was therefore necessary to subject the E165D mutant gene to random mutagenesis followed by selective screening to isolate pseudorevertants (those double mutants with increased catalytic activity relative to the E165D single mutant). The random mutagenesis procedure ensured that (i) the mutation at position 165 would be maintained, (ii)
every other position would be subjected to random variation, and (iii) double mutants (with E165D counting as one mutation) would predominate over all other types of mutants (single, triple, etc.). The selective screening method exploited the nutritional requirements of a strain of *E. coli* (DF502) harboring different triosephosphate isomerase activities. Cells with total TIM activity could grow on glycerol while those without or with partial TIM activity (as the E165D mutant) could not. Cells transformed with plasmid containing the E165D gene subjected to the random mutagenesis procedure that produced a new gene coding for an enzyme with increased activity could therefore grow on glycerol (Hermes et al., 1989). Using this strategy, six pseudorevertants have been isolated. All but one mutation occur at positions that are normally strictly conserved in all other TIM sequences. The pseudorevertant with the highest activity, E165D/S96P, involves the elimination of the hydroxyl group that binds to the catalytic carboxylate in the uncomplexed enzyme. How the side chain of Asp 165 responds to the absence of the hydroxyl group remains to be seen. Furthermore, Pro 96 is next to the important His 95. Interestingly, a residue that normally disrupts helices, proline, is placed in the D₁ helix, whose dipole moment is thought to aid in the polarization of the substrate carbonyls. Although the mutation at this position increases the activity of the enzyme, the pseudorevertant is still 10-fold less active than the wild type enzyme. Another round of random mutagenesis and selection might produce triple mutants with even higher catalytic activity.

The role of the flexible loop has also been subjected to experiment using site-directed mutagenesis. In one of the earlier, unrefined
structures of TIM, Thr 172 was modelled to suggest that its side chain hydroxyl would interact with the phosphate moiety of the substrate. Conversion of this threonine to an aspartic acid resulted in a ten-fold decrease in catalytic activity and no detectable production of methyl glyoxal using a rather insensitive assay (Chapter 5). The refined structure of the enzyme complexed to PGA (and PGH) revealed that only the backbone amide nitrogen of Gly 171 formed a hydrogen bond to the phosphate. The side chain of Thr 172 was found to be pointing away from the active site.

In independent experiments, portions of the loop were deleted from the chicken (Pompliano and Knowles, 1989) and yeast (Chapter 5) genes. In the yeast enzyme, five amino acids spanning positions 170 to 174 were deleted, while in the chicken enzyme four residues (170 to 173) were deleted. Both enzymes had very little catalytic activity. The chicken mutant has been more fully characterized (Pompliano et al., 1989). The free energy profile for this mutant indicates that all chemical species (substrate, intermediate, and product) were bound much less tightly than in the wild-type enzyme. Catalytic activity decreases by 10,000.

Interestingly, for every one molecule of DHAP that the mutant enzyme produces, five molecules of methyl glyoxal are made. In the wild-type enzyme, 10,000 molecules of GAP are processed to DHAP before even one molecule of methyl glyoxal appears. Removal of the flexible loop converts the enzyme from a triosephosphate isomerase to a weak methyl glyoxal synthase!

These results have corroborated the explanation for the function of the loop based on the x-ray structures. Removal of the loop probably allows bulk solvent to enter the active site while substrate is bound.
This has two effects: it weakens the electrostatic interactions (Glu 165, Lys 12, His 95, α-helix dipole moment) that are responsible for isomerization, and it makes less unfavorable the negative charge that must form on the bridging oxygen during the elimination reaction. Since the loop forms extensive interactions with the substrate phosphate (only one hydrogen bond, but many van der Waals contacts), a lot of binding energy is lost, so all species are bound less tightly. When substrate is bound, the phosphate group has a lot more conformational flexibility, probably allowing the bridging oxygen to access a conformation that is stereochemically more conducive to phosphate elimination.
FIGURE LEGENDS

Figure 1  The reaction catalyzed by triosephosphate isomerase. The proton that is transferred between between the two carbon atoms is shown bold typed.

Figure 2  The glycolytic pathway. TIM catalyzes the reaction at the center of this pathway and ensures that all six carbon atoms from glucose are used for energy.

Figure 3  (a) The catalytic steps showing the intermediates and transition states of the isomerization reaction. Although the intermediate is shown as two kinetically indistinguishable enediolates, it might also be an enediol. (b) Two inhibitors whose binding to TIM has been studied crystallographically. Each is shown underneath the chemical species in (a) it is thought to be an analogue of. PGA is an analogue of transition state I and PGH undergoes a partial reaction to form an analogue of the intermediate. (c) The phosphate elimination reaction proceeding from the enediolate intermediate to form methyl glyoxal and inorganic phosphate.

Figure 4  Stereoscopic view of the TIM-PGA active site showing essential catalytic residues. Glu 165 (upper left) is the catalytic base. PGA is at the center. Electrophiles His 95 and Lys 12 are on the right. The main chain atoms of helix D1, whose macrodipole moment is thought to be important for binding and catalysis are also shown.

Figure 5  The three possible roles of His 95. (a) Histidine as an imidazole stabilizing the enediolate through electrostatic
interactions. (b) The imidazolium acting as an acid to form an enediol intermediate. (c) The imidazole acting as an acid to form the enediol intermediate and an imidazolate.

Figure 6
The two base mechanism for the H95Q mutation. A substrate oxygen functions as one of the bases during the reaction (from Nickbarg and Knowles, 1988).


\[ \text{Triose phosphate isomerase} \]

\[ \text{Dihydroxyacetone phosphate} \rightarrow \text{D-glyceraldehyde 3-phosphate} \]
(a) DHAP

(b) DHAP

(c) DHAP
Parts of this Chapter were submitted to Biochemistry as part of a manuscript.

CHAPTER 2

THE STRUCTURE AND REFINEMENT OF YEAST TRIOSEPHOSPHATE ISOMERASE AT 1.9 Å RESOLUTION

Most of us don’t change opinions once we get them. Instead, we spend a lot of time looking for further proof that we’re right.

—Andy Rooney
I. INTRODUCTION

The structure of chicken muscle triosephosphate isomerase to 2.5 Å resolution was first reported by the Phillips laboratory in 1975 (Banner et al., 1975). The structure did not prove to be totally adequate for understanding the structural details of catalysis. The precise interactions between enzyme and substrate could not be determined because the enzyme structure was determined in its uncomplexed form. The crystal was eventually soaked with the substrate DHAP and data to 6.0 Å were collected. This low resolution structure revealed a conformational change in a loop composed of residues 167 to 177 but was inadequate for discerning the detailed interactions between enzyme and ligand (Phillips et al., 1977a). There was obviously a need for a structure of higher resolution with a catalytically relevant ligand in the active site. In 1981, the crystallization of yeast triosephosphate isomerase in two different crystal forms, based on the presence and absence of an inhibitor, was reported (Alber et al., 1981a). The uncomplexed crystals of the yeast enzyme were found to be of better quality than those of the chicken enzyme; these crystals diffracted to higher resolution. More importantly, high quality crystals of the enzyme complexed to the transition-state analogue PGA were finally found. This breakthrough allowed the interactions between the enzyme and the transition state to be inferred, and the structural source of catalysis to be deduced.

In this chapter, the refinement and structure of the uncomplexed enzyme to 1.9 Å resolution is described. In addition, the structure is compared to the chicken enzyme and the relationship between structure and sequence is discussed.
II. EXPERIMENTAL

A. The Initial Model.

The structure of this enzyme was solved by the isomorphous replacement method. A Cα model was built from an electron density map calculated from heavy metal derivatives to 3.0 Å resolution (Alber, 1981). This Cα structure was used to build an initial model into density calculated from data to 1.9 Å resolution. The remaining backbone and Cβ atoms were placed into density using version 6.0 of the program FRODO (Jones and Thirup, 1986; Jones, 1985) on an Evans and Sutherland PS300. This version has a function which matches the positions of a range of Cα atoms with those of highly refined structures in the Brookhaven Protein Databank. The function outputs a list of the ten structures that best match the Cα positions, and the r.m.s. deviation between the Cα’s of each of the ten structures and the segment in question. Each of these other structures can then be compared to the displayed electron density, and the coordinates for the structure whose atoms best fit the electron density can be accepted to be part of the model. It was found that it is best to examine a segment of five to ten residues before deciding which backbone structure to accept. This procedure was repeated until the backbone and Cβ positions of both subunits were determined. The side chains of each amino acid were then fitted into electron density using a variety of FRODO options.

B. Refinement.

The atomic coordinates were refined using the Restrained Parameter Least Squares method (PROLSQ) of Hendrickson and Konnert (W.A. Hendrickson, 1985). The initial model had an R-value of 43.1% for all
observed reflections between 10 to 3.3 Å. This model was subjected to crystallographic refinement in a generally repetitive manner, as follows: 1) least squares refinement proceeded until the R-value stopped decreasing; 2) at this point, the new coordinates were output and 2Fo-Fc "delete maps", in which 10-residue segments of the structure were omitted from the refinement and phase calculations, were used to manually rebuild the structure; and 3) after rebuilding, the coordinates were re-submitted for more refinement. The resolution range most influenced by bulk solvent (10 to 5 Å) was eliminated from refinement as soon as the R-factor in that shell stopped decreasing. Reflections from higher resolution shells, first to 2.5 Å and then to 1.9 Å, were eventually added to the refinement calculations. At this point the data were analyzed for systematic errors due to the data collection method. The Wyckoff scan method measures x-ray intensities at each calculated reflection without measuring the background for each reflection. Using this method, it is possible to overestimate the intensity of very weak reflections. In order to assess whether this problem existed in this data set, the R-factor was analyzed as a function of both resolution and structure factor. From the analysis, it was concluded that overestimation of weak reflections is not a problem.

Waters were added at appropriate positions by analyzing 2Fo-Fc maps. Water molecules were placed into density only if two criteria were satisfied: the contour level of the electron density must have been twice the standard deviation for the map in the asymmetric unit, and the putative water molecule must have been within hydrogen bonding distance of other polar atoms. Prior to addition of waters and refinement of individual isotropic Debye-Waller (B-) factors, the R-value had dropped
to 28.3% for all data between 5 and 1.9 Å resolution. After water molecules and individual Debye-Waller factors were included in the refinement, a weighting scheme designed to even out the R-value in each resolution range was used. A final structure using this method had an R-value of 23.1%, 27.5% if the R-factor was determined without water molecules. This structure (without the water molecules) was then used as the initial model for refinement by simulated annealing using the program XPLOR (Brunger et al., 1987). The structure that was output (again without water molecules) had an R-factor of 24.2%. The water molecules that had been removed for XPLOR refinement were appended to the XPLOR output, the structure was checked for errors, and resubmitted for additional least squares refinement, this time without the weighting scheme to even out the R-value in each resolution range. (The absence of a weighting scheme allows errors in the coordinates to be estimated using the Luzzati method (Luzzati, 1952)). In addition, the scattering contribution due to solvent was also modelled (Fraser et al., 1978) in this later version of the PROLSQ program. Eventually, the positions of the water molecules were compared with those of the complex between triosephosphate isomerase and phosphoglycolohydroxamate (Davenport et al., 1989). Of the 119 water molecules in the final model, 81 have equivalent positions in the TIM-PGH structure, which has a total of 248 water molecules. (The last model-building session in the refinement process consisted of using the 248 water molecules from the PGH/TIM structure as a guide to look for water molecules in the unbound TIM structure. Only 22 of the 119 water molecules were found by this method). The R-factor of the final model is 21.0% for 35,167 reflections from 40 to 1.9 Å resolution. Refinement statistics are given
in Table I. The progress of refinement through the various stages is summarized by the graph in Figure 1. An indication of the quality of the electron density before and after refinement is given in Figures 2 and 3.

III. RESULTS AND DISCUSSION

A. Analysis of the Structure.

The topology of TIM consists of an eight-stranded $\alpha/\beta$ barrel. This topology is one of the most common in metabolic biochemistry. To date, sixteen other functionally different enzymes with this fold have been found: pyruvate kinase (Stuart et al., 1979), enolase (Lebioda and Stec, 1988), KDGP aldolase (Mavridis et al., 1982), Taka-amylase (Matsuua et al., 1984), mandalate racemase (Neidhart, Howell, and Petsko, 1988), $\alpha$-amylase (G. Buisson et al., 1987), muscle aldolase (Sygusch et al., 1987), xylose isomerase (Farber et al., 1987; Henrick et al., 1987), glycolate oxidase (Lindqvist and Branden, 1985), RuBisCo (Schneider et al., 1986; Chapman et al., 1988), muconate lactonizing enzyme (Goldman et al., 1987), tryptophan synthase (Hyde et al., 1987), trimethylamine dehydrogenase (Lim et al., 1986), flavocytochrome $b_2$ (Xia et al., 1987), and the bifunctional enzyme N-(5'-phosphoribosyl)anthranilate isomerase-indole-3-glycerol phosphate synthase, where each domain is an $\alpha/\beta$ barrel (Priestle et al., 1987).

TIM was the first enzyme in this class and remains the prototype of what has been called the TIM barrel folding pattern. In the TIM structure, the $\beta$ barrel is composed of eight parallel $\beta$-strands involving residues 5-11, 37-41, 58-63, 89-93, 122-127, 159-165, 205-209, and 227-231.2

2 In order to conform to the numbering of triosephosphate isomerase from chicken muscle, the first residue in the yeast enzyme is designated Ala 2.
The eight $\beta$-strands have been referred to in the past by smaller case letters of the alphabet, starting with "a" for the first $\beta$-strand and ending with "h" for the last $\beta$-strand. Looking down the axis of the barrel, the barrel resembles an ellipse with the lengths of 12 Å and 18 Å for the two semiaxes. The length of the barrel itself is only about 13 Å. The barrel principally arises from the right hand twist of the $\beta$-sheet (Lasters et al., 1988). The angle made by the $\beta$-strand axes with the barrel axis is about $35^\circ$. Figure 4 contains an atomic listing of the intrachain hydrogen bonding pattern of the staggered $\beta$ strands. The barrel is situated in the protein interior; the strands are interconnected by amphipathic helices which pack against the barrel surface. As can be seen from Figure 4, the amino acids comprising the barrel are predominantly hydrophobic. Many of these hydrophobic side chains fill the interior of the barrel, and form most of the hydrophobic core of the enzyme. The only polar amino acids are found at the ends of the barrel, especially the amino terminal end, where the barrel is exposed to solvent.

As might be obvious from its name and from the Ramachandran plot in Figure 5, $\alpha/\beta$ barrels contain a fair number of $\alpha$-helices. An analysis of the structure reveals that all of the $\beta$-strands are followed by at least one helix. Using backbone hydrogen bonds as a criterion, the helical segments consist of residues 16-30, 46-54, 79-87, 95-102, 105-120, 130-137, 138-154, 177-196, 197-204, 213-221, 232-237, and 239-246.\(^3\) Since there are four more helices than there are $\beta$-strands in the overall structure, four of the $\beta$-strands are followed by a pair

---

\(^3\)The numbering of the helices begins with the residue that contributes the first carboxyl oxygen and ends with the last amide nitrogen of the helix. Using this criterion, it is inevitable that some residues participating in turns are also classified as helical.
of helices. For example, \(\beta\)-strand d is followed by helices D\(_1\) and D\(_2\), representing helices 95–102 and 105–120, respectively. All the paired helices appear very close together in sequence (95–102/105–120, 130–137/138–154, 177–196/197–204, 232–237/239–246). The first helix of these paired helices usually ends as a \(3_{10}\) helix that leads into a very tight turn to begin the next helix. Helices D\(_1\) and D\(_2\) have 3 residues between them, and for this reason a \(3_{10}\) helix at the end of helix D\(_1\) is not necessary to form a tight turn. The hydrogen bonding pattern of helices G (213–221) and H\(_1\) (232–237) allow them to be classified entirely as \(3_{10}\) helices.

Nine of the helices, A (16–30), B (46–54), C (79–87), D\(_2\) (105–120), E\(_2\) (138–154), F\(_1\) (177–196), G (213–221), H\(_1\) (232–237), and H\(_2\) (239–246), pack around the cylindrical surface of the barrel, with their main chain boundaries generally not exceeding the two ends of the barrel. Two helices, D\(_1\) (95–102) and E\(_1\) (130–137), are found at the carboxy terminal end of the barrel (where the active site is located). The amino terminal end of helix 95–102 is directed toward the active site of TIM, and it has been speculated that the helix macrodipole may provide electrostatic binding energy for the enzyme-substrate complex (Hol et al., 1978), or electrostatic stabilization of the developing enediol(ate) transition state. The presence of helices abutting the carboxy terminal end of the barrel also explains why polar residues are not found at this end of the \(\beta\)-sheet: the carboxy end of the barrel forms a cavity containing the active site, with \(\alpha\)-helix (and loop) residues actually being at the protein-water interface. The remaining helix, F\(_2\) (197–204) is present beyond the amino terminal end of the barrel. The Ca tracing of the dimer is given in Figure 6.
An analysis of the remaining amino acids reveals the various $\beta$-turns that occur in this structure. Based on the dihedral angle parameters defined previously (Chou and Fasman, 1977), ideal type 1 $\beta$-turns are found at Thr 45 and Ala 224; an ideal type 2 turn is found only at Thr 75; ideal type 3 turns are found at Ala 44, Pro 57, Val 167, Trp 168, and Pro 238. The remaining amino acids are either large loops or random coils that defy further categorization.

As one might expect, there are a number of interactions between the side chains of the barrel and the helices that surround it, and in fact, most of these interactions are hydrophobic. There are also numerous side chain interactions between the barrel and loop and turn residues. This observation should be of importance to those who have speculated that an archetypal $\alpha/\beta$ barrel protein might be used as a scaffold to design new enzymes by mutating the residues on the loops.

B. Comparison of Subunits and Estimation of Errors

TIM is a very symmetrical dimer, with non-crystallographic two-fold symmetry between the two subunits (Figure 6); the translation/rotation matrix is given in Table II. The C$\alpha$'s from the two subunits superimpose with an r.m.s. difference of 0.36 Å. The most significant difference between the two subunits when they are superimposed occurs at the amino-end of the chain. The C$\alpha$ positions for residue 2 differs by 3.2 Å in the superimposed structure. This is not very surprising since the terminal ends of proteins are usually solvent exposed and quite flexible. If residue 2 is eliminated for the purpose of the comparison, the r.m.s. difference for the two subunits

---

4 Only turns that are found to occur in both subunits, indicating a high degree of confidence in their categorization, are listed.
decreases to 0.29 Å. This number could also be considered the r.m.s. error for the whole structure. Although the two subunits have an identical sequence and therefore have a similar overall structure, part of the r.m.s. difference between the two subunits must be due to crystal field effects that each subunit experiences. Indeed, a Luzzati plot (Figure 7) indicates a coordinate error of about 0.25 Å (Luzzati, 1952). A comparison of the two active sites (all atoms in residues Asn 10, Lys 12, His 95, Ser 96, Glu 97, Cys 126, and Glu 165) gives an r.m.s. deviation of 0.30 Å. It is concluded from these comparisons that the r.m.s. error in atomic positions in this stucture is on the order of 0.3 Å.

An examination of the B-factors of the two subunits indicates a similar pattern. The differences consist of higher B-values for subunit 2 in the regions of 125-148 and 186-206 (Figure 8). High B-values have been correlated with incorrectly modeled structures (Chambers and Stroud, 1979), and these B-values might be an indication that this part of the structure is suspect. However, the quality of the electron density map argues against such an interpretation. An alternative explanation is that the differences in the B-factors are due to crystal field effects. Indeed, the amino acids at 125-148 of the first subunit make crystal contacts that are absent in the second subunit. The differences in the remaining region (186-206) are a little more difficult to assess. This region makes significant contacts with symmetry-related molecules in both subunits, and the reason for the different B-factors is not well understood.

The general features of the B-factor plot are as would be expected, with the regions of high B-factor being found on the surface of the
protein. The loop that undergoes a conformational change upon substrate binding (see following chapter), 167-176, is also a region of high B-factor but these values are, on average, not any greater than those for other loops in the protein.

C. The Active Site.

The active site of TIM, shown in Figure 9, is composed of the catalytic base, Glu 165, which has been identified by the binding of covalent active-site-directed-inhibitors (Miller and Waley, 1971; Hartman, 1971; De la Mare et al., 1972), and the electrophilic residues, Lys 12 and His 95 that are thought to polarize the carbonyl oxygens of the substrate (Alber et al., 1987a).

The carboxylic acid oxygens of Glu 165 form hydrogen bonds with the amide nitrogen and side chain hydroxyl group of Ser 96, as well as the Ne2 atom of His 95. Comparison of the conformation of Glu 165 to that of yTIM with inhibitors bound (Chapter 3 in this thesis; Davenport et al., 1989; Lolis and Petsko, 1989) reveals that the side chain moves over 2 Å upon substrate binding. The obvious implication of this observation is that serious errors could arise when model building is used to dock substrates or inhibitors in the active site of proteins. It will be difficult to predict what minor rearrangements are necessary so that the substrate is accommodated by the enzyme in its catalytically active conformation. It is becoming increasingly clear that the structures of unbound enzymes do not necessarily exist in the active conformations, and that subtle (and in some cases, major) conformational changes are required for catalysis to occur.

The charged state of the two electrophiles, Lys 12 and His 95, has been the subject of great interest. Lys 12, which has been implicated
as the amino acid responsible for the specificity of the enzyme for the natural substrates over their phosphonate analogs (Belasco et al., 1978), is involved in a hydrogen bond with the carboxylate of Glu 97. Due to the proximity of Lys 12 to Glu 97 and the usual pK$_a$ of a Lys amino group, it has generally been agreed that Lys 12 is charged. Glu 97 is also involved in a hydrogen bond with the backbone amide nitrogen of Thr 75 of the other subunit. This network of hydrogen bonds involving Lys 12 may explain why only the dimer of TIM is catalytically competent: each subunit provides a 'wall' upon which the other active site rests.

A consensus on the charged state of His 95 has been more elusive. The pK$_a$ value of a histidine imidazolium in aqueous solution is about 6.3. However, the imidazole of His 95 lies at the amino terminal end of helix $D_1$ (residues 95-102). Fersht has obtained evidence that the $\alpha$-helix dipole moment increases the pK$_a$ of an imidazole group at the carboxy terminal end of a helix (D. Sali et al.; 1988). It should logically follow that the effective positive charge at the amino terminal end due to the macrodipole moment of helix $D_1$ should decrease the pK$_a$ of His 95. Although hydrogen positions cannot be seen in electron density maps, in many cases it is rather straightforward to infer hydrogen positions based on stereochemical principles. For example, hydrogens on backbone amide nitrogens can be assigned with a high degree of confidence due to the planar requirements of a peptide bond and to knowledge of the typical nitrogen-hydrogen bond distance. We have deduced the charged state of the imidazole ring using these stereochemical considerations, as well as an analysis of the hydrogen bonding interactions of the imidazole ring. The imidazole N$\epsilon$2 of His
95 is within hydrogen bonding distance of a solvent oxygen (3.0 Å), and a carboxylate oxygen of Glu 165 (3.3 Å). Since the side chain of Glu 165 is known to be unprotonated, we conclude that the Ne2 of His 95 functions as the hydrogen bond donor and is therefore protonated. The Nδ1 atom is within hydrogen bonding distance (3.1 Å) of the backbone amide nitrogen of Glu 97. However, the requirements for a hydrogen bond between these two atoms are more stringent than other pairs of polar atoms. All the requirements are satisfied. The bond angle (C_96-N_97-Nδ1) is 128°, close to the necessary 120°. In addition, the Nδ1 atom is very close to the plane of the peptide bond formed by residues 96 and 97. In a hydrogen bonding interaction between His 95 Nδ1 and the backbone amide nitrogen of Glu 97, stereochemical considerations dictate that the amide nitrogen must be the proton donor and the Nδ1 the acceptor. The Nδ1 atom is therefore unprotonated, and the imidazole is in an uncharged state.

Unbound active sites usually refer to enzymes that are not complexed with substrate, inhibitors, or products. But even in the unbound state, solvent usually occupies the active site. In this structure there are not many crystallographically observed water molecules in the active site, therefore they must be rapidly exchanging. However, there appears to be one ordered water molecule that is present in both subunits. This water molecule (residue 602 in subunit 1 and 604 in subunit 2) is held in place by hydrogen bonds with Ne2 (3.0 Å) of His 95 and Nδ2 (3.3 Å) of Asn 10. This water molecule is displaced when substrate binds to the active site (Lolis and Petsko, 1989).

5 An interaction is defined as a hydrogen bond if two appropriate polar atoms are less than 3.3 Å apart. All distances given are the average of values in both independent subunits in the crystallographic asymmetric unit.
Much has been written in the past of the possible catalytic role of a loop near the active site of TIM (T. Alber, 1981a) that has no electron density and has thus been termed flexible. In a 6.0 Å crystal structure analysis, it has been shown that this loop (residues 167 to 176) undergoes a conformational change when DHAP is introduced into chicken TIM crystals (Phillips et al., 1977a). This movement was confirmed at 3.5 Å resolution in yeast TIM (Alber et al., 1981b). In this refined native crystal structure, the loop seems to be stabilized by two hydrogen bonds with the same loop in a symmetry related molecule. Therefore, a flexible loop in solution cannot be ruled out. Structural comparison of the bound and unbound form of the enzyme (Chapter 3) confirms that the loop moves significantly upon binding of substrate (Davenport et al., 1989; Lolis and Petsko, 1989).

D. The Subunit Interface. The major contacts between the two subunits of the TIM dimer are provided by an interdigitating loop, residues 71 to 77, that extends from one subunit into a pocket near the active site of the other subunit. Examination of the subunit interface indicates that this part of the structure is, in general, more hydrophobic than other surface areas of the monomer structure, but that it is not as hydrophobic as the globular interior of the structure (Campbell, 1988). It further seems that a large part of the specificity of one subunit for another is derived from the polar interactions at the interface. This conclusion is underscored by the fact that there are roughly 20 ordered water molecules at the subunit interface (totally shielded from bulk water), and that ten of these mediate interactions between the two subunits. It has been hypothesized that these buried
water molecules are remnants of a stage of the folding process, where the two subunits combine to make an active enzyme (Davenport, 1989).

With the exception of the interdigitating loop, there is very little sequence conservation at the interface, suggesting that monomers from different organisms would not form active heterodimers. This observation can be used to design effective and specific inhibitors of TIM. Since the active site residues of TIM are conserved among different organisms (vide infra), it is unlikely that a species-specific active-site inhibitor will be found. However, since TIM is active only as a dimer, and there is greater sequence diversity at the subunit interface among the various species, it may be possible to design a peptide or other organic molecule to bind to the subunit interface of a TIM from a pathogenic organism and prevent dimerization. The molecule would have to bind to the interface without providing the necessary "wall" to generate a proper active site.

The association constant between the two monomers has never been determined but is believed to be quite high. The crystal structure allows this number to be estimated based on buried surface area. Molecular surface calculations indicate that 575 Å²/monomer are buried on going from monomer to dimer. Campbell estimated a binding constant of $10^{14} M^{-1}$ based on the fact that the interface area is 2.5 times greater than that of the antibody Fc fragment and fragment B of protein A, whose binding constant is $10^{9} M^{-1}$ (Campbell, 1988). Using Richards's value of 20 cal/Å²/mol for the free energy of transferring an amino acid from an aqueous to less polar environment (Richards, 1977), the binding constant can be estimated to be $10^{16} M^{-1}$. If such a magnitude for the
binding constant is confirmed experimentally, it would require targeting a drug to the monomer prior to dimerization.

E. Correlation of sequence, structure and function.

It is of interest to compare this fungal TIM with the same enzyme from a higher eukaryote. We used the refined coordinates of chicken TIM furnished to us by Peter Artemiuk (personal communication). This structure has been solved to 2.5 Å resolution (Banner et al., 1975) and refined to an R-value of 17%. The enzymes from both chicken and yeast contain 247 amino acids, with a sequence identity of 53%. For the purpose of comparison, residues 4 to 56 and 57 to 248 from the chicken TIM were aligned with residues 3 to 55 and 57 to 248 from the yeast enzyme. The staggering of residues at the amino terminal end is necessary because the two sequences are out of register for the first 56 amino acids due to an insertion in the chicken sequence. An insertion at position 56 of the yeast sequence places the sequences back in register. The r.m.s. difference in coordinate position after superimposing 490 of the possible 494 Ca’s is 1.0 Å, indicating very close agreement. All atoms for the sequence positions that are strictly conserved between chicken and yeast TIM were also compared. The r.m.s. difference for the 130 residues (977 atoms) is 1.3 Å.

The sequences of triosephosphate isomerase from thirteen sources are presently known (Figure 10). Identities among these sequences range from 99% for human liver and M. mulatta to 34% for E. coli and Bacillus stearothermophilus (Figure 11). The amino acids that are strictly conserved are highlighted in the context of the complete Ca backbone (Figure 12). As is apparent from Figure 12, the conserved residues are clustered around the carboxy terminus of the barrel, where the active
site is located. Of the 44 conserved residues, 28 have atoms within 12.0 Å of WAT 602, which occupies the substrate position in the active site of subunit 1. Eleven conserved residues have atoms between 12.0 and 19.0 Å of the active site. Five of these eleven are part of the flexible loop that closes over the active site when substrate binds, and another two (Glu 129 and Thr 139) form part of a hydrogen bonding network that may be responsible for keeping the loop open when the active site is unbound. The four remaining residues that are within 12.0 and 19.0 Å of the active site form ionic interactions in hydrophobic environments of the enzyme. The side chains of Glu 104 and Lys 112 form a salt bridge that if disrupted by replacement of Glu 104 by an aspartic acid leads to a thermolabile enzyme with reduced activity (Daar et al., 1986). The side chains of Arg 189 and Glu 225 are also involved in a network that stabilizes ionic interactions in a hydrophobic environment. Of the 44 conserved residues, only five do not have any atoms within 19 Å of WAT 602. Four of these (residue 72, 73, 75, and 76) are part of the loop that interdigitates with the other subunit, and as a result are within 12.0 Å of the water molecule (WAT 604) occupying the active site in the other subunit. The remaining residue, Gly 120, is found at the surface of the protein where it terminates a helix and begins a tight turn that leads into a β-strand. Strict conservation of this residue may be important for structural reasons.

From the structural analysis of the sequence variability, the pressure to preserve the active site is clear. In order to quantitate the evolutionary pressure to maintain the active site, the r.m.s. difference between the chicken and yeast TIM atomic coordinates for
those residues that are strictly conserved among all thirteen sources was determined. The r.m.s. difference for the 44 residues (319 atoms) is 1.0 Å. If the 'flexible loop' residues (168-173) are not included in the calculation, the r.m.s. difference decreases to 0.7 Å. These numbers can be considered a composite of real differences based on residues that are not conserved but influence the conformation of conserved residues, differences in crystal contacts, and errors in the actual positions of the atoms in the structure. A better indication of how well the structure of the active site is conserved is the r.m.s. difference between what are generally acknowledged as active site residues (all atoms in Asn 10, Lys 12, His 95, Ser 96, Glu 97, Cys 126, and Glu 165). It is less than 0.4 Å for each subunit and 0.5 Å for the dimer; this number is half the overall difference of the superimposed Cα's in the chicken and yeast TIM structures.

Most of the conserved residues do not make contact with substrate when it is bound in the active site. The structure of the enzyme complexed with phosphoglycolate (Lolis and Petsko, 1989) indicates only 11 of the 44 are within van der Waals contact of the inhibitor. Most of the conserved positions form a layer around the primary active site residues (those that do make contact with substrate) and are presumably important in maintaining a precise active site structure. The analysis also indicates that there are very few positions where a specific amino acid is strictly required to maintain the structural integrity of the α/β barrel. Only five residues, Glu 104, Lys 112, Gly 120, Arg 169, and Asp 225, are implicated in maintaining the α/β barrel structure. All of this is consistent with what has been observed for other α/β barrel proteins. Although there are now structures for 17 different
proteins with this topology, very few have any sequence homology. It is clear that the analysis of strictly conserved residues is not sufficient to decipher the sequence requirements for α/β barrels.

IV. Conclusions

In this chapter, the structure of triosephosphate isomerase at 1.9 Å resolution has been reported and its active site and subunit interface described. Since triosephosphate isomerase is essential to all organisms, it is an attractive target for drug design (Wierenga et al., 1987). In light of the strong active site sequence identity among various organisms, sequence has been correlated with the structure and an alternative target (the subunit interface) for drug design proposed.

Although the structure of the uncomplexed enzyme implicates how catalysis might be achieved, it does not provide the detailed interactions between transition state and enzyme; that is the topic of the following chapter.
TABLE I

Refinement Statistics

<table>
<thead>
<tr>
<th>Restraints Applied, σ(Å)</th>
<th>Deviations Observed(^6), r.m.s Δ (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond Distance</td>
<td>0.030</td>
</tr>
<tr>
<td>Angle Distance</td>
<td>0.040</td>
</tr>
<tr>
<td>Planar 1-4 Distance</td>
<td>0.052</td>
</tr>
<tr>
<td># of bad dihedral angles(^7)</td>
<td>8</td>
</tr>
</tbody>
</table>

Resolution Breakdown

<table>
<thead>
<tr>
<th>Resolution</th>
<th># of Reflections</th>
<th>% of Data</th>
<th>Shell R-factor</th>
<th>Sphere R-factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>40.0-4.00</td>
<td>5,160</td>
<td>98.5</td>
<td>0.193</td>
<td>0.193</td>
</tr>
<tr>
<td>4.00-3.15</td>
<td>5,158</td>
<td>97.2</td>
<td>0.180</td>
<td>0.187</td>
</tr>
<tr>
<td>3.15-2.75</td>
<td>4,741</td>
<td>91.5</td>
<td>0.208</td>
<td>0.192</td>
</tr>
<tr>
<td>2.75-2.45</td>
<td>5,185</td>
<td>81.0</td>
<td>0.223</td>
<td>0.197</td>
</tr>
<tr>
<td>2.45-2.23</td>
<td>4,987</td>
<td>70.0</td>
<td>0.230</td>
<td>0.201</td>
</tr>
<tr>
<td>2.23-2.06</td>
<td>5,052</td>
<td>65.2</td>
<td>0.243</td>
<td>0.206</td>
</tr>
<tr>
<td>2.06-1.90</td>
<td>4,880</td>
<td>48.9</td>
<td>0.253</td>
<td>0.210</td>
</tr>
<tr>
<td>40.0-1.90</td>
<td>35,167</td>
<td>74.9</td>
<td></td>
<td>0.210</td>
</tr>
</tbody>
</table>

\(^6\) These Δ values are r.m.s. deviations from the corresponding values for ideal groups derived from small molecular structural studies (Sielecki et al., 1979).

\(^7\) Residues with dihedral angles significantly outside Ramachandram limits: Lys 12, Asn 28, Asn 35, His 103, Arg 303, His 403, Leu 431, Asp 522.
**TABLE II**

Orientation Matrix Between the Two Subunits in Yeast TIM

<table>
<thead>
<tr>
<th>Rotation Matrix</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-1.00000</td>
<td>-0.00168</td>
<td>-0.00198</td>
</tr>
<tr>
<td></td>
<td>-0.00251</td>
<td>0.43452</td>
<td>0.90066</td>
</tr>
<tr>
<td></td>
<td>-0.00066</td>
<td>0.90066</td>
<td>-0.43453</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Translation Vector</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>120.45</td>
<td>12.84</td>
<td>20.83</td>
</tr>
</tbody>
</table>
FIGURE LEGENDS

Figure 1. A plot of the R-factor versus refinement cycle number. The first cycle included only data from 10.0 to 3.3 Å resolution. At the eleventh cycle the data between 10.0 and 5.0 Å were deleted and data to 2.5 Å were included. At the 39th cycle data to 1.9 Å were included. Water molecules were not added and individual B-factors were not refined until the 51st cycle. The final R-factor is 0.210.

Figure 2. (a) A 2F_o-F_c delete map at 1.5σ of the initial model of the flexible loop (residues 167 to 176). More density for the atoms appears at lower contour levels. NOTE: This is not the MIR density into which the model was built. (b) A 2F_o-F_c delete map of the final model at 1.5σ.

Figure 3. (a) A 2F_o-F_c map delete map of the initial model at 1.5σ of the initial model of the active site. (b) A 2F_o-F_c delete map of the final model of the active site.

Figure 4. The hydrogen bonding pattern of the β-barrel that has been unrolled to form a single sheet. Below each β-sheet is the letter which it is represented by. The
hydrophobic nature of the barrel is evident from its sequence.

Figure 5. a. Ramachandran diagram of subunit 1. x, non-glycine residues; o, glycine residues. b. Ramachandran diagram of subunit 2.

Figure 6. A stereoscopic view of two orientations of the Ca atoms of TIM. The top view is taken down one of the barrel axes. The bottom view is shown down the symmetry axis between the two subunits.

Figure 7. Luzzati analysis of TIM. x, the R-value for different resolution ranges. The lines (from top to bottom) represents the theoretical errors of 0.15 Å, 0.20 Å, 0.25 Å, 0.30 Å, 0.35 Å, and 0.40 Å.

Figure 8. A plot of temperature factors versus residue number. The temperature factors are averaged over backbone atoms for each residue. The solid line indicates the first subunit and the dashed line indicates the second subunit. Along the bottom of the plot is the secondary structure of the sequence. α, α-helix; β, β-strand; L at 71 to 77, interdigitating loop; L at 167 to 176, flexible loop.
Figure 9. (a) Another stereoscopic view of the electron density at the active site in subunit 1. This time, the final model was subjected to seven cycles of refinement without the residues in this figure in order to eliminate any residual phase bias. The electron density is from a 2Fo-Fc omit map at 1.5 times the standard deviation. (b) The same residues as in (a) with hydrogen-bond interactions as dotted lines. The contribution of Thr 375 to the active site of subunit 1 is also shown.

Figure 10. The amino acid alignment of sequences from 13 sources: yeast (Alber and Kawasaki, 1982), Schizosaccharomyces pombe (Russell, 1985), chicken (Straus and Gilbert, 1985), rabbit muscle (Corran and Waley, 1975), human liver (Maquat et al., 1985), human placenta (Lu et al., 1984), coelacanth (Kolb et al., 1974), maize (Marchionni and Gilbert, 1986), Aspergillus nidulans (McKnight et al., 1986), Trypanosoma brucei (Swinkels et al., 1986), Escherichia coli (Pichersky et al., 1984), Bacillus stearothermophilus (Artavanis-Tsakonis and Harris, 1980), and Macaca mulatta (Old and Mohrenweiser, 1988). The numbering scheme used is that for yeast TIM, with its first residue beginning at Ala 2. The positions designated by * indicate strict conservation of amino acids.
Sequence homologies for the triosephosphate isomerases.

Stereoscopic view of two orientations of the Ca's of TIM with the highlighted side chains of strictly conserved residues. The bottom view was generated by rotating the top view 90° along the horizontal axis in the plane of the paper.
### Sequence Homologies Among the 13 Triosephosphate Isomerase Sequences

<table>
<thead>
<tr>
<th>SEQUENCE NUMBER</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>55</td>
<td>53</td>
<td>52</td>
<td>53</td>
<td>50</td>
<td>50</td>
<td>52</td>
<td>54</td>
<td>50</td>
<td>46</td>
<td>37</td>
<td>52</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>53</td>
<td>52</td>
<td>53</td>
<td>50</td>
<td>51</td>
<td>52</td>
<td>61</td>
<td>47</td>
<td>46</td>
<td>37</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>88</td>
<td>88</td>
<td>86</td>
<td>80</td>
<td>62</td>
<td>55</td>
<td>51</td>
<td>47</td>
<td>38</td>
<td>87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>98</td>
<td>93</td>
<td>83</td>
<td>63</td>
<td>54</td>
<td>52</td>
<td>47</td>
<td>38</td>
<td>98</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>92</td>
<td>83</td>
<td>62</td>
<td>54</td>
<td>53</td>
<td>47</td>
<td>36</td>
<td>99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>79</td>
<td>59</td>
<td>51</td>
<td>49</td>
<td>45</td>
<td>38</td>
<td>92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>58</td>
<td>52</td>
<td>49</td>
<td>44</td>
<td>36</td>
<td>83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>57</td>
<td>52</td>
<td>44</td>
<td>42</td>
<td>62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>48</td>
<td>46</td>
<td>40</td>
<td>54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>44</td>
<td>36</td>
<td>53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>-</td>
<td>34</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>-</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Sequence Number Code

<table>
<thead>
<tr>
<th>SEQUENCE NUMBER CODE</th>
<th>ORGANISM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>YEAST</td>
</tr>
<tr>
<td>2</td>
<td>S. POMBE</td>
</tr>
<tr>
<td>3</td>
<td>CHICKEN</td>
</tr>
<tr>
<td>4</td>
<td>RABBIT</td>
</tr>
<tr>
<td>5</td>
<td>HUMAN LIVER</td>
</tr>
<tr>
<td>6</td>
<td>HUMAN PLACENTA</td>
</tr>
<tr>
<td>7</td>
<td>COELEOCANTH</td>
</tr>
<tr>
<td>8</td>
<td>MAIZE</td>
</tr>
<tr>
<td>9</td>
<td>A. NIDULANS</td>
</tr>
<tr>
<td>10</td>
<td>T. BRUCEI</td>
</tr>
<tr>
<td>11</td>
<td>E. COLI</td>
</tr>
<tr>
<td>12</td>
<td>B. STEAROTHERMOPHILUS</td>
</tr>
<tr>
<td>13</td>
<td>M. HULATTA</td>
</tr>
</tbody>
</table>
CHAPTER 3

CRYSTALLOGRAPHIC ANALYSIS OF THE COMPLEX BETWEEN
TRIOSEPHOSPHATE ISOMERASE AND 2-PHOSPHOGLYCOLATE AT
2.5 Å RESOLUTION: IMPLICATIONS FOR CATALYSIS
I. INTRODUCTION

If the structural basis of catalysis for any enzyme is to be understood, the structures of the uncomplexed enzyme (previous chapter), and the enzyme complexed to catalytically relevant ligands (such as substrates, intermediates, products, or analogues of any chemical species on the reaction coordinate) need to be investigated. Perhaps the most important structure for understanding catalysis is that of the enzyme complexed to a transition-state analogue. In this chapter, the structure of the complex between triosephosphate isomerase and the transition-state analogue PGA is described and compared to the uncomplexed enzyme. The crystallographic results are also used to interpret various N.M.R., u.v. and pH dependence experiments in a structural context (Campbell et al., 1978; Campbell et al., 1979; Jones and Waley, 1979).

2-Phosphoglycolate has been proposed to be a transition-state analogue of TIM even though it is one methylene group shorter than any chemical species found on the reaction coordinate (Figure 1). The absence of a number of atoms relative to the true transition state does not prevent PGA from binding to the enzyme over 100 times more tightly than either DHAP or GAP, which have $K_m$s on the order of 1 mM. Although tight binding of an inhibitor provides powerful support for the designation of transition-state analogue, it should not be the only evidence. At least one tight binding inhibitor (methatrexate) that was once thought to be a transition-state analogue of an enzyme (dihydrofolate reductase) has subsequently been found to be opportunistic, taking advantage of binding interactions that have little relevance to catalysis (Mathews et al., 1978; Williams et al., 1979).
Although PGA replaces a methanol group with a much smaller oxygen atom and is therefore not a conventional transition-state analogue where only one atom might be changed, it is also not believed to be opportunistic. For these reasons, the interactions between PGA and TIM have been studied extensively using pH experiments (Hartman et al., 1975), u.v. spectrophotometry (Jones and Waley, 1979), and N.M.R. (Campbell et al., 1978; Campbell et al., 1979). Phosphoglycolohydraxamate (PGH), an inhibitor that emulates the enediolate intermediate has been synthesized and studied (Collins, 1974). Its structure complexed to TIM has recently been investigated at 1.9 Å resolution (Davenport et al., 1989).

II. Materials and Methods

A. Crystallization, Data Collection and Data Reduction.

Yeast triose phosphate isomerase (Sigma Chemical Co., Type I) was exhaustively dialyzed against a solution of 0.2 M Tris (pH 6.8) containing 1mM EDTA and 1 mM mercaptoethanol. Batch methods were used to crystallize the enzyme. Polyethylene glycol 4000 (J. T. Baker Chemical Co.) was added to a final concentration of 15-20% to a solution containing 20 mg/mL yTIM and 1.5 mM phosphoglycolic acid (Sigma). At this point the solution appeared slightly cloudy; either a small amount (5-10 µL) of water or buffer was added to redissolve the enzyme. Usable crystals (0.6 mm X 0.5 mm X 0.4 mm) were obtained in about a month at room temperature. The crystals have the symmetry of the space group P2₁ with one dimeric molecule in the asymmetric unit. The unit cell of the crystal has dimensions a=74.35 Å, b=83.97 Å, c=38.67 Å, β=99.70°. It is isomorphous with the monoclinic crystal form obtained on co-crystallization of yeast TIM with PGH (Davenport et al., 1989).
The crystals were mounted at room temperature in a quartz capillary tube with a column of mother liquor on each side to keep the crystal from drying out. The tube was also sealed with mineral oil and wax. Data were collected on two crystals on a Nicolet P3 diffractometer at room temperature, with nickel-filtered copper Kα radiation. About 10,000 reflections were collected from 44 to 3.0 Å resolution on one crystal, and 8,300 reflections from 3.15 to 2.5 Å resolution on the other crystal. The data were collected by the full integration method, scanning on ω for a total of 1°, counting for 30 seconds on the peak and 15 seconds on the background. The background was collected 1° away from the calculated peak position. The data were processed using the PROTSYS package of programs (Petsko, unpublished results) on a VAX 11/750. Radiation damage was corrected by an approximation of linear decay of intensity with exposure time. Five reflections were monitored during data collection, but only four had similar rates of radiation damage. These four were averaged, and applied to all other reflections. An empirical method was used to correct for absorption (North et al., 1968). The reduced data from the two crystals were scaled and merged to a final $R_{merge} = \frac{\sum_{hkl} |I| - \bar{I}|}{\sum_{hkl} I} \frac{I}{\sum_{hkl} I}$ of 8.4% for about 800 overlapping reflections. Sixty-four percent of the theoretical number of reflections had intensities twice the standard deviation and were kept for further work.

B. The Initial Model.

The initial model was based on the protein atoms of the incompletely refined structure of TIM-PGH, which is isomorphous to the TIM-PGA structure. The TIM-PGH model was phased by a combination of isomorphous and molecular replacement (Davenport, 1985). Ethylene
mercury phosphate (EMP), which is known to bind to Cys 126 of both monomers in the asymmetric unit of native TIM (Alber, 1981), was used to make a heavy metal derivative of the TIM-PGH crystal form after attempting to soak out the PGH. Difference Patterson maps defined the position of the heavy metal atoms in the asymmetric unit, but the derivative could not be used for isomorphous replacement phasing due to its poor quality. However, with the mercury (and presumably Cys 126) positions defined, the unrefined structure of native TIM was placed into the unit cell of TIM-PGH such that the sulfhydryls of Cys 126 were adjacent to the mercury positions. Rotation of the dimer about the axis defined by the vector between the two mercury atoms, combined with an R-factor analysis, revealed the correct orientation of the TIM dimer in this unit cell.

C. Refinement.

A 2F_o-F_{c,omit} PGH Fourier map was calculated by deleting the bound PGH of the yeast TIM-PGH structure (Davenport et al., 1989) for the structure factor and phase (e^{2\pi i c_{omit}^{\text{PGH}}}) calculation, and using the TIM-PGA observed structure factors. After minor adjustments, the structure without the coordinates for PGA was refined with PROLSQ (Hendrickson, 1985). The initial cycle of refinement included all data; the R factor was 33.1%. Six subsequent cycles in the first round of refinement with data from 10.0 to 2.8 Å resolution yielded an R factor of 24.5%. After a number of sessions involving rebuilding and refinement, the R-factor dropped to 18.5% for reflections from 5.0 to 2.5 Å. Addition of coordinates for PGA and 30 waters (which were placed in difference density of over 3 standard deviations) and additional refinement yielded a final structure whose R-factor was 16.7%. At this stage, the
reflections from 10.0 to 5.0 Å resolution were added in the refinement. The scattering contributions from the bulk solvent were modelled by a method relying on the Babinet principle (Fraser et al., 1978). The final R-factor for 10,355 reflections between 10.0 and 2.5 Å resolution is 18%. Refinement statistics are given in Table I.

III. RESULTS

It had been previously observed that the native unbound crystals of triosephosphate isomerase soaked in solutions containing either DHAP or PGA showed substantial changes in the diffraction pattern. In most cases, the crystals cracked, resulting in both weaker diffraction and unit cell changes (Johnson and Wolfenden, 1970; Alber, 1981). Conditions were found that allowed the diffusion of phosphate or DHAP into chicken TIM crystals to be studied, but only to 6 Å resolution (Rivers, 1977; Phillips et al., 1977a). In order to overcome these problems and observe the complexed enzyme at higher resolution, TIM was co-crystallized with PGA. These crystals were of a different crystal form from those of the uncomplexed enzyme but were isomorphous with co-crystals of yeast TIM and PGH. The full details describing the solution of the yeast TIM-PGH crystal form are given elsewhere (Davenport et al., 1989).

A. Comparison of Subunits and Estimation of Error.

The accuracy of the atomic coordinates in this structure can be assessed by comparison of the two subunits that comprise the asymmetric unit of the crystal. A least squares superposition of the Ca's of the two subunits leads to an r.m.s. difference of 1.0 Å. This is a much higher number than would normally be expected for two identical subunits. As a point of reference, a least squares comparison of 490...
Ca's between chicken and yeast triosephosphate isomerase, which have 53% sequence identity, also yields an r.m.s. difference of 1.0 Å. Close inspection of the superimposed subunits reveals major differences at residues 2, 55, and 56. As with most protein structures, the terminal ends are highly flexible and hard to place in the electron density map, so it is not unexpected that there is a large difference at Ala 2. Residues 55 and 56 (and the corresponding residues in the other subunit) are in a loop at the surface of the protein. These two loops on the superimposed subunits are over 6 Å apart. Such a large difference might indicate that one (or both) of the loops is misplaced. However, Fo-Fc delete maps indicate there is density at the positions of the loop in each subunit. It should be noted, however, that a Ramachandran plot (Figure 2) indicates that the conformation of the loop in the second subunit places many of the dihedral angles in those residues outside the acceptable limits. If one of the loops is incorrectly placed, then it would probably be the one of the second subunit. The only difference that might account for the two distinct conformations is a 4 Å electrostatic interaction in the second subunit between the side chain of Lys 56 and a carboxylate oxygen of a symmetry related molecule; this interaction is absent in the first subunit. Although this is a large distance for an electrostatic interaction, it can only be concluded that the different conformation of this loop in the two subunits is due to crystal field effects.

In order to eliminate the effects due to the crystal contact or to the flexibility of the amino terminal end of the protein, residues 2, 55, and 56 from both subunits were left out of the least squares superposition calculation. The r.m.s. difference of the Ca's between
the two subunits dropped to 0.6 Å. The r.m.s. difference between all atoms in the two active sites (including residues 10, 12, 95 to 97, 126, 165, 167 to 176, and PGA) is also 0.6 Å. Although these comparisons must certainly include other residues that are flexible, they cannot be identified by the experimental data for removal. Individual B-factors, which have been correlated with flexibility (Petkso and Ringe, 1984), were not refined due to the low overdeterminacy of the data (10,000 reflections for almost 4,000 atoms). This number (0.6 Å) can therefore be considered a conservative estimate for the average error in the atomic coordinates for this structure.

B. The PGA-Bound Active Site.

The enzyme makes a substantial number of hydrogen bonds to PGA in the active site. The O₁ carboxylate oxygen of PGA forms hydrogen bonds to the NZ (3.0 Å)⁸ atom of Lys 12 and the Ne2 (2.7 Å) of His 95. The O₂ carboxylate oxygen forms a hydrogen bond to the protonated (see below) carboxylate oxygens of Glu 165. The terminal phosphate oxygens are hydrogen bonded to the backbone amide nitrogens of Gly 171 (2.6 Å), Ser 211 (2.7 Å), and to a water molecule. A complete list of active-site hydrogen bonds is given in Table II.

The bridging phosphate oxygen does not make any hydrogen bonds at all. This is particularly surprising in view of the fact that the phosphonate analogs of DHAP and GAP, where the bridging oxygen is replaced by a methylene group, are poor substrates for the enzyme. Whereas the phosphonate analog of D-GAP, D-2-hydroxy-4-phosphonobutyraldehyde (CHO-CHOH-CH₂-PO₃H₂), has an enzyme-catalyzed

---

⁸ An interaction is classified as a hydrogen bond if the distance between the two polar atoms is less than 3.3 Å. The number that appears in parentheses next to the enzyme atom is the average distance to that atom in both subunits. Only the hydrogen bonds that are common in both subunits are noted here. For a complete list of all hydrogen bonds in both subunits, Table II should be consulted.
rate constant that is 800-fold lower than natural substrate (Belasco et al., 1978), the DHAP-analog, 4-hydroxy-3-oxobutylphosphonic acid (CH$_2$OH-CO-CH$_2$-PO$_3$H$_2$), hardly binds to the enzyme at a concentration as high as 10 mM (Dixon and Sparkes, 1974). Similarly, the phosphonate analog of PGA, 3-phosphonopropionate (CO$_2$H-CH$_2$-CH$_2$-PO$_3$H$_2$), has an affinity about 10$^3$-fold lower for TIM than PGA (Wolfenden, 1970; Waley, 1973). All these data led to speculation that interactions with the bridging oxygen, presumably by Lys 12, accounted for the specificity of this enzyme. The ε nitrogen of Lys 12, however, is involved in hydrogen bonds with the carboxylate oxygens of PGA and Glu 97 more than 4.0 Å away from the bridging oxygen. (It should be noted in the TIM-PGH complex, the distance between O$_3$ of PGH and NZ of Lys 12 is 3.5 Å. Although this is closer, it is not clear whether it is sufficiently close to explain the differences in binding.) What determines the specificity for the natural substrates over their phosphonate analog may have little to do with hydrogen bonding and more to do with the preference of methylene groups for the staggered conformation (Lehn, 1971; Dixon and Sparkes, 1974). As can be seen from Figure 3, the bridging oxygen adopts a conformation that is planar to the PGA carboxylate group. The energetic cost of moving a methylene group from a staggered to an eclipsed conformation may be too high for binding and catalysis to occur.

Since it is known from N.M.R. studies that PGA binds to TIM as a tri-anion (Campbell et al., 1978), it is appropriate to discuss the ionization state of the active site and the potential electrostatic interactions between enzyme and inhibitor. Based on pH titration experiments, it has been found that a proton is taken up by the enzyme
upon PGA binding (Campbell et al., 1979); this phenomenon is not observed with the binding of D-GAP, DHAP, or any other inhibitor. The presence of the negatively charged carboxylate of PGA in the active site is probably the cause of the proton uptake by the enzyme. The structure of the active site has been analyzed to try to deduce the atom that has been protonated. The only residues that interact with the carboxylate group of PGA are Lys 12, His 95, and Glu 165. Lys 12, in the unbound enzyme, is believed to be positively charged as it interacts with the carboxylate of Glu 97. Therefore, Lys 12 is incapable of accepting a proton. It is also unlikely that the ionization state of the imidazole of His 95 changes upon binding of PGA. In the unbound enzyme, the imidazole Nε1 nitrogen acts as a hydrogen-bond acceptor from the backbone amide nitrogen of Glu 97, while the Ne2 nitrogen is hydrogen bonded to a water molecule in the active site (Chapter 2). It is most likely in this arrangement that only the Ne2 nitrogen of the imidazole is protonated. In the bound form, the PGA displaces the water molecule and is now the hydrogen-bond acceptor to the imidazole Ne2 nitrogen. The Nε1 nitrogen continues to be hydrogen bonded to the backbone amide nitrogen. Glu 165, on the other hand, changes from being hydrogen bonded to the hydroxyl oxygen of Ser 96 to being hydrogen bonded to the carboxylate oxygen of PGA. Various experiments (Waley, 1972; Hartman et al., 1975; Belasco et al., 1978) have implicated a pKₐ value of below 5 for the Glu 165 carboxylate. In view of these data, it seems reasonable that Glu 165 becomes protonated to replace a potentially destabilizing electrostatic repulsion between its side chain carboxylate and the anionic carboxylate of PGA with what is thought to be a very strong bond between two carboxylates that interact through a proton (Jeffrey and
Maluszynska, 1982). The replacement of the C₁ carbon atom at GAP by the O₁ oxygen of PGA presumably accounts for some of the additional binding energy exhibited by PGA relative to the substrates. This replacement changes the (Glu 165)-C-O⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻реш  

C. Conformational Change Upon Binding of PGA.

It has been known for twenty years that TIM undergoes a conformational change upon substrate or PGA binding (Johnson and Wolfenden, 1970). A 6 Å difference electron density map of chicken TIM crystals soaked with substrate first indicated the presence of large scale motions around the active site (Rivers, 1977; Phillips et al., 1977a). Subsequently, changes in the u.v. absorbance, which were attributed to tyrosine and tryptophan residues, were observed upon binding of ligands (Jones and Waley, 1979). Besides noting the specific interactions between TIM and PGA, we are now able to describe the nature of the conformational change in greater detail.

A least squares superposition using only the Cα's of the unbound and bound form of yeast TIM reveals one significant change. A ten amino acid loop, comprised of residues 167 to 176, has moved from an open conformation in the unbound form to a closed one with PGA in the active site. Some atoms in this loop move over 8 Å to form a single hydrogen bond and numerous van der Waals contacts with the inhibitor.

The only polar interaction between the loop and PGA involves a hydrogen bond between the backbone amide nitrogen of Gly 171 and a phosphate oxygen. A superposition of only the atoms in the loop reveals that the side chains do not undergo any conformational changes during the motion from the open to the closed form. An analysis of the
conformation of the loop in the open and closed forms reveals a number
of interesting features. Residues 167 and 168 can be classified as type
3 turns, and residue 170 can be classified as a type 2 turn. Hydrogen
bonds involving the carbonyl oxygen of Pro 166 and the amide nitrogen of
Ala 169, the carbonyl oxygen of Trp 168 and the hydroxyl of Thr 172, and
the amide nitrogen of Leu 174 and the hydroxyl of Thr 172 seem to be
responsible for keeping the loop in a single conformation in both the
open and closed forms. The interplay between these three intraloop
hydrogen bonds and the available dihedral angles of L-amino acids
restricts the conformational space of the loop. Although the loop has
the freedom to move from the open to the closed form, all evidence
suggests it does so essentially as a rigid body.

An analysis of the relative position of aromatic residues in the
bound and unbound enzyme indicates that the loop movement affects
interactions among Tyr 164, Trp 168, and Tyr 208. These residues are
probably the cause of the changes in the u.v. spectrum (Jones and Waley,
1979).

To look for significant, yet subtle, conformational changes upon
substrate binding, the atoms of active site residues Asn 10, Lys 13, His
95, Ser 96, Glu 97, and Glu 165 from the PGA-complexed and uncomplexed
structure were superimposed. The r.m.s. difference for these residues
is 0.5 Å. Close examination of this least squares superposition reveals
a conformational change involving the catalytic base, Glu 165. In the
unbound structure, the carboxylate oxygens of Glu 165 are hydrogen
bonded to the backbone amide nitrogen and side chain hydroxyl group of
Ser 96, as well as to the Nε2 of His 95. With PGA bound in the active
site these side chain interactions are broken, and the carboxylate group moves over 2 Å to interact with the inhibitor.

IV. DISCUSSION

The transition-state theory of catalysis states that active sites accelerate reactions by being structurally and chemically complementary to the transition state of the reaction. Transition-state analogues are useful in enzymology because they provide clues about the factors that are responsible for catalysis by a particular enzyme. The design of these analogues is based on a mechanistic model for the reaction (Wolfenden, 1969). Structural studies of an analogue complexed to an enzyme can confirm its relevance to the presumed mechanism and reveal interactions that are important for catalysis.

A. The Origin of Tight Binding of PGA to TIM.

PGA binds to TIM at least 100 times more tightly than substrate even though some of the interactions that occur with the true transition state—most notably with the O₁ oxygen—are absent. The crystallographic results indicate that the origin of this tight binding is most certainly due to directed electrostatic interactions between inhibitor and enzyme residues, particularly His 95 and and the protonated Glu 165. It is particularly revealing that the $K_i$ is strongly dependent on ionic strength. Binding is strongest at low ionic strength, indicating the importance of electrostatic effects in this enzyme. It is also interesting to note that binding of PGA to TIM is stronger at pH 5 than at pH 7 (Hartman et al., 1975). This tighter binding at the non-physiological pH can be attributed to a higher concentration of enzyme protonated at Glu 165.
The importance of electrostatic interactions during the course of catalysis can be traced by analyzing the free energies of binding of various inhibitors. These numbers are only meant to give a qualitative description of the relative energies between various moieties in the reaction coordinate and do not reflect the actual energies during the catalytic cycle. From inhibition studies with inorganic phosphate ($K_i=6$ mM for the chicken enzyme) (Burton and Waley, 1968), it can be estimated that the phosphate binding site is responsible for about 3 kcal of the binding energy. Using an inhibition constant of 1.4 mM for the substrate analogue, glycerol-3-phosphate, (Nickbarg and Knowles, 1988) as an equilibrium constant between substrate and enzyme, it is estimated that the interactions between the enzyme and the non-phosphate portion of the substrate are worth about another 1.0 kcal of energy. The binding constant of PGA ($K_i=15$ $\mu$M) indicates that the charge redistribution that occurs during the course of catalysis (assuming that PGA provides an accurate simulation of the charge configuration at the transition state) is worth at least another 2.5 kcal in energy.

The above analysis is based on the assumption that enzyme active sites are complementary to the transition state of the reactions they catalyze. By definition, transition-state analogue inhibitors do not perfectly mimic the transition state. However, these analogues can be used to qualitatively decipher the interactions that are the source of catalysis.

B. Function of the Conformational Change.

A comparison of the PGA-complexed enzyme with the unbound enzyme has revealed essential structural features of catalysis. Figure 4 displays the changes in the active site geometry as TIM goes from the
unbound to PGA-bound state. The presence of PGA in the active site induces the catalytic base, Glu 165, to move over 2 Å, become protonated (not necessarily in this order), and form a hydrogen bond with the carboxylate oxygen of PGA. Although PGA does not exactly mimic the transition state, a similar conformational change must occur when natural substrate is bound. A more dramatic conformational change is the movement of a ten residue loop that closes over the active site. This movement is known to occur when substrate also binds (Alber et al., 1981b; 1987a).

Functional roles for conformational changes in enzymes had been postulated even before any had been observed crystallographically. In the induced-fit mechanism of catalysis, substrate specificity could be enhanced by ligand-induced conformational changes that properly align enzyme groups for catalysis (Koshland, 1958). It has since been argued that, although the induced-fit theory elegantly describes phenomena associated with allosteric enzymes, it cannot be evoked to explain substrate specificity: ligand-induced conformational changes for all substrates lower \( k_{\text{cat}}/K_{\text{m}} \) by the energy required to distort the enzyme (Fersht, 1985). More recently, it has been proposed that under certain circumstances, such as when the substrate is totally surrounded by protein in the active site, conformational changes can increase specificity and catalysis (Herschlag, 1988).

How are the active site geometry and conformational changes that occur in TIM related to catalysis? Clearly, the enzyme has rigidly fixed the factors that are responsible for polarizing the substrate carbonyl oxygens (Belasco and Knowles, 1980). His 95, Lys 12, and helix D1, whose macrodipole moment (Hol et al., 1978) induces a positive
electric field at its amino terminal end, hardly move. In polarizing
the substrate carbonyl oxygens, TIM weakens the substrate C-H bonds to
the carbonyls. The carboxylate of Glu 165 then simply swings to
position to abstract and transfer a proton from one carbon to another.
The functional role of the conformational change at position 165 is
still an open question. It is possible that the catalytic conformation
in the uncomplexed enzyme might sterically prevent productive substrate
binding. In addition, the enzyme probably has an electropositive
environment around the uncomplexed active site in order to attract the
negatively charged phosphate. Ser 96 might be used to sequester the
negatively charged glutamate away from the entrance of the active site
in order to maintain positive electropositive potential.

The role of the loop movement is a little more difficult to
decipher. It has previously been proposed that the role of this loop
might be to shield the active site from bulk water (Alber et al.,
1987a). Indeed, a Connolly surface calculation (Connolly, 1983) using a
1.4 Å probe (which approximates the size of a water molecule) indicates
that most of the substrate is shielded from bulk water. As discussed
above, electrostatic interactions are thought to be very important
during catalysis of this enzyme. Any shielding of bulk water by the
loop could effectively make the active site environment more
hydrophobic, thereby increasing the magnitude of electrostatic
interactions that occur within it. This interpretation might, at first
glance, appear to be contradicted by the ionic strength dependence of
PGA binding (Hartman, et al., 1975). However, it is not clear that the
ionic strength affects the magnitude of the electrostatic interactions
at the active site, or whether it screens the enzyme's electrostatic
gradient, which might enhance the association rate of ligands to the active site. It has already been shown that another enzyme operating at the diffusion-controlled limit, superoxide dismutase, enhances the diffusion of substrate to the active site with its electrostatic potential (Sharp et al., 1987). The ionic strength dependence of the superoxide dismutase reaction manifests itself on the rate of association between superoxide anion and the enzyme's active site (Cudd and Fridovich, 1982; Koppenol, 1981). It is not unreasonable to expect that the same might be true of TIM, although experimental evidence for this is unavailable at this time.

The closure of the loop may also play a role in preventing the phosphate elimination reaction that occurs so readily with more simple organic bases (Richard, 1984) (Figure 1c). For phosphate elimination to occur the bridging $O_3$ oxygen must either develop a negative charge or be protonated. Although there are a number of polar interactions between the enzyme and the phosphate group, it is interesting to note the absence of any positively charged groups, which could either function to protonate the bridging oxygen or to stabilize the negative charge that develops as phosphate is eliminated. By removing the electrostatic shielding powers of bulk solvent in the active site, the introduction of an uncompensated negative charge in the hydrophobic environment of the active site might be discriminated against.

Prevention of phosphate elimination might also be influenced by stereoelectronic factors (Deslongchamps et al. 1972; Kirby, 1983). In this regard, the orientation of the bridging phosphate oxygen of PGA with respect to the plane of the carboxylate has important mechanistic consequences. By analogy, it can be assumed that the bridging oxygen of
the enediol intermediate is also planar to the enediol(ate) moiety. Stereoelectronic theory predicts that the in-plane conformation observed in this complex disfavors the elimination reaction. The out-of-plane conformation that would lead to a dihedral angle of 90° for the atoms O₁-C₁-C₂-O₃ places the orbitals in the most favored positions for phosphate elimination (Figure 5). Once again, the loop’s presence is felt. In its closed position, the loop forms extensive interactions with the terminal oxygens on the phosphate. The interactions with the terminal oxygens directly influence the conformation of the bridging O₃ oxygen. The loop may therefore serve to hold the O₃ oxygen in the stereoelectronically important in-plane conformation. These observations have led to the prediction (Alber et al., 1987a) that tinkering with the loop would increase the rate of the elimination reaction dramatically, which has recently been verified by genetic deletion of four residues from the loop of the chicken enzyme (Pompliano et al., 1989).

C. Dynamics and Mechanism of the Conformational Change.

Although there now exist structures of the loop in the open and closed conformations, very little is known about the mechanism of the movement. A question that remains unanswered involves the atomic details and forces that trigger the loop to close when substrate is bound in the active site. Inspection of the superimposed bound and unbound active sites reveals a subtle difference in the backbone atoms of Glu 165. Although this difference might be within the error limits of the coordinates, if real, it may provide a clue to the mechanism of loop closure. If a shift in the backbone atoms of Glu 165 is also necessary to correctly orient the side chain carboxylate for catalysis,
this shift may be relayed to the nearby amino acids that compose the loop, triggering a further conformational change. This mechanism is now in the process of being tested by computer simulation (Joseph, 1989).

The dynamics of loop movement have been studied by computer simulation using chicken TIM coordinates and modeled DHAP (Brown and Kollman, 1987), but the mechanism or trajectory of loop movement was not described. Their study produced a TIM-DHAP structure that is qualitatively similar to some of the details in the structure observed in this study. The most important similarity involves the hydrogen bond made between the loop Gly 171 amide nitrogen and substrate phosphate oxygen. The structures of the unbound (previous chapter) and PGA-complexed TIM may improve the quality of computer simulations by providing good initial and final states of loop movement. These structures are the beginning of a series that might ultimately lead to a detailed understanding of catalysis by this enzyme.
<table>
<thead>
<tr>
<th>TABLE I</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Refinement Statistics</strong></td>
</tr>
<tr>
<td>Number of proteins atoms</td>
</tr>
<tr>
<td>of inhibitor (PGA) atoms</td>
</tr>
<tr>
<td>of water molecules</td>
</tr>
<tr>
<td><strong>Restraints</strong></td>
</tr>
<tr>
<td>Applied, $\sigma$ (Å)</td>
</tr>
<tr>
<td>Bond Distance</td>
</tr>
<tr>
<td>Angle Distance</td>
</tr>
<tr>
<td>Planar 1-4 Distance</td>
</tr>
</tbody>
</table>

Number of Bad Dihedral Angles$^9$ : 20

<table>
<thead>
<tr>
<th>Resolution Breakdown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution</td>
</tr>
<tr>
<td>10.0-4.80</td>
</tr>
<tr>
<td>4.80-3.90</td>
</tr>
<tr>
<td>3.90-3.40</td>
</tr>
<tr>
<td>3.40-3.10</td>
</tr>
<tr>
<td>3.10-2.80</td>
</tr>
<tr>
<td>2.80-2.50</td>
</tr>
<tr>
<td>10.0-2.50</td>
</tr>
</tbody>
</table>

TABLE II
Hydrogen Bond Distances Between Non-Hydrogen Atoms of PGA and TIM

<table>
<thead>
<tr>
<th>Hydrogen Bond</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O_1$ 249--N$\delta$2 Asn 10$^1$</td>
<td>3.2</td>
</tr>
<tr>
<td>$O_1$ 249--NZ Lys 12</td>
<td>3.0</td>
</tr>
<tr>
<td>$O_1$ 249--N$\epsilon$2 His 95</td>
<td>2.7</td>
</tr>
<tr>
<td>$O_2$ 249--N$\epsilon$2 His 95$^1$</td>
<td>3.1</td>
</tr>
<tr>
<td>$O_2$ 249--O$\epsilon$1 Glu 165</td>
<td>3.2</td>
</tr>
<tr>
<td>$O_2$ 249--O$\epsilon$2 Glu 165</td>
<td>2.7</td>
</tr>
<tr>
<td>$O_4$ 249--N Gly 171</td>
<td>2.6</td>
</tr>
<tr>
<td>$O_4$ 249--N Ser 211</td>
<td>2.7</td>
</tr>
<tr>
<td>$O_4$ 249--O$\gamma$ Ser 211$^2$</td>
<td>3.3</td>
</tr>
<tr>
<td>$O_5$ 249--N Gly 171$^2$</td>
<td>3.2</td>
</tr>
<tr>
<td>$O_5$ 249--N Gly 233$^1$</td>
<td>3.3</td>
</tr>
<tr>
<td>$O_6$ 249--N Ser 211$^1$</td>
<td>3.1</td>
</tr>
<tr>
<td>$O_6$ 249--N Gly 232$^2$</td>
<td>3.1</td>
</tr>
<tr>
<td>$O_6$ 249--O Wat 640$^1$</td>
<td>2.8</td>
</tr>
<tr>
<td>$O_6$ 249--O Wat 641</td>
<td>2.4</td>
</tr>
<tr>
<td>$O_6$ 249--O Wat 643</td>
<td>2.5</td>
</tr>
</tbody>
</table>

* The distance is the average of both subunits unless otherwise noted.

1 This hydrogen bond is found only in subunit 1.

2 This hydrogen bond is found only in subunit 2.
FIGURE LEGENDS

Figure 1  The reaction catalyzed by TIM, including the transition states for proton transfer. The subscripts on the carbon atoms refer to the numbering scheme for the molecule. For substrates, transition states, and intermediates, the numbering of the oxygens corresponds to the numbering of the carbon atoms to which they are attached. The numbering of the PGA carboxylate oxygens is also shown. "B:" refers to the catalytic base. Although the intermediate is depicted as a kinetically indistinguishable enediolate, it is possible it is an enediol. The phosphate elimination reaction is shown proceeding from one of the enediolates.

Figure 2  a. Ramachandran diagram of subunit 1. b. Ramachandran diagram of subunit 2.

Figure 3  A stereoscopic view of PGA in the active site of subunit 1 surrounded by residues that are within 3.5 Å. All hydrogen bonds are shown as dotted lines.

Figure 4  The conformational changes in the TIM active site upon binding of PGA. This view was generated from the orientation matrix between Cα’s (not including flexible residues – 3, 167 to 176, and 247) of the uncomplexed and PGA-bound enzyme. The structure denoted by the darker
lines is that of the PGA-bound form. The arrows indicate the conformational changes that occur upon binding of PGA, which is also shown.

**Figure 5**

Stereoelectronic effects in the TIM catalyzed reaction. The conformation that would lead to the phosphate elimination reaction is on the top. In this conformation, the bridging oxygen sits above the plane of the enediol intermediate. The sp³-hybridized orbital of atom C3 is placed at an orientation that best overlaps with the π-orbital system of the enediol, a necessary condition for phosphate elimination. On the bottom the bridging oxygen is planar to the enediol moiety. This conformation minimizes the overlap between the sp³ orbital of C3 and the π orbitals system. Hence, phosphate elimination is disfavored.
CHAPTER 4

CRYSTALLOGRAPHY OF MUTANTS OF CHICKEN TIM
I. Introduction

While crystallographic studies on the yeast enzyme were being pursued, the gene for the chicken muscle enzyme was cloned, expressed, and mutagenized at Harvard. In an effort to correlate enzymatic activity with the distance of the catalytic base from substrate, Glu 165 was substituted with an aspartic acid (Straus et al., 1985). Determination of the free energy profile of this mutant revealed that the rate of the enzymatic reaction is no longer diffusion-controlled; proton abstraction is now the rate-limiting step. This mutant enzyme has a specific activity 500-fold less than wild-type. In the thermodynamically favored direction (GAP to DHAP), $k_{\text{cat}}$ is $2.8 \text{ s}^{-1}$, down a factor of 1,500 relative to wild-type. In the reverse direction, $k_{\text{cat}}$ is down a factor of 240 to $1.8 \text{ s}^{-1}$ (Raines et al., 1986).

This mutation raises two interesting questions about catalysis. For one, what is the new conformation of the side chain at position 165? In the wild-type enzyme Glu 165 forms a hydrogen bond with the hydroxyl of Ser 96 that is 2.8 Å away. An increase of about 1 Å (the distance expected from removing a C-C bond) in the distance between the interacting atoms would effectively make the strength of such a hydrogen bond negligible. Without the hydrogen bond to stabilize its conformation, Glu 165 might be free to adopt some other conformation or perhaps remain flexible. Presumably, a different conformation for Glu 165 in the unbound state might interfere with catalysis. The more pertinent question relates to how far the reactive substrate atoms are from the carboxylate of Asp 165. Again, an increase of 1 Å is expected, assuming the substrate does not adjust its position. Modelling studies, however, suggest that even without substrate re-adjustment the distance
between the aspartic acid side chain and substrate does not necessarily have to increase (Alber et al., 1987a). This study is discussed in more detail in a later section.

The interest in this mutant increased when it was used as the starting point for a series of experiments that searched for second site suppressors (Hermes et al., 1989). The entire gene for this mutant was subjected to a "spiked oligonucleotide" method of random, \textit{in vitro} mutagenesis and was then used to transform an isomerase-minus strain of E. coli. A selection method based on differences in nutrient requirements for cells harboring different triosephosphate isomerase activities allowed colonies with increased catalytic activity to be easily identified. To date, six second site suppressors have been found on a total of five positions in the amino acid sequence: $10^{10}$, 96, 97, 167, and 233. Four of these position – 96, 97, 167, and 233 – have been identified as being strictly conserved. In one organism, alanine substitutes for glycine at position 10. The increase in activity in these pseudorevertants ranges from 1.3-fold to 19-fold the activity of the single E165D mutant (Hermes et al., 1989).

The experiments described in this chapter provide a starting point for further work on the structural and functional basis of increased activity for the pseudorevertants. The structure of the uncomplexed E165D mutant is described. However, to understand the structural source of the decreased activity of the single mutant and the increased activity of the pseudorevertants, structures of the enzymes in their

---

10 Due to an insertion in the chicken TIM sequence, position 10 corresponds to position 9 in the yeast sequence.
complexed forms are also needed. Conditions for the production of chicken TIM crystals complexed to PGA or PGH are also reported.

II. Materials and Methods

A. Chicken E165D

Crystals of the chicken TIM mutant E165D grew fortuitously from the protein stock solution left unattended at 4°C for a few months (Steve Blacklov, personal communication). The solution contained the protein at a concentration of 20 mg/mL in a buffer of 100 mM triethanolamine, pH 7.5 containing 5 mM EDTA and 60% ammonium sulfate. The crystals were isomorphous to those of the wild type enzyme. Most crystals had dimensions on the order of 0.5 mm X 0.5 mm X 0.5 mm, although there were a number that were as large as 1.0 mm in one of the dimensions.

Crystals were mounted at room temperature in quartz capillaries with a column of mother liquor on each side to keep the crystal from drying out. The capillaries were also sealed with dental wax. Zero-level 15° precession photos of the principle zones were recorded using Buerger precession cameras mounted on a sealed tube generator to confirm crystal isomorphism to the wild type enzyme. The crystals have the symmetry of the space group \(P2_1 \times 2_1\), with one dimeric molecule in the asymmetric unit. The unit cell of the crystals has dimensions \(a=105.3\ \text{Å}, b=74.5\ \text{Å},\) and \(c=61.7\ \text{Å}\

Suitability of crystals for data collection was determined on a Nicolet P3 4-circle diffractometer with a sealed tube operating at 1.2kW. Imperfections in the crystal lattice had been previously observed with the wild type enzyme (Bloomer, 1972); this mutant was no different. The most common imperfection occurred along the \(a^*\) axis and manifested itself as split peaks along reflections having indeces
(h,0,0). Crystal quality was most easily assessed by observing the peak profile of either reflection (8,0,0) or (10,0,0). All crystals screened were at least 0.5 mm X 0.4 mm X 0.3 mm in size. Only one out of 15 crystals screened had no split peak at these reflections. This crystal was used to collect data from 3.0 Å to 2.5 Å resolution. Measurements were made using the full integration method, scanning on ω for a total of 1°, counting for 45 seconds on the peak, and 15 seconds on the background collected 1° away from the calculated peak position. Five standard reflections were monitored during data collection to assess radiation damage. 300 low resolution reflections were also collected for purposes of scaling data from other crystals.

Since crystals without split peaks were difficult to find, it was decided that the crystal with the smallest peak splitting along the (8,0,0) or (10,0,0) reflection would be used to test whether the number of reflections with split peaks presented a significant problem. A shell of data between 2.9 Å and 3.5 Å was collected, processed (see below), and compared to the overlapping data from the "good quality" crystal. The merging R-factor on intensities for these two crystals was found to be 12%. It was concluded that the remaining low resolution data could be collected on this crystal without significantly compromising the quality of the structure.

The data were processed using the PROTSYS package of programs (Petsko, unpublished results) on a VAX 11-750. Radiation damage was corrected by an approximation of linear decay of intensity using the five standard reflections distributed through reciprocal space. An empirical method was used to correct for absorption (North et al., 1968). Initially, only data with background-corrected intensities
greater than two standard deviations were retained. The intensities from the two crystals merged with an R-factor of 12% for about 600 common reflections. In this reduced data set, only 37% of the possible data were observed. In order to increase the number of reflections, the data were re-processed using as a cut-off reflections having intensities greater than one standard deviation. The 700 common reflections for the data from both crystals merged with an R-factor of 13.5% on intensity. Adding the weaker data increased the total number of reflections that were observed to 52% of the theoretical number (data statistics are shown in Tables I and II).

Although it is possible to note differences in two structures using difference Fourier methods with as little as 30% of the data (Almo et al., 1989; in Bushy Stunt Virus, a Ca$^{2+}$ binding site was observed with only 1% of the data (Hajdu et al., 1989)), the use of more data gives a better impression of the quality of the structure. Inclusion of the weaker data should raise the final R-factor, but it more honestly reflects the quality of the structure.

A 2F$_o$-F$_c$ omit electron density map was calculated using the wild type chicken TIM coordinates to generate F$_c$'s and phase angles to 2.5 Å resolution (Peter Artemiuk, personal communication). The observed structure factors from the E165D mutant were used, while the calculated structure factor and phase ($e^{2\pi i \alpha}$) values were obtained from the wild-type chicken TIM coordinates in which Glu 165 from both subunits had been deleted. The map clearly showed the position of the Asp residues at position 165. The aspartic acid side chain was built into the electron density, and after minor adjustments to the local structure, the coordinates were refined using PROLSQ (Hendrickson, 1985). The

113
initial R-factor for data between 10 Å and 3.5 Å resolution was 27%. As the R-factor dropped in subsequent cycles, shells of data were incrementally added until all data to 2.5 Å resolution were included. The R-factor after 8 cycles of refinement converged to a value of 21.4%. At this point, 2Fo-Fc delete maps were used to manually rebuild parts of the structure. The above procedure was repeated three times before refinement was considered complete. The final model consists of 3,734 non-hydrogen protein atoms, two sulfate groups, and 34 water molecules. The rms deviation from ideality for bond lengths is 0.014 Å. The R-factor for data between 10 and 2.5 Å resolution is 17.6% (Table II). Considering the only average quality of the diffraction data, this is a very good final R-factor for a model with good stereochemistry.

B. In Search of Chicken TIM-PGH Crystals

As stated in the introduction, if anything is to be learned about the structural basis of the decreased catalytic activity of chicken TIM mutants and the increased activity of the pseudorevertants, structures of the bound form of these mutants need to be determined as well. Unfortunately, chicken TIM crystals have never been grown with anything other than sulfate ions in the active site. Substrate and PGA had been flowed into wild-type chicken crystals, but at the expense of crystal quality; data to only 6.0 Å resolution could be collected (Phillips et al., 1977b). Initial attempts to do the same with the mutant enzyme ended disasterously. Addition of PGA into a solution containing E165D crystals, followed by overnight incubation at 4°C to allow PGA to soak into the crystals, resulted in dissolution of the crystals. Two other strategies were approached in order to overcome this problem. Crystals of the E165D mutant were stabilized by crosslinking with gluteraldehyde,
before PGA or PGH was added. In addition, conditions to co-crystallize one of the pseudorevertants, E165D/S96P, were screened.

1. Crosslinked E165D crystals soaked with PGA or PGH. Crystals of E165D were soaked for three hours in a mother liquor solution containing 0.8% gluteraldehyde, and were then transferred to a mother liquor solution containing 0.15 M PGA. After overnight soaks, precession photos of crystals with and without PGH were taken (Figure 6). The binding of PGH causes some very small changes in the diffraction pattern. Inspection of some of the reflections (e.g., (0,14,3), (0,13,2), and (0,12,3)) reveals differences in the relative intensities, suggesting PGH may have bound.

2. Crystallization of PGH complexed to wild-type enzyme and the S96P/E165D pseudorevertant. The mutant S96P/E165D at a concentration of 20 mg/mL was obtained from Steve Blacklow (Harvard University) and dialyzed against 20 mM triethanolamine buffer (pH 7.5) containing 5 mM EDTA and 1 mM β-mercaptoethanol. Hanging drops were set up against concentrations of ammonium sulfate ranging from 55 to 80%. Each droplet was composed of 15 μL protein and 1 μL 0.15 M PGH. One crystal was found in each of the wells containing ammonium sulfate concentrations of 70 and 75%. The crystal in the 75% well (0.5 mm X 0.3 mm X 0.2 mm) was found to be that of a small molecule (G. Petsko, personal communication). Precession photos (Figure 7) of the crystal in the 70% well (0.3 mm X 0.25 mm X 0.1mm) indicated it had P2_12_12_1 symmetry and unit cell dimensions a=136.6 Å, b=73.8 Å, and c=54.5 Å.

Much larger crystals of wild-type chicken TIM (at least 0.5 mm on two dimensions) were also grown under these conditions. It has yet to

115
be established that these are isomorphous with those of the pseudorevertant.

There has been some question about whether the new crystal form has sulfate or PGH in the active site, since both are competitive inhibitors. The $K_i$'s (10mM for sulfate and 10$\mu$M for PGH) for the enzyme and the concentration of the inhibitors (3M sulfate, 6mM PGH) in the crystallization mixture suggest that PGH-complexed enzyme should be twice as concentrated as sulfate-complexed enzyme. In addition, the absence of crystals in the wells containing ammonium sulfate concentrations (55-60%) that usually crystallize the uncomplexed enzyme argues against sulfate-complexed enzyme in the crystals. The structure of this new crystal form of TIM remains to be determined.

III. Results


A number of methods have been established to estimate the errors in the coordinates of a structure. There is no one method other than that of independent multiple re-determination of a structure that is completely satisfactory.

The most commonly used methods require knowledge of the R-factor at different resolutions. In the Luzzati method, the differences between the observed and calculated structure factors are assumed to be entirely due to errors in the position of the atoms (an assumption not entirely appropriate for protein structures (see Kuriyan, 1986)). A relation between the overall error in the coordinates, and a dependence of the R-factor on resolution is derived (Luzzati, 1952). A Luzzati plot of the E165D chicken TIM structure indicates an error of about 0.25 Å (Figure 1).
The presence of identical subunits in an asymmetric unit can also be utilized for error determination, although there always exists the possibility that the error will be overestimated due to differences arising from crystal contacts. To overcome this problem it is best to eliminate all regions of the protein whose conformation is influenced by crystal contacts or are otherwise ill defined. In highly refined, high resolution structures ill-defined regions can be easily identified from their unusually high B-factors, which sometimes correspond to incorrectly modelled structures (Chambers and Stroud, 1979). The number of reflections observed in the E165D structure is less than three times the minimum number of parameters that need to be refined - the x,y,z coordinates for each of the 3,778 atoms in the model. Due to the low overdeterminancy of this structure, individual B-factor refinement is precluded.

All of the Cα's from the two subunits of the mutant superimpose with a r.m.s. difference of 0.7 Å. Close inspection of these superimposed Cα's indicates poor superposition of residues 2-3, 170-176, and 247. These regions correspond to the terminal ends of the protein and the surface loop. These regions have high B-factors and are known to be flexible in other structures of TIM (Chapter 2, this thesis). Elimination of these residues from the superposition lead to an r.m.s. difference of 0.5 Å. From the Luzzati and superposition analyses, it is safe to assume that the r.m.s. error in the atomic positions is between 0.25 Å and 0.50 Å. The least squares results from superposition of various structures (both for analysis of errors and comparisons of active sites) are summarized in Table III.
For reasons that are not entirely clear but may have to do with crystal field effects, the current model contains substantial differences in the two active sites (Figure 2). For example, the positions of the sulfate are over 1 Å apart. It is known that the two active site environments of wild type chicken TIM crystals are different. Attempts to flow DAHP into the crystal resulted in occupancy of only one active site (Phillips et al., 1977b). The explanation for this observation was that one active site resided at a crystal contact that prevented the flexible loop from closing over it. Differences in crystal environments, in addition to having steric consequences, might also affect the electrostatic potential, causing charged constituents of the active site to occupy slightly different positions.

To confirm that differences in the active sites were real and not due to any errors in model building, attempts were made to fit each active site into the density of the other subunit. Neither of the two models for the active site could be fit into the other electron density. Figure 3A and 3B shows the electron density at each active site. The only conclusion that can be drawn from this is that the enzyme has two responses to the mutation at position 165. Why each response appears separately in the two subunits (as opposed to appearing as two conformations in each subunit) is not known.

B. Subunit 1. The model and electron density for the active site of subunit 1 is shown in Figure 3A. The comparison of this active site to the wild type active site is shown in Figure 4. The most significant difference involves the conformational change of the side chain of Ser 96 toward the side chain of Glu 165. The conformational change at Ser 96 maintains the hydrogen-bond interaction (3.2 Å) between the
carboxylate (Asp 165) and hydroxyl (Ser 96) side chain atoms. Without this change in conformation, the interaction between these atoms would be sufficiently weak (3.8 Å) to allow the aspartic acid side chain to adopt some other conformation.

C. Subunit 2. The active site of subunit 2 has a different response to an aspartic acid at position 465. A hydrogen bond to Ser 396 does not exist. The conformation of the hydroxyl of Ser 396 is unchanged from wild type. Elimination of a C-C bond from Glu 465 leaves enough of a void to allow two water molecules (WAT 733 and WAT 736) to occupy positions on either side of the carboxylate oxygens of Asp 465. Comparison of this active site with that in the wild type (Figure 5) reveals the effects of this mutation are more isosteric than in subunit 1. Interestingly, the positions of the sulfates are different. This difference argues against the speculative explanation about different electrostatic potentials at the two (subunits 1 and 2) active sites. If the previous explanation had been correct, the position of the sulfate in subunit 2 of wild type and mutant enzyme would have to be the same — unless the mutation influences the electrostatic potential of the second subunit.

IV. Discussion

In site-directed mutagenesis of enzymes, active site residues are replaced, and the effects on catalysis and specificity are assessed (Knowles, 1987; Gerlt, 1987). The uncomplexed structure of such a mutant in triosephosphate isomerase has been described in this Chapter. The structural effect of the mutation E165D is local. However, more

11 To avoid confusion, the first subunit is numbered from 2 to 248 and the second subunit is numbered from 302 to 548.
than one conformation for the new side chain is observed. The hydroxyl group of Ser 96 is hydrogen bonded to Glu 165 and the backbone nitrogen of Glu 97 in both subunits of the wild type enzyme. In subunit 1 of the mutant enzyme, the side chain of Ser 96 rotates toward the carboxylate of Asp 165 to maintain that hydrogen bond, but in the process breaks the hydrogen bond to the backbone nitrogen of Glu 97. In the second subunit, the side chain of Ser 396 remains unchanged. The hydrogen bond to Glu 397 remains, but there is now no hydrogen bond to Asp 465. The side chain conformation of Asp 465 is maintained by a network of hydrogen bonds with water molecules that have filled the void left by the mutation. Changes in the distribution of bound solvent in a mutant have been observed before. For example, it has been found that mutant proteins can be stabilized when water molecules fill the void and replace the hydrogen-bond interactions of the deleted residue (Alber et al., 1987b).

Although the structure of the mutant E165D complexed to PGH has yet to be determined, interesting results have been obtained from computer simulation experiments. Ab-initio self-consistent field calculations on abstraction of a proton from hydroxyacetone by formate attempt to explain the observed rate reduction purely on the basis of the longer oxygen-carbon distance in the transition state (Alagona et al., 1986). Energy minimization of the mutant enzyme complexed to DHAP suggests otherwise (Alber et al., 1987a). Computer graphics was used to position the aspartic side chain in the otherwise wild type enzyme. Using the minimum perturbation approach (Shih et al., 1985), two conformations separated by large energy barriers were found. One conformation was discounted because the carboxylate was thought to be too far from the
substrate (>5.0 Å) to be catalytically relevant. The other conformation is a consequence of an attempt by the enzyme to release steric crowding due to placing a bulky carboxylate closer to the protein backbone. The conformation places a carboxylate oxygen 2.9 Å from the Cβ atom of DHAP, almost identical to that of wild type. However, the orientation of the carboxylate oxygen towards the substrate is different. The anti orbital of the oxygen is in a position that would make it the proton acceptor. In the wild-type enzyme, it is the syn orbital that is the proton acceptor. It has been proposed that the anti orbital is 10,000-fold less basic than the syn orbital (Gandour, 1981), and this may also be an explanation for the kinetic data. These speculations underscore the importance of determining the structure of this mutant in its complexed form.

Although a lot of information has been gathered from site-directed studies, the design of these types of experiments is biased by what the experimentalist believes is important. The use of random mutagenesis combined with selective screens circumvents the bias problem and allows nature to declare what factors are important for catalysis. In triosephosphate isomerase six pseudorevertants have been generated (Hermes, et al., 1989). Three of these pseudorevertants (E165D/E97D; E165D/V167D; and E165D/G233R) involve changes of charge around the active site and may affect catalysis by altering the electrostatic potential at the active site. Another one (E165D/S96T) introduces a methyl group into the active site, while another (G105) introduces a hydrogen bonding group. The pseudorevertant with the highest activity (E165D/S96P) removes a hydrogen bonding functional group and at the same time introduces a proline into the helix (D7) whose dipole moment is
thought to play a functional role. The factors that govern the increase in catalytic activity for these mutants will not be precisely deduced until their complexed structures are determined and compared to the complexed structure of the single mutant. The new crystal form of chicken TIM complexed to PGH should help in that goal.

While crystallization attempts were proceeding on the PGH-complexed pseudorevertant (E165D/S96P) from chicken, the same changes were being introduced into the yeast enzyme, which is known to crystallize more readily in the presence of PGH or PGA. Preliminary observations indicate that this double mutant is not a pseudorevertant in the yeast enzyme (Betsy Komives, Harvard University). If this observation is confirmed by further experimentation, it adds another parameter to the already complicated sequence/structure/function equation. Not only will the increase in activity of the chicken pseudorevertant relative to the single mutation need to be explained, but the differences in activity in the double mutants from the two organisms will also need to be addressed. The new crystal form of chicken TIM complexed to PGH should help in that goal, too.
Table I

Data Statistics

Resolution 57.7 to 2.50 Å
Reflections 8,934

A. Magnitude of F

<table>
<thead>
<tr>
<th>Min F</th>
<th>Max F</th>
<th># of Reflections</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>99</td>
<td>1,037</td>
</tr>
<tr>
<td>100</td>
<td>199</td>
<td>4,971</td>
</tr>
<tr>
<td>200</td>
<td>299</td>
<td>1,992</td>
</tr>
<tr>
<td>300</td>
<td>399</td>
<td>684</td>
</tr>
<tr>
<td>400</td>
<td>499</td>
<td>180</td>
</tr>
<tr>
<td>500</td>
<td>599</td>
<td>50</td>
</tr>
<tr>
<td>600</td>
<td>699</td>
<td>17</td>
</tr>
<tr>
<td>700</td>
<td>799</td>
<td>2</td>
</tr>
<tr>
<td>800</td>
<td>899</td>
<td>1</td>
</tr>
</tbody>
</table>

Minimum F 28
Maximum F 808
Average F 183

B. F/σ(F)

<table>
<thead>
<tr>
<th>Minimum</th>
<th>Maximum</th>
<th># of Reflections</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>396</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>1,103</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>1,003</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>733</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>617</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>506</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>409</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>413</td>
</tr>
<tr>
<td>10+</td>
<td></td>
<td>3,754</td>
</tr>
</tbody>
</table>
Table II

Refinement Statistics

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of protein atoms</td>
<td>3,734</td>
</tr>
<tr>
<td>Number of water molecules</td>
<td>33</td>
</tr>
<tr>
<td>Number of atoms in two sulfates</td>
<td>10</td>
</tr>
<tr>
<td>Total number of atoms</td>
<td>3,778</td>
</tr>
<tr>
<td>Overall B-Factor</td>
<td>10.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Description</th>
<th>Applied, $\sigma$(Å)</th>
<th>Deviations, $\Delta$(Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond Distance</td>
<td>0.020</td>
<td>0.014</td>
</tr>
<tr>
<td>Angle Distance</td>
<td>0.040</td>
<td>0.054</td>
</tr>
<tr>
<td>Planar 1-4 Distance</td>
<td>0.050</td>
<td>0.058</td>
</tr>
</tbody>
</table>

# of Bad Dihedral Angles$^{13}$ 14

<table>
<thead>
<tr>
<th>Resolution</th>
<th># of Reflections</th>
<th>% of Data</th>
<th>Shell R-factor</th>
<th>Sphere R-factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.0-5.00</td>
<td>1,569</td>
<td>78.9</td>
<td>0.211</td>
<td>0.211</td>
</tr>
<tr>
<td>5.00-4.00</td>
<td>1,579</td>
<td>75.1</td>
<td>0.144</td>
<td>0.174</td>
</tr>
<tr>
<td>4.00-3.50</td>
<td>1,333</td>
<td>64.3</td>
<td>0.149</td>
<td>0.167</td>
</tr>
<tr>
<td>3.50-3.00</td>
<td>1,097</td>
<td>29.7</td>
<td>0.178</td>
<td>0.168</td>
</tr>
<tr>
<td>3.00-2.75</td>
<td>1,540</td>
<td>52.1</td>
<td>0.192</td>
<td>0.172</td>
</tr>
<tr>
<td>2.75-2.50</td>
<td>1,522</td>
<td>36.0</td>
<td>0.199</td>
<td>0.176</td>
</tr>
<tr>
<td>10.0-2.50</td>
<td>8,660</td>
<td>52.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^{12}$ The $\Delta$ values are r.m.s. deviations from the corresponding values for ideal groups derived from small molecular structural studies (Sielecki et al., 1979).

$^{13}$ Residues with dihedral angles significantly outside Ramachandran limits: Arg 4, Lys 13, Asp 36, Asn 65, Ala 169, Ile 170, Thr 172, Lys 174, Thr 175, Cys 126, Ala 246, Lys 313, Val 496, Ala 546.
<table>
<thead>
<tr>
<th>Structures</th>
<th>Superimposed Atoms</th>
<th>R.M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subunit 1, E165D</td>
<td>Cα: all</td>
<td>0.74 Å</td>
</tr>
<tr>
<td></td>
<td>Cα: 4-169, 177-246</td>
<td>0.50 Å</td>
</tr>
<tr>
<td></td>
<td>all atoms: 13,95,97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>backbone atoms: 96,165</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>cTIM</td>
<td></td>
</tr>
<tr>
<td>E165D</td>
<td>Cα: 2-548</td>
<td>0.32 Å</td>
</tr>
<tr>
<td></td>
<td>Cα: 2-248</td>
<td>0.33 Å</td>
</tr>
<tr>
<td></td>
<td>Cα: 302-548</td>
<td>0.28 Å</td>
</tr>
<tr>
<td></td>
<td>all atoms: 13,95,97,313,395,397</td>
<td></td>
</tr>
<tr>
<td></td>
<td>backbone atoms: 96,165,396,465</td>
<td>0.41 Å</td>
</tr>
<tr>
<td></td>
<td>all atoms: 13,95,97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>backbone atoms: 96,165</td>
<td>0.43 Å</td>
</tr>
<tr>
<td></td>
<td>all atoms: 313,395,397</td>
<td></td>
</tr>
<tr>
<td></td>
<td>backbone atoms: 396,465</td>
<td>0.33 Å</td>
</tr>
</tbody>
</table>
FIGURE LEGENDS

Figure 1  A Luzzati plot of the output master data file from PROLSQ refinement. The reference lines indicate the data distributions expected from errors of 0.15 Å, 0.20 Å, 0.25 Å, 0.30 Å, 0.35 Å, and 0.40 Å (bottom to top).

Figure 2  The superimposed active site residues (Lys 13, His 95, Ser 96, Glu 97, Glu 165) from both subunits of the chicken E165D mutant using least square methods.

Figure 3  (a) An Fo-Fc delete map of the active site of subunit 1 contoured at 2.0 σ. (b) A 2Fo-Fc delete map of the active site of subunit 2 contoured at 1.5 σ. In both figures, Asp 165 is at the upper left, Lys 13 at the bottom and the imidazole of His 95 at the middle. The two waters adjacent to the carboxylate oxygens of Glu 465 are also shown in Figure 3b, however one water has no density. Its density comes up at lower contour levels.

Figure 4  Comparison of the mutant and wild type active sites in subunit 1.

Figure 5  Comparison of the mutant and wild type active sites in subunit 2.

Figure 6  Precession photo of gluteraldehyde crosslinked E165D crystal soaked in mother liquor containing PGA.

Figure 7  Precession photo of gluteraldehyde crosslinked E165D crystal soaked in mother liquor without PGA.

Figure 8  Precession photo of co-crystal of PGH and the pseudorevertant S96PE165D.
LUZZATI ANALYSIS, UNCOMPLEXED CHICKEN E165D

R VALUE

0.0 0.1 0.2 0.3 0.4 0.5

RESOLUTION (1/2D)
CHAPTER 5

MOLECULAR BIOLOGY OF YEAST TRIOSEPHOSPHATE ISOMERASE
I. INTRODUCTION

The combination of x-ray crystallography and mutagenesis, whether random or site-directed, is powerful. It allows the roles of individual amino acids at the active site to be defined precisely (Knowles, 1987; Gerlt, 1987). The two techniques have also been effectively used to enhance catalytic efficiency or alter specificity. The number of enzymes of known structure that have been studied by mutagenesis is small but rapidly growing. This chapter describes the cloning, expression, and mutagenesis of yeast triosephosphate isomerase. The Appendix contains the purification and partial kinetics of one of the mutants.

The following discussion is meant to alert others who work on this project of the availability of the mutagenic primers listed in Table I, and is motivated by the hope that these mutants will eventually be made and characterized.

In order to test the effects of mutations at the subunit interface on protein stability and enzyme activity, mutagenic primers to convert Thr 45 to Cys and to Asp were synthesized. The unrefined structure of yeast TIM (the structure being used at the time these experiments were designed, indicated that Thr 45 interacted with its symmetry-related partner in the other subunit. It was hoped that the cysteine mutation would be oxidized in the purified protein to form a disulfide bond, and stabilize the protein to denaturation. The mutation to an aspartic acid was meant to test the effects of placing a potential negative charge in the hydrophobic subunit interface. It has subsequently been found that

14 Some enzymes in this category are Tyr RNA synthetase, dihydrofolate reductase, subtilisin BPN, trypsin, aspartate aminotransferase, staphylococcal nuclease, carboxypeptidase, triosephosphate isomerase, aspartate transcarbamylase, alkaline phosphatase, ribonuclease, cytochrome c peroxidase, α-lytic protease.
replacement of buried asparagines by aspartic acid destabilizes the enzyme (Ahern et al., 1987; Casal et al., 1987).

One other mutant was designed to have a dramatic effect on the structure. Ala 234 is found in the middle of a short helix which has been proposed to bind the phosphate. The residues at the amino terminal end of this helix (Gly 232 and Gly 233) form hydrogen bonds with the terminal oxygens on the phosphate. It is believed that the microdipole moment of the helix also contributes to phosphate binding (Alber, 1981). In order to test this hypothesis without introducing a negative charge at the amino terminal end, it was decided to mutate a residue in the middle of the helix (Ala 234) to a proline, a residue that abhors being in the middle of a helix (Richardson and Richardson, 1988). This substitution was thought to be potentially very informative. The response of a helix to the introduction of a proline in its center could be assessed. If the helix were found to be significantly distorted, it is likely that the strength of the microdipole moment would have been diminished, and the effect on binding and catalysis would have been determined.

Another region of interest for site-directed studies was the flexible loop. There was speculation that the loop functioned to prevent phosphate elimination (Alber et al., 1981b). Mutations were designed to test this hypothesis. In an earlier yeast TIM structure, the side chain of Thr 172 was modelled in a conformation that indicated the hydroxyl would interact with the phosphate in the closed form. Based on this structure, threonine was mutated to an aspartic acid in anticipation that the two negative charges (one in the loop and one on the substrate) would prevent the loop from closing over the substrate.
Subsequent refinement and rebuilding produced a structure that indicated the side chain at this position was directed away from the substrate and into the solvent. Partial kinetic characterization of this mutation is discussed in a later section.

Another oligomer was designed to remove five residues from the middle of the loop (residues 170 to 174). A similar mutation has been made independently and recently characterized at Harvard (Pompliano et al., 1989). The results of that mutant will also briefly be discussed in a later section.

The other mutations were designed to probe the role of strictly conserved residues. It had been speculated that the side chains of Cys 126 and Ser 211 were positioned on either side of Glu 165 to act as anchor for the two essential conformations of catalysis - that of abstraction of a proton from substrate, and reprotonation of the intermediate. A number of mutations could have been planned, all in the category of "gee-whiz mutagenesis - the change and see what happens approach" (Knowles, 1987). Removal of the functional group (to Ala) was an obvious mutant to make. Another mutation was to change the polarity of the interactions (Cys to Ser and Ser to Cys) to see what happens. It is now believed that both these residues are too far from the carboxylate to form interactions of any significance. Nonetheless, the reasons for the strict conservation remains undetermined.

The remaining mutation - Ser 96 to Ala - was designed to remove a functional group from a strictly conserved residue. The hydroxyl of Ser 96 forms a hydrogen bond to the carboxylate of Glu 165 in the uncomplexed enzyme. Upon substrate binding, this interaction is broken and the side chain of Glu 165 moves 2 Å toward the substrate. Although
the exact nature of this hydrogen bond on the overall rate or mechanism of the reaction is not known, this position is one of the sites that leads to pseudoreversion of the E165D mutant in chicken TIM. Interestingly, the same double mutant in the yeast enzyme is not a pseudorevertant. Inspection of the wild type yeast structure reveals that the only residue with which Ser 96 interacts that is not common to both the chicken and yeast enzymes is Ser 100, which is a histidine in chicken TIM. Perhaps additional mutagenesis on this residue might convert the yeast double mutant into a pseudorevertant.

The above introduction indicates that a full agenda for mutagenesis was in place. All that remained was to overexpress the enzyme.

II. EXPERIMENTAL PROCEDURES

A. Materials

E. coli strains JM101 (Δlac, pro, SupE, thi1, F', proAB+, lac i+, lac z ml5tra Δ36) and TG1 were obtained from J. Messing and Amersham Corp., respectively. DF502 ([Δ(rha, pfkA, tpi) pfkB1, his-, pyrD-, edd-, F-, str']) was a construct of Dr. Dan Fraenkel (Alber & Kavasacki, 1982).

The plasmids pTP1-12, pUC18, pKK223, and pBSyTPI were obtained from R.C. Davenport, J. Messing, M. Distefano, and Patti Lodi, respectively.

Restriction endonucleases, bacterial and calf intestine alkaline phosphatase, T4 polynucleotide kinase, and T4 DNA ligase were purchased from IBI, Boehringer Mannheim, and England Biolabs and were used under conditions specified by the manufacturer.

Klenov fragment of DNA polymerase I, M13 sequencing primer, deoxyadenosine and deoxycytosine 5'-(α-35S)thiophosphate (>600 Ci/mmol) and adenosine 5'-[γ-32P]triphosphate (>5000 Ci/mmol) were
purchased from Amersham Corp. Deoxyribonucleotide triphosphates and
dideoxyribonucleotide triphosphates were either from Amersham Corp. or
Pharmacia.

Analytical-grade agarose and low melting point agarose was from
BRL. For in-gel ligations NuSieve GTG agarose from FMC Corporation was
used.

Solid media for bacterial growth (yeast extract, tryptone, and
agar) were from Difco.

Inorganic salts, PEG 6000, and glycerol for growth media were
purchased from Mallinckrodt or Baker and sterilized prior to use.

Anhydrous Ampicillin and Streptomycin Sulfate were purchased from
Sigma and usually stored at -20° C as 25 mg/mL and 20 mg/mL solutions,
respectively.

D,L-glyceraldehyde-3-phosphate diethyl acetal and yeast
glyceraldehyde-3-phosphate dehydrogenase were purchased from Sigma.

Buffers and Media:
TE: 10 mM Tris, 1 mM EDTA
10X Buffer A: 0.2 M Tris-HCl, 0.1 M MgCl₂, 0.5 M NaCl, 0.01 M DTT, pH
    7.5.
10X Buffer B: 0.2 M Tris-HCL, 0.1 M MgCl₂, 0.1 M DTT, pH 7.5.
Prehybridization Solution: 3 mL of 20X SSC, 1 mL 100X Denhardt’s, 0.2 mL
    10X SDS, 5.8 mL H₂O.
Hybridization Solution: 3 mL of 20X SSC, 1 mL 100X Denhardt’s, 6.0 mL
    H₂O.
20X SSC: 3 M NaCl, 0.3 M sodium citrate, 10 mM EDTA, pH 7.2
100X Denhardt’s Solution: 2% Bovine serum albumin, 2% polyvinyl
    pyrrolidone, 2% Ficoll.
Media were prepared according to recipes in Maniatis et al. (1982).

B. Methods

1. Growth of Bacteria. Cells used in this work were grown in either LB, 2 YT, or minimal (M9) media with appropriate antibiotics and nutrients. Agar plates were prepared according to established laboratory procedures (Maniatis et al., 1982).

2. Preparation of M13 phage and single-stranded DNA. M13 phage and single stranded DNA for use as mutagenesis templates were prepared using a modification of the procedure described in the M13 cloning manual (Amersham Corp.). One liter of M13 phage-inoculated JM101 was grown overnight in either M9 or LB media. The solution was centrifuged at 5,000 rpm for 30 minutes. Polyethylene glycol (PEG) 6000 and NaCl were added to the supernatant at a final concentration of 5% and 0.5 M, respectively, and left for at least 1 hour at 4°C. The solution was centrifuged again at 5,000 rpm and the pellet was resuspended in 4 mL TE. PEG 6000 and NaCl were added to final concentrations of 4% and 0.5 M, respectively, and left for 30 minutes at 4°C. The solution was centrifuged at 10,000 rpm for 15 minutes and the pellet was resuspended in 4 mL TE. The DNA was extracted twice with phenol and once with chloroform. Finally, 400 μL of 3 M sodium acetate and 8 mL of ethanol were added to precipitate the DNA on ice. After centrifugation at 10,000 rpm for 30 minutes, the DNA pellet was dissolved in 0.5 mL TE and quantitated spectrophotometrically as described (Maniatis et al., 1983).

Growth and purification of single stranded DNA from plaques was as described in the M13 cloning manual.

3. Synthesis and purification of oligonucleotides. Biosearch procedures were followed in the synthesis and purification of
oligonucleotides. Oligonucleotides were synthesized on either a Biosearch Sam One DNA synthesizer using phosphotriester chemical methods or on a Biosearch 8600 DNA synthesizer using phosphoramidite chemistry. The solid phase product from the Sam One synthesizer had to be treated with 1,1,3,3-tetramethyl guadinine and 2-pyridine aldoxime, after which, 1 mL of concentrated ammonium hydroxide was added to the resin. The product from the 8600 DNA synthesizer is automatically deblocked and detached from the support, and in a solution of ammonium hydroxide. All steps thereafter are the same regardless of which chemistry is used in the synthesis. The ammonium hydroxide solution containing the DNA was heated at 55° C for 5 hours and lyophilized using a Savant Speed-Vac. The target oligonucleotide was purified from smaller contaminating oligonucleotides using preparative polyacrylamide gel electrophoresis, followed by elution of the corresponding band with 100 mM ammonium bicarbonate (50° C, 10 minutes, followed by 37° C, overnight). The DNA was desalted using a Sep-pak C_{18} column. To confirm that the instruments were synthesizing the desired sequences, some of the oligonucleotides were sequenced by the Maxam and Gilbert method (1980). The oligonucleotides synthesized for this project are shown in Table I.

4. Mutagenesis reactions.

(a) T172D. Mutagenesis was performed using the standard two primer procedure (Zoller and Smith, 1984). All mutagenesis primers were 5'-phosphorylated with kinase. The annealing reaction was done with the phosphorylated M13 sequencing primer (10 pmol; Amersham Corp.), the mutagenic 15mer (10 pmol), single-stranded M13yTPI template DNA (0.5 pmol) and 1 μL buffer A in a total volume of 10 μL. This solution was heated at 55° C for 5 minutes and then left at room temperature for 5
minutes to anneal. Ten μL of an extension/ligation solution containing 4 μL 2 mM dNTP's, 1 μL 10 mM rATP, 1 μL 10x buffer B, 3 units T4 DNA ligase, and 2 units of Klenow fragment were added to the annealing mixture and allowed to incubate at 15° C for 14 hours. Ten μL of the mutagenesis solution were then used to transform 100 μL CaCl₂-treated JM101. After growth for 18 hours on M9 plates, 92 plaques were individually grown in 1 mL M9 liquid media. The phage particles were separated from JM101 by mini-centrifugation. Dot blot hybridization was used to screen for mutants by applying 200 μL of each phage sample on nitrocellulose, baking at 80° C in vacuo for 1-2 hours, washing for 1 hour at 67° C with prehybridization solution, and hybridizing with the [γ-³²P]-labelled mutagenesis primer. Hybridizations as high as 50° C indicated there were five potential mutants from a total of 72. One of these was used to re-infect JM101, from which another plaque was isolated and sequenced to confirm the mutation.

(b) C126S. The optimum annealing ratio (primer/template) for this reaction was found to be 100. Mutagenesis was accomplished using 1 pmol ssM13yTPI template and 95 pmol phosphorylated Ser 126 in a 10 μL solution containing 15 pmol M13 sequencing primer and 1.5 μL Buffer A. The exact procedures as in the Asp 172 mutant were repeated except that the mutagenesis solution was diluted 5-fold for the transformation. 72 plaques were grown, and six of these were presumed mutants from dot blot hybridization. JM101 was re-infected and a single plaque was grown and sequenced to confirm the mutation.

(c) ΔLoop(170-174). The deletion mutagenesis protocol developed at Genentech was used to delete the nucleotides coding for amino acids 170 to 174 (Adelman et al., 1983). Annealing of the oligomer, extension
and ligation were done under the same conditions as for the T172D and C126S mutants. An aliquot of the reaction mixture was saved as a control for subsequent transformation. The remaining mixture was phenol extracted, and the purified DNA treated with S_1 nuclease to digest the single-stranded DNA coding for the loop region to which the oligomer was not complementary or DNA which had not been completely extended. This procedure enriched the mixture for the mutated genotype. This mixture and the non-S_1 nuclease-treated mixture were used to separately transform JM101. The non-S_1 nuclease-treated mixture produced 25 plaques, one of which was identified as the deletion mutant by autoradiography. The S_1 nuclease-treated mixture produced 50 colonies, six of which were positive for the deletion mutant. One of these was used to transfec JM101, from which DNA from a single colony was isolated and sequenced to confirm the mutation.

5. Preparation of double-stranded DNA. Double-stranded DNA, either from a plasmid or from the replicative form of M13 phage, was prepared by the method of Birnboim and Doly (1979). The DNA was resuspended in an appropriate volume of TE, 1 μg RNAse added, and the solution left at room temperature for at least 30 minutes. The DNA was extracted with phenol and chloroform, precipitated with ethanol, and redissolved in TE. The DNA purified by this procedure was pure enough for all subsequent manipulations.

6. Ligation and transformations. A number of different ligation methods were used during the course of this work. The method of choice was one developed at BRL for fragments electroeluted from gels (King and Blakesley, 1986) or of Dumais and Nochumson (1987) for "in-gel ligations".
Competent cells were prepared according to the method of Dagart (1979) and used immediately or stored at -70° C.

III. RESULTS

A. Cloning the yTPI Gene Into M13mp18

The original plasmid containing the yTPI was constructed using a series of manipulations (Figure 1) (Davenport, 1986). A 3-kilobase Pst I-Hind III fragment of pTPIC10 (Alber and Kawasaki, 1982) containing the yTPI gene was inserted into the homologously cleaved pUC18. The Pst I site of the resulting plasmid fell between the 3' end of the lac promoter and the start site of the yTPI gene. To decrease the number of bases between the 3' end of the lac promoter and the yTPI gene, the plasmid was linearized with Pst I, and digested with Bal31 nuclease. After blunt-ending with S1 nuclease, the terminal ends were ligated together, and the mixture of derivative plasmids was used to transform DF502. Amp' colonies were selected on LB plates and assessed for TIM activity. Plasmid from one of these colonies with TIM activity was purified. Unfortunately, the plasmid that was purified for further work had a problem. A very large portion of the plasmid, including the 5' end of the yTPI promoter, most of the lac mRNA coding region, the lac Z Shine-Delgarno sequence, and most of the lac operator were deleted (Davenport, 1986). Not only was this deletion probably the cause for levels of expression for the enzyme that were lower than anticipated, but it also made cloning into M13mp18 for mutagenesis less straightforward since the multiple cloning site had also been deleted.

The gene was transferred into Sma I-digested (which forms blunt ends) M13mp18 DNA by digesting the plasmid containing the gene with Hae II and blunt ending with Klenow fragment (Figure 2). The resulting DNA
was used for some of the site-directed mutagenesis even though this construction eventually proved to be inadequate for cloning into a high expression vector. The nearest, convenient restriction site on the 5' end of the \textit{yTPI} gene was 400 bases away from the start site. To overcome the problem of having too many bases between the \textit{yTPI} start site and any potential promoter, a mutation was made ten bases upstream of the \textit{yTPI} AUG codon to introduce an Acc I site. The resulting mutagenesis vector could be conveniently cloned into any plasmid's multiple cloning site (Figure 2).

\textbf{B. Construction of Various Expression Vectors}

During the course of this project, a number of expression vectors were constructed. Previous to this thesis work, plasmids pTPIC10, p12, and pUCyTPI were used. From none of these could more than 10 mg of protein be purified from a 10 L culture. For most experiments 10 mg of enzyme is usually a sufficient amount, but crystallization of yeast TIM has consistently been successful only with batch methods which require two milligrams of enzyme per trial, and not all trials produce quality crystals. In light of results from other laboratories that were obtaining hundred milligram quantities of other proteins, it seemed worth the effort to try to increase the yields.

The gene, with the additional 400 bases upstream of the start site, was cloned into pUC18 to form pUCyTPI. After the Acc I mutation was made in M13yTPI, the gene was cloned into pKK-223. Surprisingly, this new construction, with 400 bases removed between the start site and the more robust tac promoter, improved expression by only about 50%. It was not until M13yTPI/Acc was further mutagenized (to introduce more
restriction sites) and cloned into the PBS phagemid by Patti Lodi (Harvard University) that expression was substantially increased.

1. pUCyTPI(T172D). This mutant was cloned directly into pUC18 from M13yTPI. The mutant gene was excised by Ava I and Cla I and inserted into the Ava I, Acc I, phosphatase-treated pUC18. The resulting expression vector had 400 extra bases between the lac promoter and the AUG start site, but was nevertheless used to purify a few milligrams of TIM (see Appendix I).

2. PBS-Δ. Part of the gene containing the deleted region was cloned into PBSyTPI using unique restriction sites within the gene. The BglII, EcoRV fragment containing the deleted region was cleaved from M13yΔ, and purified from an agarose gel. The large fragment from the homologously cleaved PBSyTPI was also purified from an agarose gel, into which the mutant region was ligated.

IV. APPENDIX TO CHAPTER 5: PURIFICATION AND PRELIMINARY CHARACTERIZATION OF LOOP MUTANTS

A. Purification of T172D

Cells containing this mutant in pUCyTPI (pUCyT172D) were grown in 10 L of LB/Amp containing 1% glycerol and 0.5% lactate. After 11 hours of growth at 37° C the cells attained an OD₆₀₀ of 2.2 and were harvested.

The cells were lysed in 100 mL phosphate buffer (10 mM, pH 7.0) containing 7% glycerol, 0.3 mL 10% triton X-100, and 50 mg lysozyme and 0.7 mg DNAse. After centrifuging, the 65%-95% ammonium sulfate fraction was dialyzed against PBS, applied to a DEAE-cellulose (Sigma) column,
and eluted with PBS. The pooled fractions containing TIM activity were concentrated before subjection to immunoadsorption chromatography on immobilized rabbit antibodies specific for yeast TIM. The protein was eluted with 3 M KSCN, dialyzed immediately against phosphate buffer, and shown to be greater than 95% pure by SDS-PAGE.

B. Purification of Δ-loop

Cells containing pBSΔ were grown overnight in 2 L LB/Amp/Str to an OD_{600} of 2.1. SDS-polyacrylamide gel of the crude extract indicated that expression levels were lower than in the medium which was used as the inoculant (Figure 3), but this may have been due to non-optimization of growth conditions. This phenomenon has been observed by others and causes at the genetic level cannot be ruled out. The cells were suspended in 10 mL TE (containing 1 mM mercaptoethanol) and lysed with a French press. The cell debris was ultracentrifuged at 30,000 rpm, for 2 hours at 4° C. Ammonium sulfate was added to the supernatant, and the 60%-90% fraction was dialyzed against Tris, applied to an FPLC column, and eluted with a potassium chloride gradient (0-150 mM). The fraction which usually corresponds to TIM had no detectable activity in a crude assay. SDS-PAGE indicated a number of high molecular weight impurities that needed to be removed prior to more rigorous kinetic characterization.

C. Preliminary Kinetics of the T172D Mutant

TIM activity in the direction of DHAP formation was measured under two slightly different conditions. At 28° C, one reaction buffer contained 100 mM triethanolamine, pH 7.9, 5 mM EDTA, 0.2 mM NADH, 0.017 mg/mL glycerophosphate dehydrogenase, and 5.8 ng/mL mutant. At 30° C, the other reaction buffer contained 66 mM triethanolamine, pH 7.6, 3.3
mM EDTA, 0.02 mg/mL glycerophosphate dehydrogenase, and 10.2 ng/mL mutant. Data were also collected for the wild-type enzyme (Sigma), and each experiment was performed twice. The kinetic data are summarized in Table II.

The differences between these numbers are a bit too large to be accounted for by the different experimental conditions, so they should be regarded tentatively. It is probably safe to conclude, however, that $k_{cat}/K_m$ is down about an order of magnitude relative to the wild type enzyme. It appears that both $k_{cat}$ and $K_m$ contribute to the decrease in the pseudo second-order rate constant. Since the loop is known to form interactions with the substrate phosphate, it is not surprising that binding of the Michaelis complex is weakened. The presence of a negative charge on the loop should adversely affect the energetics of interaction between loop and phosphate. It might seem surprising that the mutation also lowers $k_{cat}$. But if the loop really functions to exclude bulk solvent from the active site and enhance the electrostatic interactions within it, disruption of this function could very well affect $k_{cat}$. Indeed, experiments with the chicken loop deletion mutant (residues 170 to 173 were deleted) prepared at Harvard indicate that $K_m$ is much higher, $k_{cat}$ is very much decreased (down 10,000-fold), and large amounts of methyl glyoxal are produced (Pompliano et al., 1989).
<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-TCC-TGG-AGT-AGC-AGC-GGC-CCA-GAC-TFF-TCC-3'</td>
<td>Loop Deletion</td>
</tr>
<tr>
<td>5'-ACC-CAA-CTT-GGA-AGC-CA-3'</td>
<td>Loop Sequencing Primer</td>
</tr>
<tr>
<td>5'-CAA-ACC-GTC-ACC-AAT-3'</td>
<td>Thr 172 \rightarrow Asp</td>
</tr>
<tr>
<td>5'-TAA-GTA-GCA-AGC-TGG-3'</td>
<td>Thr 45 \rightarrow Cys</td>
</tr>
<tr>
<td>5'-GTC-TAA-GTA-GTC-AGC-TGG-AGG-3'</td>
<td>Thr 45 \rightarrow Asp</td>
</tr>
<tr>
<td>5'-C-ACC-GAT-ACT-CAA-GAT-G-3'</td>
<td>Cys 126 \rightarrow Ser</td>
</tr>
<tr>
<td>5'-CAA-AGA-AGG-ACC-ACC-3'</td>
<td>Ala 234 \rightarrow Pro</td>
</tr>
<tr>
<td>5'-T-TCT-TTC-GGC-GTG-ACC-C-3'</td>
<td>Ser 96 \rightarrow Ala</td>
</tr>
</tbody>
</table>

The following oligonucleotides exist but their sequence has not been confirmed:

Cys 211

Pro 96

Cys 172\textsuperscript{15}

\textsuperscript{15} Should be sequenced. There is evidence that this primer does not contain the correct sequence.
<table>
<thead>
<tr>
<th>ENZYME</th>
<th>T(°C)</th>
<th>pH</th>
<th>Tris (mM)</th>
<th>K_m\text{GAP} (mM)</th>
<th>k_{cat} (sec(^{-1}))</th>
<th>k_{cat}/K_m (M(^{-1})s(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>wild type</td>
<td>28</td>
<td>7.9</td>
<td>100</td>
<td>0.66</td>
<td>3730</td>
<td>5.65 \times 10^6</td>
</tr>
<tr>
<td>T172D</td>
<td>28</td>
<td>7.0</td>
<td>100</td>
<td>2.47</td>
<td>784</td>
<td>3.17 \times 10^5</td>
</tr>
<tr>
<td>wild type</td>
<td>30</td>
<td>7.6</td>
<td>66</td>
<td>0.72</td>
<td>6360</td>
<td>8.83 \times 10^6</td>
</tr>
<tr>
<td>T172D</td>
<td>30</td>
<td>7.6</td>
<td>66</td>
<td>1.23</td>
<td>509</td>
<td>4.14 \times 10^5</td>
</tr>
</tbody>
</table>
FIGURE LEGENDS

Figure 1  Cloning of yTPI from pTPIC10.

Figure 2  Cloning of yTPI into various expression vectors.

Figure 3  Expression levels of yTPI from crude extracts of E. coli. From left to right: lanes 1-5, pBSΔ from different colonies of DF502; lane 6, pBSyTPI in DF502; lane 7, pBSΔ (2L growth, inoculated with cells from lane 1); lane 8, pBS-Ser126 (2L growth); lane 9, pBSyTPI (2L growth); lane 10, yTIM (Sigma).
REFERENCES


158
Wyckoff, H.W., Doscher, M., Tsernoglou, D., Inagami, Y., Johnson L.N.,
J. Mol. Biol. 27, 563-578.
Xia, Z.-X., Shamala, N., Bethge, P.H., Lim, L.W., Bellamy H.D., Xuong,
2629-2633.

TRANSITION-STATE ANALOGUES IN PROTEIN CRYSTALLOGRAPHY:
PROBES OF THE STRUCTURAL SOURCE OF ENZYME CATALYSIS

Elias Lolis and Gregory A. Petsko
Department of Chemistry
Massachusetts Institute of Technology
Cambridge, MA 02139

Shortened Title: TRANSITION-STATE ANALOGUES
Send Proofs to: Gregory A. Petsko, Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139 (Telephone 617-253-1837)
INTRODUCTION

Although enzymes have been studied extensively since their discovery over 150 years ago (1), it is arguable that not a single enzyme mechanism is fully understood in all its detail. The source of the tremendous rate acceleration and specificity of enzymes is gaining increasing attention due to their industrial and medical applications. The ability to make novel enzymes with new activities using recombinant DNA technology has added to the urgency of understanding enzyme reactions.

The goal of this review is to describe what has been deduced about the source of catalytic power for a number of enzymes, using as guides the structures of the enzymes complexed to transition-state analogues (TSAs). We start with an overview of transition-state theory and its application to enzyme-catalyzed reactions. For a more detailed discussion of this topic other reviews (2-7) should be consulted. We proceed to consideration of TSAs as probes for the interactions that occur at enzyme active sites. Finally, we describe the interactions between TSAs and specific enzymes, and consider the implications for the mechanism of catalysis.

PERSPECTIVES AND SUMMARY

An enzyme reaction mechanism can be studied and understood on at least three different levels. In the kinetic mechanism the order of binding of substrates and dissociation of products, and kinetic parameters such as $k_{cat}$ and $K_m$ are determined. The chemical mechanism involves a description of the various chemical species, such as intermediates and transition states, that are formed during the course of a reaction. Free energy profiles, which describe the energetics of the chemical pathway, can be
obtained from studies of the chemical and kinetic mechanisms. In principle, a free energy profile should also describe the nuclear and electronic changes the substrate undergoes during conversion to product (the reaction coordinate), but in reality these changes are very difficult to determine. The profile says nothing about how the enzyme facilitates this transformation other than by lowering the energetic barrier to the product. The *structural mechanism* seeks to understand how the enzyme lovers this energy and is able to nudge the substrate along the pathway that leads to product. A variety of techniques have been applied to study this problem. Chemical modification and mutagenesis experiments can identify amino acids in the active site that participate in catalysis. In ideal circumstances, specific roles can be assigned to these amino acids, and combined with kinetic analyses, energies of interactions (with substrate, transition state, and product) can be determined. The precise arrangement of atoms in an active site is determined by x-ray crystallography. The structure and mechanism of some enzymes using this tool have recently been reviewed (8). Structures of enzymes complexed to catalytically relevant ligands can also reveal the interactions that are responsible for catalysis. Comparisons of various forms of an enzyme - for example, the complexed and uncomplexed structures - might reveal the dynamic components of catalysis. It is only when the kinetic, chemical, and structural components have been studied that a detailed enzyme reaction mechanism can be drawn.

**TRANSITION-STATE THEORY AND ENZYME CATALYSIS**

Transition-state theory, also known as the theory of absolute reaction rates, has long been used to account for the rates of chemical reactions
(9-10). Its formulation is based on equilibrium thermodynamics and vibrational energy relationships. It is assumed that the rate of any chemical reaction depends on the decomposition rate of the transition state, the highest energy state along the potential energy surface corresponding to the chemical entity in which bonds are being made and broken. It is assumed that in the transition state, atoms involved in bond making and bond breaking are actually in-flight, and will move forward to product if the appropriate vibration drives them in the correct direction. The decomposition rate of the transition state is simply its vibrational bond frequency, \( v \). The rate constant for the chemical reaction can therefore be written as

\[
k = \kappa v [S^\#]
\]

where \([S^\#]\) is the concentration of the transition state and \( \kappa \) is the transmission coefficient. (The transmission coefficient represents various "anomalies" that are observed during the study of reaction rates. For example, due to some feature of the potential energy surface, a substrate, having passed through the transition state, might subsequently be returned back to substrate rather than forward to product. In this review, such anomalies are ignored and \( \kappa \) is assumed to be 1.) The vibrational frequency can be defined as

\[
v = \frac{kT}{h}
\]

where \( k \) is Boltzmann's constant, \( T \) is the absolute temperature, and \( h \) is Planck's constant. Using the well known equation relating equilibrium constants and free energies (\( \Delta G^\# = -RT\ln K^\# \)), the concentration of the transition state can be written in terms of the substrate ground state concentration \([S]\):

\[
[S^\#] = [S] \exp(-\Delta G^\#/RT)
\]
This expression can now be incorporated into equation (1) to yield

\[ k = \kappa \nu [S] \exp(-\Delta G^\# / RT) = \kappa (kT/h)[S] \exp(-\Delta G^\# / RT) \]  \hspace{1cm} (4)

In 1946, Pauling suggested that an enzyme is able to accelerate a chemical reaction by having evolved an active site that is exquisitely complementary to the transition state (11). That is, although an active site must bind substrate, it is in the transition state (relative to all other chemical species in the overall transformation) that the binding interactions with the enzyme are maximized.

Pauling's conclusion was mathematically derived in 1963, when transition-state theory was applied to enzyme catalysis using a thermodynamic cycle (Scheme 1) (6). \( K_n^\# \) and \( K_e^\# \) are the equilibrium constants between ground state and transition state in the uncatalyzed and catalyzed reaction, respectively. \( K_s \) defines the dissociation constant for substrate from the enzyme, and \( K_T \) defines the dissociation constant for transition state from the enzyme.

Using the subscripts e and n for the enzymatic and non-enzymatic reactions, respectively, and equation (4), the ratio of the catalyzed and uncatalyzed rate constants is

\[
\frac{k_e}{k_n} = \frac{\kappa_e \nu_e [S] \exp(-\Delta G_e^\# / RT)}{\kappa_n \nu_n [S] \exp(-\Delta G_n^\# / RT)} = \frac{K_e^\# \kappa_e \nu_e}{K_n^\# \kappa_n \nu_n} \]  \hspace{1cm} (5)

It is generally assumed the factor \( \kappa \nu \) is very similar in both the catalyzed and uncatalyzed reactions, so these terms can be factor out. Applying the equivalence \( (K_s K_n^\# = K_e^\# K_T) \) from the thermodynamic cycle yields

\[
\frac{k_e}{k_n} = \frac{K_e^\#}{K_n^\#} = \frac{K_S}{K_T} \]  \hspace{1cm} (6)

165
This illustrates that the enzyme binds the transition state much more tightly than it binds the substrate. The relationship between the dissociation constants for the substrate and transition state from the enzyme is of the same order of magnitude as the relationship between the rates of the catalyzed and uncatalyzed reaction (7).

In order to understand the source of catalysis for any particular enzyme, it is therefore necessary to understand the source of the binding power of the enzyme for the transition state.

Implications and Paradoxes of Transition-State Theory

There are certain aspects of transition-state theory as it applies to enzyme reactions that require further discussion. For example, there appears to be a paradox between the derivation of tight-binding for the transition state and the existence of multiple transition states in an enzyme free energy profile. In a single substrate-single product reaction, for every intermediate, $I$, that occurs on the reaction coordinate, there are $I+1$ transition states. This is in addition to the transition states corresponding to substrate binding and product dissociation, which for this discussion can be disregarded since the enzyme active site can do little about their energetics.\textsuperscript{16} To which transition-state is the enzyme active site complementary?

Knowles and Albery have discussed in detail the factors that increase the efficiency of an enzyme (12), and this bears on the issue of transition-state complementarity. These factors are discussed in terms of how enzyme groups interact with the various chemical species

\textsuperscript{16} This comment may not apply to all enzymes. For example, there is evidence that copper- and zinc-containing superoxide dismutase may use an electrostatic potential to enhance the rate of productive binding of superoxide (120).
along the reaction coordinate. Residues involved in *uniform binding* interact equally strongly with all species (substrate, transition states, intermediates, and products) in the reaction coordinate. The functional groups involved in *differential binding* discriminate in the binding between substrate and product. Those residues that are involved in *catalysis of an individual step* preferentially bind one of the transition states in the reaction. No catalytic advantage is gained from a mutation that preferentially binds the transition-state of a fast chemical step. An advantage is only gained when the barrier of the rate-limiting step is lowered. Evolutionary pressure has been exerted, and continues to be exerted, to lower the highest transition state, although mutations that lower the energy of a non-rate-determining step would not be deleterious. With that caveat in mind, the assumption can be made that an enzyme active site is probably most complementary to the highest transition state along the reaction coordinate.

The existence of multiple transition states raises questions about how the enzyme can also bind the transition states to which it is not complementary, as well as the other chemical species - substrates and intermediates - that must be bound. There is circumstantial evidence to support two explanations. For the situation where the active site is relatively rigid, it can be assumed that all chemical species along the reaction coordinate must bear a close resemblance to one another, so the enzyme has no difficulty in binding all of them. The strain applied by the active site due to its complementarity to one of the transition states forces the substrate into a conformation toward that transition state, where all favorable interactions are finally maximized. Alternatively, all protein atoms are known to contain thermal motions, and some enzymes
are known to undergo large conformational changes during catalysis (13). To what extent these thermal motions (atomic vibrations) and conformational changes are coupled to the reaction coordinate is not known, but it is likely that the flexibility allows the active site to adapt to each species along the reaction coordinate.

PROBING ENZYME-TRANSITION-STATE INTERACTIONS: THE USE OF TRANSITION-STATE ANALOGUES In 1969, Wolfenden recognized another implication of transition-state theory. He noted that although transition states have a short lifetime (about $10^{-12}$ seconds) and are therefore difficult to study directly, molecules that mimic the transition state of an enzyme-catalyzed reaction should bind tightly to the enzyme and could be used as a model of the transition state (7). Solution studies have revealed that many of these TSA inhibitors bind to the enzyme several orders of magnitude greater than substrate. A correlation has also been made between the $K_i$ values of various thermolysin TSA inhibitors differing in the side chain at the $P_{2}'$ position and the $k_{cat}/K_m$ values of their corresponding substrates. The plot of $k_{cat}/K_m$ vs. log $K_i$ shows a linear relationship with a slope indicating that the structural modification at the side chain produces the same incremental change in binding energy for the TSA and the actual transition state (14). None of the inhibitors, however, approach the expected binding of the transition-state based on equation (6). That this is the case is not surprising. It is impossible to get a perfect analogue since bonds are being made and broken in the true transition state. If the enzyme is capable of discriminating between the subtle differences of the substrate ground state and transition state, it is no wonder it can also discriminate between the actual transition state and an analogue of it.
Enzymologists are fond of using the pseudo-second-order rate constant $k_{\text{cat}}/K_m$ as a measure of transition-state stabilization: a mutant with altered $k_{\text{cat}}/K_m$ is often interpreted as having reduced or enhanced binding to the activated complex. But it is important to realize that in some cases this pseudo-second-order rate constant refers to a pseudo transition state, one that does not appear anywhere on the free energy profile. Only when the rate-determining step is of significantly higher energy than all other transition states along the reaction coordinate, does $k_{\text{cat}}/K_m$ refer to a real transition state. If the reaction has a number of transition states of approximately equal energy then the overall rate will reflect contributions from all of them. For the purposes of this review, a transition-state analogue is of interest if it resembles any of the highest free energy points in a reaction, regardless of their relative magnitudes. We are interested in the interactions that are responsible for catalysis, and any molecule that provides information about these interactions is relevant.

The mechanistic significance of the tight binding of transition-state analogues is quite evident. The design of an inhibitor is based on a guess as to the mechanism of the reaction, or on various experiments that suggest what the mechanism or transition state might be. The tight binding of an analogue provides powerful support for the mechanism on which its design is based, but it cannot be the only evidence upon which the designation TSA rests. Some tightly bound inhibitors are known to be opportunistic, taking advantage of binding interactions that have little relevance to catalysis. For example, methotrexate was thought to be a TSA of dihydrofolate reductase, binding with a $K_i$ of 58 pM. Crystallographic analysis revealed that it binds to
the enzyme in an orientation that is inverted with respect to dihydrofolate itself (15-16).

Studies with TSAs have also lead to unexpected results that revealed even more details about enzyme mechanisms. For example, the observation of slow binding kinetics for a tightly bound TSA might be indicative of a conformational change during the catalytic process (3). Although slow binding kinetics might appear paradoxical, it is quite consistent with how enzymes operate. Tight binding of a transition state by an enzyme is not a sufficient condition for catalysis to occur; substrate must first be bound, and the transition state, once reached, must decompose to products. Enzymes must therefore bind the transition state in the context of first binding the substrate (17). During binding of a substrate, both the enzyme active site and substrate must be desolvated (18-19). In some cases, the enzyme or substrate also undergoes a conformational change during the catalytic process. These two factors can decrease the binding kinetics of a TSA (or any other inhibitor) relative to the substrate. For example, the inhibitor may not be as effective as substrate at desolvating itself or the active site. Alternatively, when an enzyme that undergoes a conformational change during catalysis is presented with a ligand in a configuration that normally only develops after substrate binding, binding of the ligand may be slow. In a later section, the structural source of slow binding kinetics of a TSA is discussed.

A recent, novel application of TSAs is the generation of catalytic antibodies (20-21). Cleverly designed TSAs coupled to a hapten are used as antigens to elicit an antibody response. Many antibodies bind the TSA tightly, but some are even capable of catalyzing the reaction that
the TSA mimics. The subtle differences between the TSA and the actual transition state are once again apparent from the inability of catalytic antibodies to accelerate chemical reactions to the extent anticipated from the strong binding to the analogue. Since not all the antibodies that bind the TSAs are catalytic, it would be of great interest to understand the structural reasons for their failure. In the rush to produce new catalytic antibodies of practical significance, this question has been, in our view, somewhat neglected. Nonetheless, the ability to generate these catalytic antibodies is another vindication of transition-state theory, and of the utility of TSAs to understand enzyme catalysis.

The importance of x-ray structures of TSAs bound to enzymes is clear from the above discussion: such interactions are the basis of catalysis. When combined with structures of the uncomplexed enzyme and of the enzyme complexed to substrates, intermediates, and products (or their analogues), a complete picture of the mechanism of action, including any conformational changes, should emerge. In addition, as with the case of methotrexate, wrong assumptions can be corrected.

From a more practical point of view, the structures of TSAs complexed to clinically important enzymes or their homologues can help assess the physical forces that are present at active sites, and can be used to generate more potent inhibitors (22). In fact, two clinically useful inhibitors of angiotensin converting enzyme, captopril (23) and enalapril (24), were developed partly from crystallographic studies on the homologous enzyme, carboxypeptidase A.

Despite the importance of studying transition state–enzyme interactions at the atomic level, there is a paucity of data on this
topic. This is due to a very serious limitation of the crystallographic method: not all proteins can be crystallized. To date, about 100 enzyme structures have been determined. Only a handful of these structures have a TSA bound to the active site. Some of these structures, and their implications for catalysis, are discussed in the following section.

THE STRUCTURAL SOURCE OF CATALYSIS FOR VARIOUS ENZYMES

In this section the assumed structural basis of catalysis for a number of enzymes is given. The mechanism of action for each enzyme is based on a large number of experiments covering many disciplines. We describe each mechanism in the context of the structural information, focusing on the active site and the interactions that are made with TSA inhibitors.

Proteases

The most studied and best understood enzymes are the proteases. These enzymes may be classified according to their active site structures. The serine proteases contain a "catalytic triad" of serine, histidine, and aspartic acid. The aspartyl proteases contain a pair of aspartic acids at the catalytic site. The thiol peptidases contain a catalytically active cysteine which plays a similar role to the serine of the serine proteases. (The thiol proteases will not be discussed here, since structures of this enzyme complexed to a TSA do not exist. However, structures of uncomplexed papain (25) and actinidin (26), and of an acylenzyme intermediate (27) suggest that the mechanism of action of these proteases is similar to that of serine proteases.) Finally, the zinc proteases contain a metal group that is absolutely essential for catalysis. Although all these enzymes catalyze the same reaction, hydrolysis of amide and ester bonds, different mechanisms are used.
THE SERINE PROTEASES: CHYMOTRYPSIN, ELASTASE, KALLIKREIN, PROTEASE A, SUBTILISIN, AND TRYPSIN Subtilisin is the first enzyme studied to high resolution (>3.0 Å) with a TSA bound in its active site (28). In that study, a covalent bond between His 64 of the catalytic triad and the peptide chloromethyl ketone inhibitor was reported. It was not until the structure was refined that the transition-state nature of the inhibitor was revealed: the ketone carbonyl carbon atom of the inhibitor also formed a covalent bond to Oγ of the catalytic Ser 221 to form a model of the tetrahedral transition state (29). Since then, a number of other serine proteases have been studied crystallographically, some with TSAs also bound to the active site (30-39). The mechanism of action of the enzymes are the same. Each enzyme, however, has its own substrate specificity which is determined by the binding subsites.

Kinetic studies of chymotrypsin have nicely demonstrated how binding energy is used to decrease the activation energy of the reaction (40). Increasing the size (or hydrophobicity) of the side chain occupying the S₂′ subsite increases $k_{cat}/K_m$ not by decreasing $K_m$, but by increasing $k_{cat}$. This clearly suggests that the subsites not only determine the substrate specificity for the enzyme, but also interact more strongly with the transition state than with the ground state of the substrate. How this is accomplished at a site that is relatively far removed from the scissile bond is still unresolved. It will not be resolved until the differences in interactions between the substrate ground state and transition state at the S₂′ subsite can be assessed. Here we only focus on the interactions and environment around the scissile bond that have been revealed by some recent TSA-enzyme complexes (33-34).
The reaction is thought to proceed through a covalent, tetrahedral intermediate with the side chain hydroxyl of Ser 195 (Figure 1a). Many spectroscopic experiments that have purported to prove the existence and accumulation of the intermediate have subsequently been shown to be artefactual (41). Hence, the intermediate has never been shown to accumulate and probably has a structure very close to the transition state of the reaction. The strongest evidence for the tetrahedral transition state is the crystallographic analysis of a number of inhibitors bound to these enzymes. While some inhibitors do not form a covalent bond to the serine, they do bind to the active site in a distorted conformation that mimics the tetrahedral transition state (31,42). Two TSAs that do form covalent adducts are the natural inhibitor chymostatin and the boron-containing inhibitor, phenylethane boronic acid (PEBA). The structure of chymostatin bound to protease A indicates that the Oγ of Ser 195 converts the aldehyde group of the inhibitor into the tetrahedral hemiacetal (Figure 1b). Similarly, Ser 195 of chymotrypsin makes a covalent bond with PEBA to form a tetrahedral adduct (Figure 1c) (33). The binding of these inhibitors is consistent with the mechanistic proposal for this enzyme (Figure 1a). The substrate binds to the active site such that the scissile carbonyl oxygen forms hydrogen bonds to the amide nitrogens of Gly 193 and Ser 195. Binding of substrate also desolvates the active site. The side chain of Asp 102 is thought to raise the pKₐ of His 57, while maintaining its catalytic conformation (43). Although Asp 102 was once believed to be involved in a double proton transfer mechanism (122), it is now believed the charged side chain stabilizes the charges on His 95 and the transition state (123). Progress toward formation of the
charged tetrahedral transition state increases the strength of the interactions between the substrate oxygen and enzyme groups. Conversion of the substrate double bond to the longer, tetrahedral single bond of the transition state places the oxygen closer to the amide nitrogens of Gly 193 and Ser 195 in the "oxyanion hole." More importantly, the interactions change from uncharged-uncharged hydrogen bonds to the stronger charge-uncharged hydrogen bonds (44). Interactions at the $S_1'$ subsite also contribute to stabilization of the transition state. The side chain of the leaving group forms more extensive interactions with the subsite after substrate is converted to the tetrahedral intermediate. Electronic rearrangement of the tetrahedral intermediate leads to elimination of the leaving group and formation of the more stable acylenzyme intermediate. Due to potential overlap between the leaving group's amino moiety and the acylenzyme's carbonyl carbon, the leaving group is forced to diffuse out of the active site and is replaced by water. Deacylation occurs after nucleophilic attack of the acylenzyme intermediate by a water molecule. The unstable tetrahedral intermediate that is again formed readily collapses to the carboxylic acid product and a free Ser 195 hydroxyl.

ZINC-CONTAINING PROTEASES: CARBOXYPEPTIDASE A AND THERMOLYSIN

Crystallographic characterization of an enzyme mechanism is well-developed with the zinc-proteases, carboxypeptidase A (CPA) and thermolysin (TLN). CPA has been crystallographically characterized in its uncomplexed form (45) and complexed to slowly-hydrolyzed substrates (46), substrate analogues (47), TSAs (48-49), "reaction-coordinate analogues" (intermediate analogues) (50-53), a natural inhibitor (54), and hydrolysis products (48,54-55). Similarly, thermolysin has been
studied with a wide range of inhibitors (60-70) that have contributed to the understanding of its reaction mechanism.

The sequence and overall tertiary structures of these two zinc-containing enzymes are unrelated, yet they have elements in common in their active site and mechanism. Here the mechanism of CPA is discussed and the similarities to the TLN-catalyzed reaction are noted. The structural source of slow-binding inhibition of a transition-state analogue to TLN is also discussed.

Carboxypeptidase A (CPA) is a zinc-containing exopeptidase. It hydrolyzes C-terminal amino acids with preference for large hydrophobic side chains. The enzyme has been the subject of kinetic (56), crystallographic, and mutagenesis (57-58) experiments that have caused a great deal of controversy regarding its enzymatic mechanism (8). A large number of high resolution structures of enzyme-substrate and enzyme-inhibitor complexes, along with genetic and chemical information have been used recently to propose a new reaction mechanism (Figure 2a) (59).

A model of the Michaelis complex is provided by the structure of the substrate-analogue, (−)-2-benzyl-3-(p-methoxybenzoyl)propanoic acid, bound to the active site (47). Residues at the active site include Arg 71, Arg 127, Asn 144, Arg 145, Glu 270, Tyr 248, Zn²⁺, and a zinc-bound water molecule. The structure of this Michaelis-complex analogue reveals the scissile carbonyl to be hydrogen bonded to the guanidinium of Arg 127. The zinc-bound water molecule is also in position to be deprotonated by Glu 270 to attack at the π' orbital of the carbonyl. Comparison of this complex (and all other complexed structures) and the uncomplexed structure reveals a conformational change. It was
originally thought that Tyr 248, which moves from its position in the unbound enzyme to form a hydrogen bond to the substrate amide nitrogen, donates a proton to the amino leaving group to facilitate its departure (54). However, a mutant in which this residue is changed to a Phe retains wild-type activity with altered substrate affinity (57-58). It is, therefore, now believed that Tyr 248 is only important for substrate binding.

Crystallographic analysis of four reaction-coordinate analogues (50-53), two aldehydes and two ketones, has also provided a model for the intermediate of the CPA-catalyzed reaction. These reaction-coordinate analogues undergo a chemical step in the active site. This is in contrast to a TSA, which is a preformed structural homologue of the transition state. The CPA reaction-coordinate analogues bind to the enzyme in their hydrated forms. Although this is not surprising for the two electrophilic aldehydes, it is unexpected for the two ketones. These ketones are not very electrophilic at their carbonyl carbon, and it has been estimated that in solution less than 0.2% is hydrated (59). The enzyme promotes the binding of the gem-diolate analogues through polarization of the hydrated oxygens by Arg 127 and zinc. This is also the interaction that presumably stabilizes the oxyanion transition state during the course of the proteolytic reaction.

For the tetrahedral intermediate to collapse to products, the amino leaving group must be protonated. As mentioned previously, while Tyr 248 was once favored for this role (54), the structure of the enzyme complexed to the transition-state analogue, N-[[[(benzyloxy carbonyl)amino]methyl]hydroxyphosphinyl]-L-Phe, has implicated another residue. This preformed, tetrahedral phosphonamidate
inhibitor was studied crystallographically at both pH 8.5 (49), where it is intact at the active site, and at pH 7.5 (50), where it exists as the hydrolysis products. It is not known whether the enzyme participates in the hydrolysis of the inhibitor or whether the reaction is due only to the more acidic pH. Nevertheless, the intact inhibitor–enzyme complex suggests that the amino leaving group is protonated by Glu 270 and not Tyr 248. Comparison of the inhibitor structures at the 2 pH's reveals the changes that are thought to occur between transition-state and product (49). Both the Gly–PO₃(H)⁻ moiety and the Phe side chain in the Pᵢ' pocket have moved significantly upon product formation, possibly indicating a trajectory toward product release. Consistent with this view is the change in the position of Tyr 248 from a closed conformation to a disordered one, which would allow the product to diffuse from the active site.

Thermolysin has also been studied with a wide range of inhibitors bound to its active site (60-69). This enzyme has a MW of 35,000, and binds four calcium ions that are essential for its thermostability. Although there is no sequence or structural homology between TLN and CPA, as mentioned above, they do have similarities in the active site and mechanism of action: the zinc ion and glutamate, which have similar roles in catalysis, are structurally superimposable. The similarity ends there. The secondary functional groups, those responsible for substrate binding and alignment, occupy different positions and are represented by different amino acids. Nevertheless, structural studies and kinetic investigations have resulted in a mechanism of action for the enzyme (Figure 3a) not very much different from that of CPA. We focus on structural studies (63) of two putative TSAs: one that binds
slowly and another that binds with normal kinetics (71), and their implications for catalysis.

\[ N-[[1-[(phenylmethoxycarbonyl)amino]-2-phenylethyl]methoxyphosphinyl]-L-Leu-L-Ala-methyl ester (Z-F^PLA) \] binds to thermolysin with high affinity (0.068 nM), but a small second order rate constant (1000 M\(^{-1}\)s\(^{-1}\)). Its corresponding substrate, ZFLA, is hydrolyzed at a faster rate (6.7 \times 10^5 M\(^{-1}\)s\(^{-1}\)) than the inhibitor binds. An analogue of the inhibitor, where Gly is substituted for Phe (ZG^PLL) (the substitution of Leu for Ala can be ignored), binds to the enzyme 100-fold less tightly (9.1 nM), but with a more reasonable second order rate constant (>10^5 M\(^{-1}\)s\(^{-1}\)). The structure of the two inhibitor–enzyme complexes reveals different modes of binding (Figure 3b and 3c). The presence of the phenyl group on ZF^PLA allows it to make more extensive interactions with the enzyme. The phenyl side chain also restricts the rotation about the N-C\(_\alpha\) bond. These two factors constrain the carbonyl oxygen of the carboxbenzoxyl moiety to make a hydrogen bond with the amide nitrogen of Trp 115. For this interaction to occur, a water molecule present in the uncomplexed enzyme (70) must be displaced. In addition to these changes, ZF^PLA binds to the enzyme in the presumed transition–state form: the two phosphorus oxygens are ligands to the pentacoordinate zinc. ZG^PLL, on the other hand, binds to the enzyme in what is thought to be a non-productive mode, in which the zinc continues to have tetrahedral coordination. But more importantly, the carbonyl oxygen is not hydrogen bonded to the amide nitrogen of Trp 115. Rather, the carbonyl is rotated about 120°, and the water molecule, which is not expelled, mediates a hydrogen bond between the substrate's amide nitrogen and the amide nitrogen of Trp 115. The difference in the
binding mode is believed to be due to the absence of the phenyl side chain and the greater flexibility around the C-N bond for the ZG\textsubscript{PL}LL inhibitor. The difference in the kinetics of binding of the two inhibitors has been attributed to the water molecule which is displaced in one case and remains bound in the other. For steric reasons, the ZP\textsubscript{LA} inhibitor cannot bind to the active site with the water molecule present. The presence of a bulky tetrahedral phosphonamide group, together with a carbonyl group of restricted motion, occludes the active site and prevents the solvent molecule from escaping. Only during the rare occasion when the water molecule is already displaced can the inhibitor bind. The smaller ZG\textsubscript{PL}LL inhibitor has no such problem. It can bind "comfortably" with the water molecule trapped between itself and the enzyme. There is circumstantial evidence to suggest that for catalysis to occur, the water molecule must also be displaced by the substrate (62–63). This is, to our knowledge, the first indication that displacement of solvent molecules can be a source of the slow binding phenomenon, as first suggested by Rich (18).

**ASPARTIC PROTEINASES** Another set of proteases not believed to form covalent, tetrahedral intermediates during catalysis are the apartic proteinases, which function optimally at low pH. This class of enzymes has two similarities to the zinc-proteases. Like CPA, the aspartic proteinases include clinically important enzymes (22). And they also function by bringing the two reactants, the peptide and a water molecule, in close proximity. Nucleophilic attack by the water molecule on the carbonyl carbon to form the tetrahedral intermediate is thought to be assisted by general base/general acid catalysis by two aspartyl groups. A study of the complex between rhizopuspepsin and a reduced
peptide inhibitor, D-His-Pro-Phe-His-PheY(CH₂-NH)Phe-Val-Tyr, where the oxygens in the transition state are replaced by protons, illustrates these points (Figure 4) (72).

The catalytic site of the uncomplexed enzyme (73) is predominantly defined by Asp 218, a protonated Asp 35, and a water molecule that hydrogen bonds between them. Analysis of the reduced inhibitor-enzyme complex suggests that when a substrate binds to the active site, the scissile bond is distorted by the enzyme, causing it to depart from planarity. This distortion causes the nitrogen atom to change from a trigonal configuration to a pyramidal geometry, reducing the double bond character, and therefore the strength, of the peptide bond. The distortion from planarity also allows the carbonyl oxygen to be polarized by hydrogen bonds to Oδ₁ of Asp 35 and Oγ of Ser 38. The carboxylate of Asp 218 facilitates nucleophilic attack by the bound water on the carbonyl group, by proton abstraction. Binding interactions between substrate and enzyme (specifically with Asp 35 and Ser 38) are maximized as the carbonyl is converted to the tetrahedral transition-state. Electronic rearrangement of the intermediate to products is facilitated by the transfer of a proton from Asp 218 (perhaps the same proton abstracted from the attacking water molecule, or one exchanged with bulk solvent) to the peptide nitrogen leaving group. Diffusion of the products restores the hydrated environment for another catalytic cycle.

Triosephosphate Isomerase

Triosephosphate isomerase (TIM) is a glycolytic enzyme that catalyzes the reversible interconversion of dihydroxyacetone phosphate (DHAP) and D-glyceraldehyde-3-phosphate (GAP) (Figure 5a). It is composed of two
identical subunits of MW 26,000. Although there is no cooperativity between the two active sites, and each active site is composed of residues exclusively from one subunit, only the dimer is active.

The reaction catalyzed by the enzyme is a very simple one: a proton is transferred from one carbon atom to another. Various studies have revealed that (i) an enediol(ate)\(^{17}\) is formed during the course of the reaction (74); (ii) the enzymatic reaction proceeds at or near the diffusion-controlled rate (75); and (iii) a catalytic base, Glu 165, is responsible for abstracting and transferring the proton (76-77). Abstraction of a proton from either GAP or DHAP in solution using much simpler organic bases as catalysts results in the production of methyl glyoxal and inorganic phosphate. \textit{In the non-enzymatic reaction}, phosphate elimination from the enediol(ate) proceeds more rapidly than reprotonation to complete the isomerization reaction (78). On the enzyme, the reverse is true: the isomerization is catalyzed by almost ten orders of magnitude relative to the uncatalyzed rate, while the phosphate elimination reaction is greatly disfavored.

Structural studies of TIM must address a very intriguing question: how is it that this enzyme is able to direct the chemical flux (at the diffusion-controlled rate) towards DHAP and away from methyl glyoxal? The structure of the enzyme complexed to the two inhibitors phosphoglycolate (PGA), a transition-state analogue, and phosphoglycolohydroxamate (PGH), an intermediate analogue, begins to answer this question (79-80). How these inhibitors are related to the transition state and intermediate of the reaction is illustrated in Figure 5.

\(^{17}\) It has yet to be resolved whether the intermediate is an enediol or an enediolate
ACCELERATION OF THE ISOMERIZATION REACTION Both PGA and PGH, and presumably the substrate, bind to the enzyme in the cis conformation (Figure 6). There are three components to the active site that are thought to be responsible for the isomerase activity: (i) an active site environment that excludes bulk water and any of its effects; (ii) electrophilic groups Lys 12, His 95, and the dipole moment of α-helix D₁; and (iii) the catalytic base, Glu 165.

The relative hydrophobicity of, and accessibility of bulk water to, the active site is believed to be very important for the isomerization reaction. Although individual water molecules are trapped at certain positions, most of the inhibitor (except for a small area around the phosphate) and all of the catalytic active site residues (most importantly Lys 12, His 95, and Glu 165) are absolutely inaccessible to bulk water. In the absence of electrostatic screening by bulk water (81), the strength of electrostatic interactions within the active site is magnified.

Abstraction of a proton from a carbon atom is a very difficult reaction to accomplish. In terms of acid/base chemistry, the pKₐ of this carbon acid (either GAP or DuAP) must be over 15, while that of the catalytic base (Glu 165) in the uncomplexed enzyme is 4.0 (82). For catalysis to occur efficiently, the discrepancy between these two pKₐ's must be significantly reduced in the ES complex. The structure of the enzyme complexed to the PGH inhibitor suggests how the enzyme does this (80).

On one side of the inhibitor is located the electrophilic residues Lys 12, His 95, and the macrodipole moment of helix D₁ (residues 95 to 102). These functional groups are found on the side of the inhibitor
that contains the $O_1$ and $O_2$ oxygens (for the atomic numbering scheme, see Figure 5). Polarization of the substrate carbonyl oxygen ($O_1$ in GAP and $O_2$ in DHAP) weakens the adjacent C-H bond. (Polarization of the substrate by the enzyme has been observed chemically (83) and spectroscopically (84).) The $N\varepsilon 1$ nitrogen of His 95 is within hydrogen bonding distance of both the $O_1$ and $O_2$ oxygens, allowing this one atom ($N\varepsilon 1$) to polarize both DHAP and GAP. The imidazole group of His 95 lies at the amino terminal end of helix D. The macrodipole moment of this helix could play a supporting electrophilic role in that it effectively places a partial positive charge at its amino terminal end, where the substrate $O_1$ and $O_2$ oxygens reside (85). Lys 12 is also found very near the substrate oxygens, and is thought to contribute to the positive electrostatic environment on the side of the active site that polarizes the carbonyl oxygens.

Polarization of the carbonyl oxygens makes proton abstraction by Glu 165 more facile. The $pK_a$ of Glu 165 is raised from 4.0 to 6.0 upon substrate binding (124), presumably due to desolvation of the active site. This change of $pK_a$ further decreases the difference in $pK_a$ between the chemically reacting groups (Glu 165 and the substrate C-H). The carboxylate of the side chain is placed so that it can easily transfer a proton between the $C_1$ and $C_2$ carbons.

DECELERATION OF THE PHOSPHATE ELIMINATION REACTION For phosphate elimination to occur, either a transient negative charge must form on the $O_3$ oxygen bridging the phosphorus atom to the enediol portion of the intermediate, or the oxygen must become protonated. It is probably no accident that there are no enzyme groups hydrogen bonding to this oxygen. Without the electrostatic screening effects of bulk water, the
introduction of an uncompensated negative charge in the hydrophobic active site is greatly disfavored.

The conformation of the O₃ oxygen relative to the plane of the enediol moiety also affects the course of the chemical reaction. Stereoelectronic theory (86) predicts that phosphate elimination is more prone to occur when the orbital overlap between the enediol π electrons and the sp³ electrons of O₁ is maximized. This is the situation only when O₃ is as far out of the enediol plane as is stereochemically allowed. Phosphate elimination is disfavored when the orbital overlap between these electrons is minimized. This is the situation when the O₃ oxygen lies within the enediol plane. The structure of both PGA and PGH complexed to the enzyme indicates that the O₃ oxygen adopts the in-plane conformation (Figure 6), thereby minimizing the propensity to eliminate phosphate.

Comparison of the complexed and uncomplexed (87) forms of this enzyme also reveals two conformational changes. A ten residue loop (amino acids 167-176) is initially in an "open" conformation to allow substrate access to the active site. Once substrate has bound, this loop moves over 8 Å to close over the substrate and active site, and exclude bulk water. This functional role for the loop is supported by a modelling experiment in which the open conformation of the loop replaces the closed conformation in the PGA-complexed enzyme. Calculation of water accessibility (a Connolly surface (125)) indicates that bulk solvent would be able to enter deeply into the active site of this "hybrid" TIM.

In its closed form, the loop forms extensive interactions with the terminal oxygens on the substrate phosphate. The position of the
terminal oxygens directly influence the conformation of the O₃ bridging oxygen. The loop may therefore also serve to hold the O₃ oxygen in the stereoelectronically important in-plane conformation. Genetic deletion of this loop results in a mutant enzyme that has 10,000-fold less isomerase activity, but is capable of producing large amounts of methyl glyoxal (88).

The second conformational change involves the side chain of Glu 165. In the unliganded form, the carboxylate is hydrogen bonded to the hydroxyl of Ser 96. When the active site is occupied by inhibitor, this group move 2 Å to poise itself for catalysis. Why this conformational change is essential is unclear. One possibility is that Ser 96 must anchor the carboxylate group away from the middle of the active site. Otherwise, without substrate or inhibitor in the active site, the carboxylate could form interactions with His 95 and Lys 12, and prevent substrate from binding productively.

Aspartate Transcarbamylase

Aspartate transcarbamylase is an allosteric enzyme that catalyzes the condensation of carbamyl phosphate with L-aspartate to produce N-carbamyl-L-aspartate. This highly regulated enzyme reaction commits the organism to the biosynthesis of pyrimidines. Binding of cytidine triphosphate (CTP), the end product of the pyrimidine pathway, signals that enough pyrimidines have been synthesized and decreases catalytic activity. The end product of the purine pathway, adenosine triphosphate (ATP), stimulates catalytic activity in order to maintain the balance between purines and pyrimidines. Structural studies of this enzyme have aimed at not only determining the source of the catalytic activity, but also at understanding the regulatory mechanism (89).
The enzyme is large and complex. It is composed of six catalytic polypeptide chains (MW 33,000 each) that form two trimers, and six regulatory chains (MW 17,000 each) that form three dimers. The catalytic trimers and regulatory dimers combine to form a molecule with D₃ symmetry. High resolution structures of the unliganded enzyme (90), the enzyme complexed to the negative effector CTP (91), and the enzyme complexed to N-phosphonacetyl-L-aspartate (PALA) have been determined (92-94). PALA can be considered both a substrate analogue, since it contains many of the functional groups of the two substrates, and a TSA, because the amide bond (that is formed in the product) has already been made although the phosphate analogue (the phosphono group) has not been eliminated (Figure 7) (95). A complex of carbamyl phosphate, succinate (an analogue of aspartate without the amino group) and the enzyme has also been studied crystallographically (96). Modelling of an amino group to the carbamyl phosphate-succinate-enzyme x-ray structure (at the computer graphics level) produces a picture of the catalytic step: the attack of the amino group on the carbonyl carbon. Based on these structural (and other biochemical) studies (see (89) and references therein), a model of the catalytic and regulatory mechanism has been proposed.

THE CATALYTIC MECHANISM The condensation reaction that occurs involves nucleophilic attack of the lone pair electrons of aspartate's amino group on the carbonyl carbon of carbamyl phosphate to form a tetrahedral intermediate (Figure 7) (94). The phosphate group is eliminated as the intermediate collapses to product. The enzyme facilitates this reaction by (i) bringing the two substrates together in the correct orientation; (ii) assisting in the removal of the proton on aspartate’s ammonium ion
to free the nitrogen electrons for nucleophilic attack; and (iii) facilitating phosphate elimination from the intermediate.

Once the reacting species are bound, it is believed aspartate's ammonium ion is deprotonated either by His 134, which is about 4 Å away in the aspartate-modelled structure, by the phosphate (of carbamyl phosphate), or by a water molecule in the active site. The negative charge that develops as the tetrahedral transition-state forms is stabilized by hydrogen bonds from Thr 55, Arg 105, and His 134. Once again the strength of the hydrogen bond is increased in the transition-state as one of the participants is converted from an uncharged to a charged atom (44). Phosphate elimination to form products is promoted by an interaction between Arg 54 and the anhydride oxygen of carbamyl phosphate. This interaction can stabilize the developing negative charge on this atom as the phosphate is eliminated. This is in contrast to the case of TIM, where the absence of such an interaction disfavors phosphate elimination.

THE ALLOSTERIC MECHANISM The catalytic mechanism just described occurs only after a large number of conformational changes have been initiated. The regulatory mechanism of aspartate transcarbamylase is a classic example of the Monod-Wyman-Changeux theory of allostery (121). The unliganded enzyme exists in the tense (T) state. The binding of carbamyl phosphate to one of the six active sites causes a local conformational change that increases the active site affinity for aspartate. This conformational change allows Arg 105 to interact with carbamyl phosphate. Once aspartate is also bound, a number of more extensive conformational changes are initiated that convert each of the other active sites into the relaxed (R) state.
The active site is formed at the interface between two catalytic chains, with amino acids from both chains contributing to the active site (97). Within each catalytic chain, there is an aspartate-binding domain and a carbamyl phosphate-binding domain. Upon binding of the two substrates, a series of conformational changes induce the two domains to move closer to each other by about 2 Å and adopt the R state. The conformational changes involve two segments of the protein. Residues 70 to 75 (the 80's loop), which are associated with the active site, have Cα movements of about 5 Å. Another segment, the loop composed of residues 225 to 245 (the 240's loop), moves about 8 Å. In the T form, the 240's loops a pair of catalytic subunits lie essentially side by side, but after the conformational change are placed on top of each other. During the conformational change, a number of electrostatic interactions are made and broken. Some of the new electrostatic interactions between loop and nonloop residues free other functional groups to interact with the substrates and facilitate catalysis. For example, in the T state Glu 50 forms an ion pair with Arg 105. The conformational change to the R state breaks this interaction and makes a new one among Glu 50, Arg 234, and Arg 167. The new position of Arg 167 (stabilized by Glu 50) allows it to also interact with the α-carboxylate of aspartate and orient the substrate for the catalytic reaction. In addition, Arg 105 (which interacted with Glu 50 in the T state) is now free to interact with the substrate phosphate group, and to also stabilize the oxanion intermediate (94).

These tertiary changes also cause changes in the quaternary structure, which force the other active sites to adopt the R state. The real source of the quaternary changes is the steric constraints the loop
must overcome as it undergoes its conformational change. The loop disrupts symmetry-related links between two catalytic chains. Since all the catalytic chains in the T state are partly held together by interactions between their respective 240's loop, disruption of one intersubunit interaction is transmitted to the other two catalytic pairs. The conversion to the R state results in a holoenzyme where the catalytic trimers have moved apart by about 12 Å and rotated relative to each other by 10° (±5° along the molecular threefold axis). The regulatory dimers have also reoriented by 15° about the molecular twofold axis in order to maintain the interactions between the catalytic and regulatory chains.

Inhibition by CTP and activation by ATP is still under active investigation. It appears that binding of these heterotropic effectors transmits conformational changes across 60 Å from the allosteric site to the nearest active site (98).

Ribonuclease

Ribonuclease (RNAse) contains 124 amino acids and has a molecular weight of about 14,000. It hydrolyzes RNA with a preference for pyrimidine bases on the 3' side of the substrate. A general acid/general base mechanism was implicated from pH dependence studies (99). An "in-line" mechanism with a cyclic phosphate intermediate (Figure 8) (101) was formulated from chemical modification, kinetic, and N.M.R. experiments as well as a crude x-ray structure (100). An alternative, "adjacent" mechanism was also formulated based on N.M.R. and kinetic experiments and structure-activity relationships of model substrates (102). The two models differ on the role of His 12. In the in-line mechanism, His 12 functions directly in catalysis acting as a base in the first step of
the reaction, and as an acid in the second. In the adjacent mechanism, His 12 functions only to keep His 119 properly positioned for its catalytic role. Structural studies of RNAse (103-105) and other experiments (106-107) have supported the in-line mechanism.

The structure of the enzyme complexed to a substrate analogue (UpcA), in which the 5'-oxygen that forms part of the link between the two nucleic acids is replaced by a methylene, revealed the precise atomic arrangements in the active site (103). A number of electrophilic groups (Lys 7, His 12, Lys 41, Lys 66, and His 119) occupy the binding cleft of RNAse. The evolutionary conserved Thr 45 is responsible for the pyrimidine specificity: its amide nitrogen and side chain hydroxyl can form hydrogen bonds with either uracil or cytosine. His 12 is primed to act as a base and remove the proton from the 2'-hydroxyl of the pyrimidine base. The deprotonated 2'-oxygen adds to the phosphate forming a 2',3'-cyclic phosphate intermediate. Lys 41 stabilizes the pentacoordinate trigonal bipyramidal transition state to this intermediate by interacting with one of the equatorial oxygens. Elimination of the leaving group is promoted by transfer of a proton from His 119 to the leaving group oxygen, which occupies an apical position on the pentacoordinate transition state. A water molecule displaces the 2' oxygen in the second half of the reaction. His 119 deprotonates this water molecule, which attacks the cyclic phosphate intermediate. Reprotonation of the 2'-oxygen by His 12 forms the 3' phosphate product.

The simplicity of this mechanism has recently been questioned by the structure of a TSA inhibitor complexed to RNAse (108), and N.M.R. studies (109). The structure of the enzyme complexed to the
pentacoordinate adduct of uridine vanadate indicates that the
interactions of the transition state with Lys 41 and His 12 are reversed
(Figure 8b). His 12 is shown to interact with one of the equatorial
oxygen atoms on the cyclic pentacoordinate moiety. Lys 41 forms a strong
hydrogen bond to the 2'-oxygen. This interaction is also found in the
N.M.R. study. While it seems improbable that Lys 41 functions as a
base, these studies suggest that the description of the proton transfer
route may need to be adjusted. Further crystallographic experiments are
in progress in order to resolve this ambiguity (110).

Ribulose-bisphosphate Carboxylase (Rubisco)
The most abundant enzyme in the world, Rubisco, catalyzes the initial
step in the Calvin cycle of carbon dioxide fixation, a process that
stores the energy trapped by photosynthesis. Some of this energy is
lost, however, because Rubisco also catalyzes the initial oxygenation
step in photorespiration, thereby limiting crop yields. Interest in the
mechanism of this enzyme centers around understanding the factors
responsible for these two reactions, and trying to improve the
efficiency at which carbon dioxide competes with oxygen for reaction
with ribulose bisphosphate.

Various biochemical studies have revealed that two molecules of
carbon dioxide are required for catalysis to proceed. One of these
molecules is an activator, forming a carbamate bound to lysine and being
stabilized by a magnesium ion. The substrate ribulose-1,5-bisphosphate
binds to the ternary complex (enzyme-CO₂-Mg²⁺) and reacts with a second
carbon dioxide. Oxygen replaces the second carbon dioxide when the
monoxygenase activity is observed. The catalytic mechanism that has
been formulated based on a large number of studies (for reviews see
references 111 and 112) involves five distinct chemical steps: enolization, carboxylation, hydration, deprotonation, and protonation. One of the most appealing features in studying this enzymatic reaction is that each of the first four steps can be monitored independently of the overall reaction. For example, the six carbon reaction intermediate, 2'carboxy-3-keto-D-arabinitol 1,5-bisphosphate, is stable enough to be made synthetically. When fed to the enzyme, its hydrolysis to product or its decarboxylation to yield the 2,3-enediol(ate) (the reaction in the reverse direction) can be monitored. This feature allows mutants unable to catalyze the overall reaction to be analyzed, and permits conclusions to be made about the functional role of individual residues (112). The structure of the unliganded enzyme (113-114) has been used as a guide to make mutants that have allowed active site residues to be assigned roles in catalysis. More recently, the structure of the enzyme complexed to the transition-state analogue, 2-carboxy-D-arabinitol 1,5-bisphosphate (CABP), has been solved to 2.8 Å (115) (Figure 9).

The structure confirms that the activating CO₂ at the end of the carbamylated Lys 191 is one of the ligands to the metal ion. Much evidence has been gathered to suggest that the metal ion plays a very important catalytic role. Catalysis has been shown to be sensitive to the nature of the metal ion (116). Mutation of one of the ligands to the metal ion (Asp 193 to Asn) results in an enzyme that cannot catalyze the primary chemical reaction (enolization), but can fragment the six carbon intermediate to form products (112). The TSA-enzyme complex suggests why this occurs. The hydroxyl on the C₂ atom of the TSA was originally the ketone of the substrate, and is one of the ligands to the
$\text{Mg}^{2+}$. It is likely that the ketone oxygen of the substrate is also in close proximity to the metal ion and is polarized by it. This polarization weakens the adjacent C-H bond, and makes abstraction of the proton by the catalytic base more facile (a situation similar to that of triosephosphate isomerase). Replacement of Asp 193 by Asn alters the affinity of the enzyme for the metal ion, and possibly the metal ion's position in the active site. Even a small change in the position of $\text{Mg}^{2+}$ might affect the enzyme's ability to polarize the carbonyl and accomplish the enolization reaction. Hydrolysis of the intermediate is not affected because it occurs on the C$_3$ atom which is unperturbed by the $\text{Mg}^{2+}$.

The enzyme provides an environment in which the strength of electrostatic interactions between $\text{Mg}^{2+}$ and other atoms are optimized. Upon binding of CABP, the metal ion is totally shielded from solvent. Electrostatic interactions also appear to be very important in binding the two phosphate groups on the TSA. CABP binds in an extended conformation leaving the two phosphorus atoms about 10 Å apart. Phosphate binding site I is composed of two lysines, main chain NH groups, and the amino terminal end of a short helix. An arginine (Arg 281) and two histidines (His 291 and His 321) make up the second phosphate binding site.

The identity of the catalytic base that causes the enolization reaction is not immediately obvious from the TSA-enzyme complex. Previous chemical modification (117) and mutagenesis (118-119) experiments implicate Lys 166, which interacts with the the P$_1$ phosphate of the TSA inhibitor. The nitrogen atom could be built outside of electron density to within 3 Å of the C$_3$ atom (115). Since CABP mimics
the transition state of the hydration step, Lys 166 could still be the catalytic base if a conformational change is invoked as the reaction proceeds. The structure also suggests other residues that are in close proximity and might function as bases: His 321, Lys 329, Ser 368, and the nitrogen of the carbamoylated Lys 191. This ambiguity illustrates the importance of using more than one TSA-enzyme complex to understand a multistep enzyme reaction. While high resolution structures of the other steps of the reaction are awaited, this example also indicates the power of TSA-enzyme complexes when combined with site-directed mutagenesis and kinetic analysis.

CONCLUSIONS

It is impossible, by definition, to observe an enzyme-transition-state complex directly by X-ray crystallography, and will be for the foreseeable future. Transition-state analogue inhibitors, which bind competitively to enzymes and have something of the stereochemistry and charge configuration of the activated complex, can give the structural biochemist some insight into the interactions between the active site residues and this transient species. Protein crystallographers have made use of these molecules to aid in unravelling the structural basis of enzyme catalytic power and specificity. In no case have they provided all of the important information about an enzymatic reaction; structural studies of weaker-binding substrate analogues and of active-site-directed mutants have been needed as well. Yet in many cases TSAs have revealed unexpected interactions and dramatic conformational changes that were not observed when substrate-like molecules were used.

This fact, as we have seen, is not surprising. Catalysis demands that the transition state be stabilized relative to the ground-state
Michaelis complex, so some extra interactions ought to form as the activated complex develops. Since the transition state is nearly always unusual in geometry and more charge-separated than the ground state, we expect that an enzyme, which is a relatively flexible organic molecule, will be able to adapt its structure to form more favorable interactions. The case studies examined in this review suggest that this is often the case, but they also indicate that many enzymes have active sites that are preformed to be optimally complementary to the transition state of the reaction they catalyze. We conclude, therefore, that a transition-state analogue should be the first choice of the crystallographer when selecting an inhibitor to bind to a crystalline enzyme. Biochemists should be encouraged to continue to develop these molecules, for they often represent the best alternative to direct observation of the productive enzyme-substrate complex. And even when the Michaelis complex can be observed, transition-state analogues, by virtue of their differences in geometry and charge configuration, are likely to give essential additional information. For this reason, we have avoided and will continue to avoid commenting on the controversy about whether certain molecules are "really" transition-state analogues. Such debates usually focus on questions of relative binding constants and other quantitative factors. Issues like these are important, but somewhat peripheral to the crystallographer in search of information about how the enzyme does its job. Even an imperfect TSA, one that might strictly be termed a substrate or a product analogue if all the data were available, can show what interactions the enzyme CAN make. We prefer an operational definition: a transition-state analogue is any competitive inhibitor that has stereochemical and/or electrostatic features similar
to those thought to occur in one of the transition-states of the reaction. Perhaps, given this relaxed view of what constitutes a useful TSA, biochemists will provide more of them. We structural enzymologists need all the help we can get.
REFERENCES

110. Rasmussen, B., Petsko, G.A. 1989. personal communication
\[ \begin{align*}
E + S & \xrightarrow{K_n} E + S^* \\
K_{S \uparrow \uparrow} & \quad K_{T \uparrow \uparrow} \\
E S & \xrightarrow{K_e} ES^*
\end{align*} \]

Scheme 1
Figure 1  
(a) The catalytic mechanism of serine proteases. (b) A stereoscopic view of (a portion of) chymostatin A complexed to Ser 195 of protease A (34). His 57 and Asp 102 (lower right) as well as two water molecules (bottom left) that are thought to be part of the oxyanion hole are also shown. (c) A stereoscopic view of phenylethane boronic acid forming a tetrahedral complex with chymotrypsin (the bond between Oγ of Ser 195 and boron is not shown) (33). The positions of His 57 (middle) and Asp 102 (right) are also shown. In both (b) and (c) the complexes are thought to be analogues of the tetrahedral transition state. Coordinates are from the Brookhaven Databank.

Figure 2  

Figure 3  
(a) The reaction catalyzed between TLN, and the interactions in the active site. The three steps of the reaction involve nucleophilic addition to the carbonyl to form a tetrahedral intermediate, protonation of the amide nitrogen, and cleavage of the peptide bond. (b) Interactions between TLN and the slow-binding inhibitor
The inhibitor carbonyl oxygen makes a hydrogen bond to the amide nitrogen of Trp 115. (c) The interactions between TLN and the inhibitor ZF^{LA}. The inhibitor carbonyl is rotated relative to that of ZF^{LA}, and the interaction between the inhibitor amide nitrogen and Trp 115 amide nitrogen is mediated by a water molecule. The absence and presence of the water molecule in the two structures of (b) and (c) is believed to be source of the difference in the binding kinetics for these two inhibitors. Reprinted with permission from Acc. Chem. Res. 1988, 21,333-340. Copyright 1988 American Chemical Society.

Figure 4
A stereoscopic view of the rhizopuspepsin active site (Asp 35 and Asp 218) complexed with the reduced peptide inhibitor D-His-Pro-Phe-His-Phe[CH$_2$-NH]Phe-Val-Tyr (for clarity, only the portion around the scissile bond is shown) (72). Superimposed on the complex is the position of the water molecule between Asp 35 and Asp 218 in the uncomplexed enzyme. This water molecule is not present in the inhibited structure, but appears to be in perfect position to be activated and attack the peptide bond carbonyl. Coordinates are from the Brookhaven Databank.

Figure 5
(a) The reaction catalyzed by TIM. Two intermediates are shown, but they are kinetically indistinguishable. Although they are shown as enediolates, the intermediate may be an uncharged enediol. The numbering of the substrate oxygen atoms follows the numbering of the
carbon atoms to which they are bonded. (b) The two inhibitors of TIM, 2-phosphoglycolate (PGA) and phosphoglycolohydroxamate (PGH), that have been studied crystallographically. PGA is believed to be an analogue of a portion of transition state 1. PGH is believed to undergo deprotonation by Glu 165 to form an analogue of the intermediate. The two inhibitors are drawn just below the chemical species of Figure 5a they are thought to mimic. (c) The phosphate elimination reaction from the enediolate intermediate to form methyl glyoxal and inorganic phosphate.

Figure 6
PGH bound to TIM, with active site residues His 95 and Lys 12 (right), and Glu 165 (upper left). This orientation shows the cis conformation of the O₁ and O₂ oxygens of TIM, and the planarity of the O₃ oxygen with the enediol moiety.

Figure 7
(a) The reaction catalyzed by Aspartate transcarbamylase and the presumed intermediate that is formed. (b) The chemical structure of the inhibitor N-phosphonacetyl-L-aspartate (PABA).

Figure 8
(a) The in-line mechanism of ribonuclease. A detailed description is given in the text. (b) A stereoscopic view of the uridine vanadate TSA complexed to the active site of RNAse (108). His 12 is shown on the right, His 119 on the lower left, Lys 41 just above His 119, and Thr 45 on the upper left. Coordinates are from the Brookhaven Databank.
Figure 9  A schematic diagram of the interactions between activated spinach Rubisco and the TSA 2-carboxy-D-arabinitol-1,5-bisphosphate (CABP). Reprinted by permission from Nature, Vol. 337, pp. 232. Copyright (c) 1989 Macmillan Magazines Limited.