

**X-RAY CRYSTALLOGRAPHIC STUDIES
OF THE R65Q MUTANT OF YEAST
PHOSPHOGLYCERATE KINASE**

Thesis by

Timothy M. McPhillips

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1997

(Submitted March 17, 1997)

c 1997

Timothy M. McPhillips

All Rights Reserved

ABSTRACT

The structure of a ternary complex of the R65Q mutant of yeast 3-phosphoglycerate kinase (PGK) with magnesium 5'-adenylylimido-diphosphate (Mg-AMP-PNP) and 3-phospho-D-glycerate (3-PG) was determined by X-ray crystallography to 2.4 Å resolution. The structure was solved by single isomorphous replacement, anomalous scattering and solvent flattening, and has been refined to an R-factor of 0.186, with rms deviations from ideal bond distance and angles of 0.009 Å and 1.82°, respectively. The structure represents the first view of a ternary complex of PGK with both substrates. The 3-PG is located in a 'basic patch' of residues on the N-terminal domain, while the Mg-AMP-PNP interacts with residues on the C-terminal domain. The overall structure is an 'open' configuration with the two substrates separated by ~11 Å across the interdomain cleft. The structure is inconsistent with the hinge-bending model of PGK function which predicts a closed configuration for the ternary complex. Based on the available structural, biochemical, and site-directed mutagenesis data, it is proposed that the basic patch primarily represents the site of anion activation and not the catalytically active binding site for 3-PG. The active site of PGK is proposed to be the combination of the nucleotide binding site on the C-domain and a possible 3-PG binding site between the domains, near the N-terminus of helix 14 and the sidechain of arginine-38. The kinetic properties and crystal structure of the R65Q mutant of yeast PGK are consistent with this proposal.

TABLE OF CONTENTS

ABSTRACT	iii
LIST OF TABLES	vii
LIST OF FIGURES	ix
LIST OF APPENDICES	xiii
CHAPTER 1. Introduction to Phosphoglycerate Kinase and the R65Q	
Mutant of the Yeast Enzyme	1
A. Biochemistry of Phosphoglycerate Kinase	2
A.1. Reaction Catalyzed and Physiological Roles	2
A.2. Kinetics and Anion Activation of Phosphoglycerate Kinase	3
A.3. Mechanism of Phosphoryl Transfer and Role of Metal Ion	5
B. Structure of Phosphoglycerate Kinase	6
B.1. Molecular Weight, Aggregation State, and Sequence	6
B.2. Tertiary Structure	7
B.3. Substrate Binding Sites	8
C. PGK Function and Regulation	9
C.1. The Hinge-Bending Hypothesis	9
C.2. Site-Directed Mutagenesis Experiments	10
C.3. Structural Studies of Yeast R65Q PGK	12
References	14
CHAPTER 2. Structure Determination of the R65Q Mutant of Yeast	
Phosphoglycerate Kinase Complexed with Adenylylimido- diphosphate and 3-Phosphoglycerate at 2.4 Å Resolution	28
A. Experimental Procedures	29
A.1. Protein Expression and Purification	29
A.2. Crystallization	30
A.3. Data Collection	30
B. Phasing Procedures	31

B.1. Phase Combination Using <i>Structure Factor Utilities</i> (SFU)	31
B.2. SIRAS Phasing Using PHASIT	34
B.3. Electron Density Maps	36
B.4. Molecular Replacement Phasing	37
B.5. Combination of MR and SIRAS Phases	39
B.6. Why Molecular Replacement Failed	40
B.7. Improvement of SIRAS Phases by Solvent Flattening	41
C. Model Building and Refinement	44
C.1. Initial Model Building	44
C.2. Crystallographic Refinement	46
C.3. Phase Combination with Partial Model Phases	47
C.4. Determination of Water Structure	49
C.5. Refinement of the 3-Phosphoglycerate Molecule	49
C.6. Treatment of Disordered Regions	50
C.7. Quality of the Final Model	50
References	51

CHAPTER 3. Structure of the R65Q Mutant of Yeast Phosphoglycerate

Kinase Complexed with Adenylylimido-diphosphate

and 3-Phosphoglycerate 99

A. Conformation of the Polypeptide Chain	100
A.1. Yeast R65Q PGK	100
A.2. Differences with the 3PGK Model	101
A.3. Comparison with Equine, Porcine, and Bacterial PGK	103
A.4. Orientation of the Domains in Different PGK Structures	104
B. Binding Mode of AMP-PNP	106
B.1. Binding site and conformation of AMP-PNP	106
B.2. Specific Interactions Between AMP-PNP and the Protein	107
C. Binding Mode of 3-PG	109
C.1. Basic Patch Residues	109
C.2. Interactions between 3-PG and the Protein	109

C.3. Comparison with the Porcine Structure	110
C.4. Alternative Orientations of 3-PG	112
D. Conclusion	113
References	115
CHAPTER 4. A New Model of PGK Function and Activation	137
A. Implications of the Yeast R65Q Structure	138
B. Significance of the Conformational States of PGK	139
B.1. Crystallographically Observed Domain Motions	139
B.2. PGK Dynamics in Solution	142
B.3. Role of Conformational States in PGK Function	144
C. The Catalytic Binding Site for 3-PG	145
C.1. Conformations of the Interdomain Region	145
C.2. The N-Terminus of Helix 14	147
D. The Basic Patch	149
D.1. An Activatory Role for the Basic Patch	149
D.2. Effect of Anion Binding on the Basic Patch Structure	150
D.3. The R65Q Basic Patch Mutation	152
E. A Closed Conformation of PGK	154
E.1. Making Sense of the Trypanosomal PGK Structure	154
E.2. PGK Dynamics and the Trypanosomal PGK	154
E.3. The Active Site of PGK and the Trypanosomal Structure	157
E.4. Anion Activation and Trypanosomal PGK	158
References	159
APPENDICES	168

LIST OF TABLES

Table 1.1 Values of equilibrium constants shown in Figure 1.2	19
Table 1.2 Kinetic parameters for ADP-forming reaction of PGK carried out in the absence and presence of activatory concentrations of sulfate ion	20
Table 1.3 Fractions of the PGK complexes shown in Figure 1.3 at equilibrium as determined by EPR, and the resulting equilibrium constants between complexes	21
Table 1.4 Crystallization conditions and unit cell parameters for structures of PGK	22
Table 1.5 Effect of site-directed mutagenesis on activity and anion-activatory behavior of yeast PGK	23
Table 2.1 Data collection statistics for yeast R65Q PGK complexed with AMP-PNP and 3-PG	53
Table 2.2 RMS and average deviations between the various phase sets and the 1QPG model phases	54
Table 2.3 Single isomorphous replacement and anomalous scattering phasing statistics for the CMNP derivative using the program PHASIT	55
Table 2.4 Characteristics of models refined during model-building and refinement stage of structure solution	56
Table 2.5 TNT refinement scheme for the first model	57
Table 2.6 X-PLOR refinement scheme for model 7	58
Table 2.7 Final refinement statistics for model 1QPG	59
Table 3.1 Constituent residues of secondary elements in yeast phosphoglycerate kinase	117
Table 3.2 Results of superimposing highly refined PGK coordinates from other species on the yeast 3PGK and R65Q models	118

Table 3.3 Rotational and translational components of coordinate transformations relating the C-domain orientations in PGK from four different crystal structures	119
Table 3.4 Hydrogen bond distances between Mg-AMP-PNP and yeast R65Q PGK	120
Table 4.1 Comparison of the rotation and translation axes optimal for domain closure with the observed rotations and translations	160
Table 4.2 Interprobe distance distribution parameters for yeast PGK mutant (C97S, Q135C, S290C) modified with 5-(2-((iodoacetyl)-aminoethyl)amino)naphthalene-1-sulfonic acid (IAEDANS) and 5-iodoacetamidofluorescein (IAF)	161

LIST OF FIGURES

Figure 1.1 Role of PGK in glycolytic pathway	24
Figure 1.2 Substrate/product binding equilibria for PGK	25
Figure 1.3 Possible mechanism of phosphoryl transfer in PGK emphasizing the role played by the divalent cation	26
Figure 1.4 Sequence alignment for PGK from yeast, horse , and <i>B. stearothermophilus</i>	27
Figure 2.1 Crystals of yeast R65Q PGK complexed with AMP-PNP and 3-PG	60
Figure 2.2 RMS and average errors in the phase sets used in solving the structure	61
Figure 2.3 RMS phase error versus resolution for a number of phase sets used in the initial structure solution	62
Figure 2.4 The $v=0$ Harker section of the difference Patterson used in locating the Hg(II) heavy atom in the CMNP derivative data set	63
Figure 2.5 a-b Region 1 of SIR-phased map	65
Figure 2.5 c-d Region 2 of SIR-phased map	66
Figure 2.5 e-f Region 1 of SIRAS-phased map	67
Figure 2.5 g-h Region 2 of SIRAS-phased map	68
Figure 2.5 i-j Region 1 of MR-phased map	69
Figure 2.5 k-l Region 2 of MR-phased map	70
Figure 2.5 m-n Regions 1 and 2 of MR-phased map with MR model	71
Figure 2.5 o-p Region 1 of SIRAS*MR-phased map	72
Figure 2.5 q-r Region 2 of SIRAS*MR-phased map	73
Figure 2.5 s-t Region 1 of FLAT-phased map	74

Figure 2.5 u-v Region 2 of FLAT-phased map	75
Figure 2.5 w-x Region 1 of Model_13-phased map	76
Figure 2.5 y-z Region 2 of Model_13-phased map	77
Figure 2.6 Comparison of the MR model (the refined molecular replacement solution based on the 3PGK coordinates) with the final 1QPG coordinates	78
Figure 2.7 Flowcharts illustrating solvent flattening procedure using PHASES	79
Figure 2.7 (a) Overview of the procedure showing how solvent mask calculation and solvent flattening operations are performed alternately to arrive at the final phase set	79
Figure 2.7 (b) Solvent mask calculation procedure	80
Figure 2.7 (c) Solvent flattening procedure used for each solvent mask	81
Figure 2.8 Superposition of C _{α} traces of C-domains in the 1QPG and 3PGK models	82
Figure 2.9 A detailed view of Figure 2.8 showing residues 230-260 of the C-domain in the 1QPG and 3PGK models	83
Figure 2.10 Flowcharts illustrating methods used in solving the structure	84
Figure 2.10 (a) How the solvent flattened map was calculated starting from the native and derivative data sets	84
Figure 2.10 (b) How the first three models were built and refined starting from the solvent flattened map and resulting in the electron density map into which model 4 was built	85
Figure 2.10 (c) How model building and refinement proceeded from the map into which model 4 was built up to the final 1QPG coordinates	86
Figure 2.11 Characteristics of the various models	87
Figure 2.11 (a) Residues included in each model	87
Figure 2.11 (b) Number of atoms included in each model	88

Figure 2.11 (c) Percentage of non-alanine, non-glycine residues for which all side-chain atoms were included	89
Figure 2.11 (d) R-factor resulting from refinement of each model	90
Figure 2.12 a-b FLAT-phased map	92
Figure 2.12 c-d Model_1*FLAT-phased map	93
Figure 2.12 e-f Model_2*FLAT-phased map	94
Figure 2.12 g-h Model_3*FLAT-phased map	95
Figure 2.12 i-j Model_13-phased map	96
Figure 2.13 Stereoview of omit-refine difference density for 3-PG bound to the basic patch on the N-domain of yeast R65Q PGK	97
Figure 2.14 Ramachandran diagram for the final refined coordinates	98
Figure 3.1 (a) Stereoview of the overall structure of yeast R65Q PGK.	
(b) Magnified view with secondary structural elements labeled	121
Figure 3.2 Ramachandran diagram for yeast coordinates 3PGK	122
Figure 3.3 Direction of the observed, average interdomain rotation axis in PGK, and comparison with the optimal rotation axis for domain closure by 'hinge-bending'	123
Figure 3.4 Stereoview of Mg-AMP-PNP bound to the nucleotide binding site on the C-domain of yeast R65Q PGK	124
Figure 3.5 Atomic naming scheme for Mg-AMP-PNP	125
Figure 3.6 Final $2F_o - F_c$ density for the triphosphate group and Mg^{2+} ion of Mg-AMP-PNP bound to the protein	126
Figure 3.7 Hydrogen bonding interactions between the adenine ring of the Mg-AMP-PNP and the protein	127
Figure 3.8 Primary interaction between the protein and the sugar moiety of the Mg-AMP-PNP	128

Figure 3.9 Interaction between the triphosphate moiety of the Mg-AMP-PNP and the N-terminus of helix 13	129
Figure 3.10 Interaction between the γ -phosphate of the Mg-AMP-PNP and the sidechain of asparagine-334	130
Figure 3.11 Stereoview of 3-PG bound to the basic patch on the N-domain of yeast R65Q PGK	131
Figure 3.12 Atomic naming scheme for 3-PG	132
Figure 3.13 Distances between the 3-PG molecule and neighboring protein atoms in the basic patch	133
Figure 3.14 Comparison of two alternative orientations of 3-PG in the basic patch with the orientation chosen for the 1QPG model	134
Figure 3.14 (a) 1QPG model	134
Figure 3.14 (b) First alternative orientation	135
Figure 3.14 (c) Second alternative orientation	136
Figure 4.1 (a) Distance between the centroids of the N- and C-domains of the transformed coordinates of yeast R65Q PGK as a function of rotation angle. (b) Distance between substrates in the transformed coordinates of yeast R65Q PGK as a function of rotation angle	162
Figure 4.2 Location of probes for fluorescent energy transfer experiments	163
Figure 4.3 Superposition of the C α traces of the yeast R65Q and porcine PGK structures showing the interdomain region	164
Figure 4.4 The result of modeling 3-PG into the porcine PGK coordinates adjacent to the N-terminus of helix 14	165
Figure 4.5 Positions of arginines 65, 121, and 168 in the yeast R65Q, horse, pig, and bacterial PGK structures	166
Figure 4.6 Shift of helix 5 between the bacterial and porcine PGK structures	167

LIST OF APPENDICES

APPENDIX 1	Command and Geometry Definition Files for Refinement of PGK Using TNT	168
APPENDIX 2	Command and Geometry Definition Files for Refinement of PGK Using X-PLOR	173

Chapter 1

Introduction to Phosphoglycerate Kinase and the R65Q Mutant of the Yeast Enzyme

Phosphoglycerate kinase (PGK) plays key roles in both glycolysis and photosynthetic carbon fixation. PGK catalyzes the interconversion of 1,3-bisphospho-D-glycerate (1,3-BPG) and ADP with 3-phospho-D-glycerate (3-PG) and ATP. The enzyme is activated by anions, including sulfate and the substrates 3-PG and ATP. Divalent cations such as Mg^{2+} are essential to the phosphoryl transfer. PGK is a monomeric protein of 45 kD organized in two domains of approximately 200 residues each. Crystal structures of the enzyme from yeast, horse, pig and bacteria show that the C-terminal domain binds nucleotide in a hydrophobic pocket, while the N-terminal domain binds 3-PG in a 'basic patch.' Since the two binding sites are separated by over 7 Å across the interdomain cleft, it has been proposed that substrate binding triggers a 'hinge-bending' motion that results in domain closure and enables turnover. However, this hypothesis predicts that the basic patch is part of a transitory catalytic site, and this prediction is contradicted by site-directed mutagenesis studies which indicate that the primary role of this site is anion regulation—not catalysis. The R65Q mutant of the yeast enzyme is representative of these mutations. It has a significant level of catalytic activity that is not enhanced in the presence of anions.

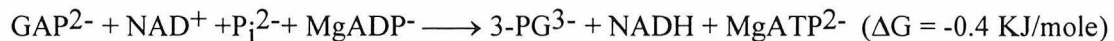
A. Biochemistry of Phosphoglycerate Kinase

A.1. Reaction Catalyzed and Physiological Roles

Phosphoglycerate kinase catalyzes a key phosphorylation step in glycolysis. Figure 1.1 summarizes the glycolytic pathway and shows the role played by PGK. The enzyme produces the first ATP molecules generated by this pathway. In this role, PGK facilitates the transfer of the anhydride phosphate of 1,3-bisphospho-D-glycerate (1,3-BPG) to MgADP, yielding 3-phospho-D-glycerate (3-PG) and MgATP (Scopes, 1978a). Equilibrium at standard conditions strongly favors the formation of ATP ($K_{eq} = 2.0 \times 10^3$ at 25°C) because the products have a much lower free energy ($\Delta G^\circ = -18.8$ kJ/mole). However, physiological concentrations of the substrates make the ATP-forming reaction unfavorable *in vitro*:

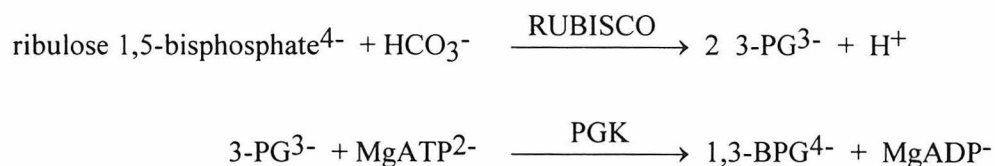


But, the reaction proceeds in this direction *in vitro* because it is tightly coupled to the previous, exergonic step in the glycolytic pathway, the reaction catalyzed by glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Since the equilibrium associated with GAPDH generates 1,3-BPG exergonically at physiological conditions ($\Delta G = -1.7$ kJ/mole), the concerted operation of GAPDH and PGK is more closely balanced:



These ΔG values are based on substrate concentrations found in rabbit muscle (Mathews & Van Holde, 1996).

In addition to generating ATP in glycolysis, PGK performs the *reverse* function in an important carbon fixing pathway associated with photosynthesis (Scopes, 1978a). This MgADP-forming reaction follows the reaction carried out by ribulose 1,5-bisphosphate carboxylase (RUBISCO),



The bicarbonate carbon is eventually incorporated into plant carbohydrates.

Most mechanistic studies have focused on the MgADP generating reaction for convenience (Scopes, 1978a). Solutions of 1,3-BPG are relatively unstable, while frozen solutions of 3-PG can be kept indefinitely. In addition, coupling PGK catalysis to the oxidation of NADH to NAD⁺ by glyceraldehyde-3-phosphate dehydrogenase (GAPDH) provides a simple spectroscopic assay for the PGK system. The assay is easier to carry out if the 1,3-BPG consumed by GAPDH is a product of the reaction being monitored, rather than a reactant. Consequently, the glycolytic reactants MgADP and 1,3-BPG are usually referred to as the products of PGK catalysis.

A.2. Kinetics and Anion Activation of Phosphoglycerate Kinase

Considerable study has been made of the ADP-forming kinetics of PGK (Scopes, 1978a). The substrates 3-PG and ATP activate the enzyme at high concentrations, yielding non-linear kinetic plots for both substrates (Scopes, 1978b). These observations have been explained in terms of two binding sites for each substrate, an

activatory and a catalytic site. Non-substrate anions such as sulfate appear to bind to both sites, activating the enzyme at low concentrations, and inhibiting it at high concentrations (Larsson-Raznikiewicz & Arvidsson, 1971; Schierbeck & Larsson-Raznikiewicz, 1979; Scopes, 1978b). Linear kinetics are observed for both substrates at sulfate concentrations greater than 20 mM (Scopes, 1978b). Apparently the activatory effects of the substrates are eliminated when non-substrate ions activate the enzyme.

The enzyme can bind and release substrates and products in any order. Figure 1.2 shows a subset of the equilibria involving the various substrates and products, and Table 1.1 lists approximate values for the equilibrium constants shown in the figure (Schmidt et al., 1995). Catalytic turnover can occur when the two substrates, 3-PG and ATP, are bound to the enzyme simultaneously. Since the dissociation constant for 1,3-BPG is an order of magnitude smaller than for ADP, 1,3-BPG is observed to accumulate on the enzyme (Schmidt et al., 1995). Consequently, release of 1,3-BPG is the rate-limiting step in the ADP-forming reaction (Scopes, 1978b).

Table 1.2 lists kinetic parameters for the ADP-forming reaction of PGK in the presence and absence of activatory concentrations of sulfate. The turnover number (k_{cat}) for the enzyme is approximately 600 s^{-1} when activated (Mas et al., 1987), and about 180 s^{-1} unactivated (Schmidt et al., 1995). The presence of the sulfate increases the K_{m} values for both substrates by a factor of approximately 2.5. However, since the k_{cat} values are also increased by a factor of about 3.5, the catalytic efficiency ($k_{\text{cat}}/K_{\text{m}}$) of the enzyme is increased overall by anion activation.

A.3. Mechanism of Phosphoryl Transfer and Role of Metal Ion

Partly due to the complex kinetic behavior of PGK, the precise mechanism of the phosphoryl transfer catalyzed by the enzyme is unknown. However, the following three things are known: (1) A phosphoenzyme intermediate is not involved (Larsson-Raznikiewicz & Schierbeck, 1974). The phosphoryl group is transferred directly between the two substrates. (2) One amino acid, arginine-38, is essential to catalysis, as demonstrated by site-directed mutagenesis (Sherman et al., 1992). The role played by this residue in catalysis is not known, however. (3) Divalent cations such as Mg^{2+} are absolutely required for turnover (Larsson-Raznikiewicz, 1964).

A recent EPR investigation of the internal equilibrium of PGK in the presence of Mn^{2+} demonstrates the catalytic importance of the divalent cation (Latwesan, D., 1992). The three equilibria expressed in Figure 1.3 summarize Latwesan's interpretation of the data. Complexes I and II represent PGK:3-PG:MnATP complexes. The distance between the two substrates is larger in complex I than in complex II. Complexes III and IV represent the analogous PGK:1,3-BPG:MnADP complexes, with the products further apart in complex IV. The EPR experiments measured the relative concentrations of complexes II, III, IV at equilibrium in solution. Complex I was not populated at measurable concentrations at equilibrium. Table 1.3 lists the observed fractions of each complex and the resulting values for the equilibrium constants K_2 and K_3 .

The equilibria K_1 and K_3 represent conformational changes in the enzyme or movement of the substrates/products between multiple binding sites, while equilibrium K_2 represents the actual phosphoryl transfer reaction. Interestingly, both products

appear to coordinate the Mn^{2+} in complex III. In the ATP-forming direction, the cation appears to facilitate phosphoryl transfer by bringing the ADP and 1,3-BPG anions together. This scheme does not indicate whether the Mn^{+2} similarly binds the two substrates in the ADP-forming reaction. However, in both reactions, the metal ion presumably stabilizes the transition states associated with nucleophilic attack on the transferred phosphorous atom.

A study of PGK activity using Rh^{+3} as the cation supports this role for the metal (Pappu et al., 1994). $\text{Rh}(\text{H}_2\text{O})_4\text{ATP}$ is an exchange inert substrate of PGK that shows single-turnover kinetics. Following phosphoryl transfer, RhADP:1,3-BPG remains bound to the enzyme, indicating that both substrates are within the coordination sphere of the metal ion during the transfer. The authors of this study concluded that the metal aligns the substrates for phosphoryl transfer and activates the electrophile by acting as an electron sink, consistent with the scheme in the EPR study described above. The authors also point out that the trapping of an intermediate analogous to complex III of Figure 1.3 on the protein indicates that a protein ligand may also be coordinating the metal at this stage, holding it onto the protein.

B. Structure of Phosphoglycerate Kinase

B.1. Molecular Weight, Aggregation State, and Sequence

Studies of PGK from many species, spanning prokaryotes to mammals, show that the enzyme consists of a single polypeptide chain of approximately 400 residues. For example, the yeast enzyme has 415 residues and a molecular weight of 45 kD

(Watson et al., 1982). The protein is monomeric under most conditions (Larsson-Raznikiewicz, 1970). Figure 1.4 shows a structure-based alignment of the sequences of PGK from the species for which crystal structures are available: yeast, horse, pig, and *B. Stearothermophilus*. Although no sequence is available for porcine PGK, the 2.2 Å crystal structure of the porcine enzyme indicates it is nearly identical to the equine sequence (Harlos et al., 1992).

The residues are numbered in Figure 1.4 according to the yeast sequence. As indicated in the figure, 166 residues, or 40% of the yeast sequence, are identical among the sequences shown. A total of 68 residues, or 16% of the yeast sequence, are identical in 23 known species (Watson & Littlechild, 1990). The distribution of conserved residues over the entire sequence indicates that the tertiary structure and function of the enzyme are conserved across all known species as well (Joao & Williams, 1993).

B.2. Tertiary Structure

Phosphoglycerate kinase has been studied crystallographically since the 1970s. Early work on the yeast (Wendell et al., 1972; Bryant et al., 1974) and equine (Blake et al., 1972) enzymes revealed a bilobal structure, with the single polypeptide chain of PGK folding into two globular domains of approximately equal size, the N-domain and C-domain. Each domain is an α/β , twisted open-sheet structure composed of a single parallel β -sheet sandwiched between α -helices. Approximately 30 of the ~400 residues make up two polypeptide linkages connecting the two domains.

Currently, crystal structures of the enzyme are available from four species, horse (Blake & Rice, 1981), pig (Harlos et al., 1992; May et al., 1996), yeast (Watson et al., 1982; McPhillips et al., 1996), and the bacterium *B. stearothermophilus* (Davies et al., 1994). Table 1.4 lists the structures reported at this time, including crystallization conditions and substrates bound to the enzyme. Note that when the current work was initiated, the only complete coordinate set available from the Protein Data Bank was the yeast 3PGK model (Watson et al., 1982). Since the completion of the R65Q yeast model reported here, coordinates for the equine, porcine, and bacterial structures have been deposited with the PDB or made available directly from the authors. A preliminary structure of PGK from a fifth species, the trypanosome *T. brucei*, has been recently announced (Bernstein, 1997).

B.3. Substrate Binding Sites

An intriguing observation made by early investigators was the large distance between the apparent substrate binding sites of PGK. Crystallographic studies of the equine and yeast enzymes demonstrated that nucleotide substrates or analogs bind in a hydrophobic pocket on the surface of the C-domain, with the phosphate groups of the nucleotides proximal to the interdomain region (Blake and Evans, 1974; Watson et al., 1982). More recently, the crystal structure of porcine PGK complexed with 3-PG shows that this substrate binds to the N-domain in a 'basic patch' consisting of arginines 21, 38, 65, 121 and 168 and histidines 62, 167, and 170 (Harlos et al., 1992). Although the site is on the surface of the N-domain facing the nucleotide binding site across the

interdomain cleft, all highly refined crystal structures of PGK agree that the peripheries of the nucleotide binding site and the basic patch are separated by at least 7 Å (Harlos et al., 1992), too far for direct phosphoryl transfer. This raises the obvious question: How does PGK work?

C. PGK Function and Regulation

C.1. The Hinge-Bending Hypothesis

The large distance between the two apparent substrate binding sites and the similarity between the domain organization of PGK and hexokinase, an enzyme known to exhibit dramatically different open and closed conformations (Bennett, Jr. & Steitz, 1978), led to the suggestion that PGK also has two, very different, open and closed conformations (Blake & Rice, 1981). A large 'hinge-bending' motion, transforming the enzyme from the open to the closed conformation, could bring the substrates of PGK into proximity before phosphate transfer (Banks et al., 1979).

This 'hinge-bending' hypothesis may be summarized as follows for the ADP-forming reaction: ATP and 3-PG bind to the C-domain and the basic patch on the N-domain, respectively. Subsequent to both substrates binding, the protein undergoes a significant conformational change, closing the gap between the domains, bringing the two substrates together, and forming a single, albeit transient, catalytic site. The positively charged sidechains contributed by the basic patch participate in the phosphoryl transfer by stabilizing the intermediates. After the reaction has occurred, the

protein again undergoes a conformational change, opening up the active site and allowing products to dissociate.

One aspect of the ‘hinge-bending’ hypothesis has limited experimental support. Small angle X-ray scattering (Pickover et al., 1979; Timchenko & Tsyuryupa, 1982), fluorescence spectroscopy (Mouawad et al., 1990; Desmadril et al., 1991), NMR (Tanswell et al., 1976; Fairbrother et al., 1990), and other experiments have confirmed that structural adjustments do occur on substrate binding. Although these spectroscopic studies have not clearly described either the nature or the magnitude of these conformational changes, the recently reported structure of *T. Brucei* PGK purports to represent a “closed” conformation (Bernstein, 1997) in which the distance between the binding sites is shorter than in previous structures.

However, the hinge-bending hypothesis does more than predict that PGK undergoes large conformational changes. Most important, it predicts that the basic patch is part of the active site of PGK. Observations of multiple conformations of the enzyme alone cannot prove the basic patch is part of the active site. Biochemical data must also agree. However, the biochemical data for PGK do not support a catalytic role for the basic patch.

C.2. Site-Directed Mutagenesis Experiments

If the basic patch is part of the active site, mutating residues that bind substrate in the basic patch will dramatically affect activity. Table 1.5 summarizes the results of all mutagenesis experiments that have been reported for PGK. These investigations show

that although mutagenesis of the basic patch histidines 62, 167, 170 and arginines 21, 65 and 168 diminishes the ability of anions to activate PGK (Fairbrother et al., 1989a, 1989b; Walker et al., 1989; Sherman et al., 1990, 1991, 1992), near-total elimination of catalytic activity is observed only for mutations of arginine-38 (Sherman et al., 1992). The failure of the other mutations to dramatically reduce activity suggests that the primary role of the basic patch cannot be catalytic. Rather, the biochemical evidence indicates that the basic patch is primarily involved in anion activation (Sherman et al., 1990).

Not only do most basic patch mutations fail to severely reduce PGK activity, but some mutations actually increase enzyme activity under certain conditions. For example, the R65Q mutant has a maximum velocity, $V_{\max}$, nearly twice that of the wild type enzyme. But, since the K_m values of the mutant for 3-PG and ATP are increased by factors of 2 to 6 relative to the wild type values, the mutant has a lower catalytic efficiency (Sherman et al., 1991). It is not surprising that the mutation reduces efficiency given that it also alters the enzyme enough to completely eliminate anion activation. Thus, the kinetic behavior of the R65Q mutant shows that arginine-65 is not essential for catalysis, but is absolutely required for anion activation, and lends support to the proposal that the basic patch primarily represents the site of anion activation and not of catalytic activity.

There are, however, two alternative explanations for the kinetic properties of the R65Q mutant that are consistent with the hinge-bending hypothesis. One is that the mutation lowers the affinity of the active site for the product 1,3-BPG. Assuming that

the product-dissociation step is rate-limiting, the overall reaction would proceed more quickly because product leaves faster. This possibility is unlikely, however, because product inhibition studies indicate that the affinity of 1,3-BPG for R65Q is the same as for the wild type (Sherman et al., 1991).

The other possible explanation is that the mutant does not assume the completely open conformation exhibited by the wild type enzyme. If the structure of the 'open' form of the mutant resembled the 'closed' form of the wild type enzyme, the conformational change required to close the mutant would be smaller, which could result in increased activity. The loss of anion activatory effects could be explained by postulating that binding anions to the wild type enzyme favors domain closure. This would be unnecessary for a partially closed mutant, so anions would not affect the activity of the mutant. Investigation of this possibility required the structure determination of the R65Q mutant of PGK.

C.3. Structural Studies of Yeast R65Q PGK

To further understand PGK function, activation, and dynamics, the following structural studies were performed:

(1) To assess the structural consequences of the R65Q substitution, the structure of the R65Q mutant of yeast PGK in the presence of 3-PG, Mg^{+2} , and an ATP-analog, AMP-PNP was determined crystallographically. Although it was assumed when the project was started that the structure of wild-type PGK from yeast was known, it was later

discovered that the available coordinates for yeast PGK had serious problems. Thus, an additional benefit of this work was the accurate determination of the structure of the yeast enzyme.

(2) To further understand the dynamics of PGK, the available structures of PGK, including the newly determined R65Q structure, were compared. The relationships between the different crystallographically observed domain orientations were determined and analyzed. Evidence for a ‘hinge-bending’ motion that could close the interdomain space and bring the basic patch and nucleotide binding pocket together was not found. Furthermore, a potential catalytic binding site for 3-PG was discovered, and an alternative mechanism of PGK activity and activation was proposed.

References

- Banks, R.D., Blake, C.C.F., Evans, P.R., Haser, R., Rice, D.W., Hardy, G.W., Merrett, M., Phillips, A.W. (1979) Sequence, structure and activity of phosphoglycerate kinase: a possible hinge-bending enzyme. *Nature* **279**:773-777.
- Bennett, Jr., W.S., Steitz, T.A. (1978) Glucose-induced conformational change in yeast hexokinase. *Proc. Natl. Acad. Sci. USA* **75**:4848-4852.
- Bernstein, B.E., Michels, P.A.M., Hol, W.G.J. (1997) Synergistic effects of substrate-induced conformational changes in phosphoglycerate kinase activation. *Nature* **385**:275-278.
- Blake, C.C.F., Evans, P.R., Scopes, R.K. (1972) Structure of horse-muscle phosphoglycerate kinase at 6 Å resolution. *Nature New Biology* **235**:195-198.
- Blake, C.C.F., Evans, P.R. (1974) Structure of horse muscle phosphoglycerate kinase. Some results on the chain conformation, substrate binding and evolution of the molecule from a 3 Å Fourier map. *J. Mol. Biol.* **84**:585-601.
- Blake, C.C.F., Rice, D.W. (1981) Phosphoglycerate kinase. *Phil. Trans. R. Soc. Lond. A* **293**: 93-104.
- Bryant, T.N., Watson, H.C., Wendell, P.L. (1974) Structure of yeast phosphoglycerate kinase. *Nature* **247**:14-17.
- Burgess, R.J., Pain, R.H. (1977) The reversible unfolding of yeast 3-phosphoglycerate kinase. *Biochem. Soc. Trans.* **5**:692-694.
- Davies, G.J., Gamblin, S.J., Littlechild, J.A., Dauter, Z., Wilson, K.S., Watson, H.C. (1994) Structure of the ADP complex of the 3-phosphoglycerate kinase from *Bacillus stearothermophilus* at 1.65 Å. *Acta Cryst.* **D50**:202-209.
- Desmadril, M., Minard, P., Ballery, N., Gaillard-Miran, S., Hall, L., Yon, J.M. (1991) Conformational changes in yeast phosphoglycerate kinase upon ligand binding: Fluorescence of a linked probe and chemical reactivity of genetically introduced cysteinyl residues. *Proteins Struct. Func. Genet.* **10**:315-324.
- Fairbrother, W.J., Graham, H.C., Williams, R.J.P. (1990) An NMR study of anion binding to yeast phosphoglycerate kinase. *Eur. J. Biochem.* **190**:161-169.

- Fairbrother, W.J., Hall, L., Littlechild, J.A., Walker, P.A., Watson, H.C., Williams, R.J.P. (1989a) Site-directed mutagenesis of histidine 62 in the 'basic patch' region of yeast phosphoglycerate kinase. *FEB Lett.* **258**:247-250.
- Fairbrother, W.J., Walker, P.A., Minard, P.A., Littlechild, J.A., Watson, H.C., Williams R.J.P. (1989b) NMR analysis of site-specific mutants of yeast phosphoglycerate kinase. An investigation of the triose-binding site. *Eur. J. Biochem.* **183**: 57-67.
- Graham, H.C., Williams, R.J.P., Littlechild, J.A., Watson, H.C. (1991) A proton-NMR study of a site-directed mutation (His388→Glu) in the interdomain region of yeast phosphoglycerate kinase. Implications for domain movement. *Eur. J. Biochem.* **196**:261-269.
- Harlos, K., Vas, M., Blake, C.F. (1992) Crystal structure of the binary complex of pig muscle phosphoglycerate kinase and its substrate 3-phospho-D-glycerate. *Proteins Struct. Func. Genet.* **12**:133-144.
- Joao, H.C., Taddei, N., Williams, R.J.P. (1992) Investigating interdomain region mutants Phe194→Leu and Phe194→Trp of yeast phosphoglycerate kinase by H-1-NMR spectroscopy. *Eur. J. Biochem.* **205**:93-104.
- Joao, H.C. & Williams, R.J.P. (1993) The anatomy of a kinase and the control of phosphate transfer. *Eur. J. Biochem.* **216**:1-18.
- Larsson-Raznikiewicz, M. (1964) Kinetic studies on the reaction catalyzed by phosphoglycerate kinase. I. The effect of Mg²⁺ and adenosine 5'-triphosphate. *Biochim. et Biophys. Acta* **85**:60-68.
- Larsson-Raznikiewicz, M. (1970) Electrophoretic purification as well as some physical and chemical characterizations of phosphoglycerate kinase from yeast. *Eur. J. Biochem.* **15**:574-580.
- Larsson-Raznikiewicz, M., Arvidsson, L. (1971) Inhibition of phosphoglycerate kinase by products and product homologues. *Eur. J. Biochem.* **22**:506-512.
- Larsson-Raznikiewicz, M., Schierbeck, B. (1974) Studies on a possible phosphoryl-enzyme intermediate in the catalytic reaction of yeast phosphoglycerate kinase. *Biochem. Biophys. Res. Comm.* **57**: 627-634.
- Latwesen, D. (1992) EPR investigation of the internal equilibrium of 3-phosphoglycerate kinase. *Ph.D. Thesis Chap. 4-5*. University of Wisconsin-Madison.

- Mas, M.T., Bailey, J.M., Resplandor, Z.E. (1988) Site-directed mutagenesis of histidine-388 in the hinge region of yeast 3-phosphoglycerate kinase: Effects on catalytic activity and activation by sulfate. *Biochemistry* **27**:1168-1172.
- Mas, M.T., Resplandor, Z.E., Riggs, A.D. (1987) Site-directed mutagenesis of glutamate-190 in the hinge region of yeast 3-phosphoglycerate kinase: Implications for the mechanism of domain movement. *Biochemistry* **26**:5369-5377.
- Mathews, C.K., Van Holde, K.E. (1996) *Biochemistry*. New York: The Benjamin/Cummings Publishing Company, Inc. 1996
- May, A., Vas, M., Harlos, K., Blake, C. (1996) 2.0 Å resolution structure of a ternary complex of pig muscle phosphoglycerate kinase containing 3-phospho-D-glycerate and the nucleotide Mn adenylylimidodiphosphate. *Proteins Struct. Func. Genet.* **24**:292-303.
- McPhillips, T.M., Hsu, B.T., Sherman, M.A., Mas, M.T., Rees, D.C. (1996) Structure of the R65Q Mutant of Yeast 3-Phosphoglycerate Kinase Complexed with Mg-AMP-PNP and 3-Phospho-D-Glycerate. *Biochemistry* **35**:4118-4127.
- Minard, P., Bowen, D.J., Hall, L., Littlechild, J.A., Watson, H.C. (1990) Site-directed mutagenesis of aspartic-acid 372 at the ATP binding-site of yeast phosphoglycerate kinase—over-expression and characterization of the mutant enzyme. *Protein Eng.* **3**:515-521.
- Mouawad, L., Desmadril, M., Perahia, D., Yon, J.M., Brochon, J.C. (1990) The effects of ligands on the conformation of phosphoglycerate kinase: fluorescence anisotropy decay and theoretical interpretation. *Biopolymers* **30**:1151-1160.
- Pappu, K.M., Gregory, J.D., Serpersu, E.H. (1994) Substrate activity of Rh(III)ATP with phosphoglycerate kinase and the role of the metal ion in catalysis. *Arch. Biochem. Biophys.* **311**:503-508.
- Pickover, C.A., McKay, D.B., Engelman, D.M., Steitz, T.A. (1979) Substrate binding closes the cleft between the domains of yeast phosphoglycerate kinase. *J. Biol. Chem.* **254**:11323-11329.
- Schierbeck, B., Larsson-Raznikiewicz, M. (1979) Product inhibition studies of yeast phosphoglycerate kinase evaluating properties of multiple substrate binding sites. *Biochim. et Biophys. Acta* **568**:195-204.
- Schmidt, P.P., Travers, F., Barman, T. (1995) Transient and equilibrium kinetic studies on yeast 3-phosphoglycerate kinase. Evidence that an intermediate containing 1,3-bisphosphoglycerate accumulates in the steady state. *Biochemistry* **34**:824-832.

- Scopes, R.K. (1978a) 3-Phosphoglycerate kinase. *Enzymes* **8**:335-351.
- Scopes, R.K. (1978b) The steady-state kinetics of yeast phosphoglycerate kinase. Anomalous kinetic plots and the effects of salts on activity. *Eur. J. Biochem.* **85**:503-516.
- Sherman, M.A., Dean, S.A., Mathiowetz, A.M., Mas, M.T. (1991) Site-directed mutations of arginine 65 at the periphery of the active site cleft of yeast 3-phosphoglycerate kinase enhance the catalytic activity and eliminate anion-dependent activation. *Protein Eng.* **4**:935-940.
- Sherman, M.A., Fairbrother, W.J., Mas, M.T. (1992) Characterization of the structure and properties of the His 62→Ala and Arg 38→Ala mutants of yeast phosphoglycerate kinase: An investigation of the catalytic and activatory sites by site-directed mutagenesis and NMR. *Protein Sci.* **1**:1-9.
- Sherman, M.A., Szpikowska, B.K., Dean, S.A., Mathiowetz, A.M., McQueen, N.L., Mas, M.T. (1990) Probing the role of arginines and histidines in the catalytic function and activation of yeast 3-phosphoglycerate kinase by site directed mutagenesis. *J. Biol. Chem.* **265**:10659-10665.
- Tanswell, P., Westhead, E.W., Williams, R.J.P. (1976) Nuclear-magnetic-resonance study of the active-site structure of yeast phosphoglycerate kinase. *Eur. J. Biochem.* **63**:249-262.
- Timchenko, A.A., Tsyuryupa, S.N. (1982) Large scale structural changes in yeast phosphoglycerate kinase on binding of substrate. *Biophysics* **27**:1065-1069.
- Walker, P.A., Joao, H.C., Littlechild, J.A., Williams, R.J.P., Watson, H.C. (1992) Characterization of yeast phosphoglycerate kinase modified by mutagenesis at residue-21. *Eur. J. Biochem.* **207**:29-37.
- Walker, P.A., Littlechild, J.A., Hall, L., Watson, H.C. (1989) Site-directed mutagenesis of yeast phosphoglycerate kinase. The 'basic patch' residue arginine 168. *Eur. J. Biochem.* **183**:49-55.
- Watson, H.C., Littlechild, J.A. (1990) Isoenzymes of phosphoglycerate kinase—evolutionary conservation of the structure of this glycolytic enzyme. *Biochem. Soc. Trans.* **18**:187-190.

Watson, H.C., Walker, N.P.C., Shaw, P.J., Bryant, T.N., Wendell, P.L., Fothergill, L.A., Perkins, R.E., Conroy, S.C., Dobson, M.J., Tuite, M.F., Kingsman, A.J., Kingsman, S.M. (1982) Sequence and structure of yeast phosphoglycerate kinase. *EMBO J.* **1**:1635-1640.

Wendell, P.L., Bryant, T.N., Watson, H.C. (1972) Low resolution structure of yeast phosphoglycerate kinase. *Nature New Biology* **240**:134-136.

Table 1.1 *Values of equilibrium constants shown in Figure 1.2. The first four values are dissociation constants for PGK and the indicated substrate/product. The fifth value represents the internal equilibrium between reactant and product bound forms of PGK. Table from Schmidt et al., 1995.*

Equilibrium Constants	Values
$K_{\text{ADP}} (\mu\text{M})^{\text{a}}$	40 ± 5
$K_{\text{PG}} (\mu\text{M})^{\text{a}}$	33 ± 3
$K_{\text{ATP}} (\mu\text{M})^{\text{b}}$	15 ± 3
$K_{\text{BPG}} (\mu\text{M})^{\text{b}}$	0.016 ± 0.004
$K_{\text{int}}^{\text{b}}$	0.11 ± 0.02

^aScopes, 1978b.

^bSchmidt et al., 1995.

Table 1.2 *Kinetic parameters for ADP-forming reaction of PGK carried out in the absence and presence of activatory concentrations of sulfate ion.*

Variable Substrate	Conditions	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($\mu\text{M}^{-1} \text{s}^{-1}$)
3-PG ^a	[ATP] = 5 mM, [Mg ²⁺] = 1 mM pH = 7.5 T = 25°C	160 ± 6	170 ± 25	0.94 ± 0.16
ATP ^a	[3-PG] = 5 mM, [Mg ²⁺] = 1 mM pH = 7.5 T = 25°C	195 ± 5	125 ± 15	1.6 ± 0.2
3-PG ^b	[ATP] = 5 mM, [Mg ²⁺] = 1 mM [SO ₄ ²⁻] = 50 mM pH = 7.5 T = 25°C	607 ± 60	500 ± 50	1.2 ± 0.2
ATP ^b	[3-PG] = 10 mM, [Mg ²⁺] = 1 mM [SO ₄ ²⁻] = 50 mM pH = 7.5 T = 25°C	598 ± 60	300 ± 630	2.0 ± 0.3

^aSchmidt et al., 1995.

^bMas et al., 1987.

Table 1.3 *Fractions of the PGK complexes shown in Figure 1.3 at equilibrium as determined by EPR, and the resulting equilibrium constants between complexes.*^a

Parameter	Value
Complex I (%)	0
Complex II (%)	40±2
Complex III (%)	20±1
Complex IV (%)	40±2
K ₂	0.5
K ₃	2.0

^aLatwesen, 1994.

Table 1.4 *Crystallization conditions and unit cell parameters for structures of PGK. Entry codes are given for structures that have been deposited with the Protein Data Bank.*

Source	Substrates Present	Precipitant Used	Resolution	Space Group	Unit Cell Parameters
yeast ^a (3PGK)	ATP (soak)	60-80% saturated (NH ₄) ₂ SO ₄	2.5 Å	C2	a = 126.6 Å b = 54.4 Å c = 93.0 Å β = 133.9°
yeast R65Q ^b (1QPG)	3-PG, Mg-AMP-PNP	11% PEG-8000	2.4 Å	C2	a = 96.0 Å b = 82.3 Å c = 70.1 Å β = 122.2°
horse ^c	none	50-60% saturated (NH ₄) ₂ SO ₄	2.5 Å	C2	a = 50.8 Å b = 106.9 Å c = 36.3 Å β = 98.6°
pig ^d	3-PG	12-14% PEG-8000	2.0 Å	P2 ₁	a = 47.8 Å b = 110.1 Å c = 36.6 Å β = 93.7°
pig ^e	3-PG, Mn-AMP-PNP (soak)	12-21% PEG-1500	2.0 Å	P2 ₁	a = 47.8 Å b = 110.1 Å c = 36.6 Å β = 93.7°
<i>B. stearo-thermo-philus</i> ^f (1PHP)	MgADP	30-40% PEG-600	1.65 Å	P2 ₁	a = 40.5 Å b = 74.0 Å c = 68.5 Å β = 99.8°
<i>T. brucei</i> (trypano-some) ^g	3-PG, MgADP	?	2.8 Å	P2 ₁ 2 ₁ 2 ₁	a = 113.2 Å b = 116.0 Å c = 171.3 Å 4 mol/asu

^aWatson et al., 1982.

^bMcPhillips et al., 1996.

^cBlake & Rice, 1981.

^dHarlos et al., 1992.

^eMay et al., 1996.

^fDavies et al., 1994.

^gBernstein et al., 1997.

Table 1.5 *Effect of site-directed mutagenesis on activity and anion-activatory behavior of yeast PGK. Relative activities are relative to a single wild-type maximum velocity and are based on the V_{max} values listed in Joao & Williams, 1993. Different mutants are likely to have been assayed under different conditions. See original references for details.*

Mutation	Anion Activation	Relative Activity (%)	Mutation	Anion Activation	Relative Activity (%)
R21K ^a	NO	50	R168M ^f	YES	20
R21M ^a	NO	13	R168K ^f	SLIGHT	40
R21A ^b	NO	27	R168Q ^b	NO	18
R38A ^b	NO	0	H170D ^f	?	100
R38Q ^b	NO	0	E190Q ^g	NO	6
H62Q ^c	YES	212	E190D ^g	NO	9
H62A ^d	NO	85	F194L ^h	YES	66
R65Q ^e	NO	165	F194W ^h	YES	50
R65S ^e	NO	163	D372N ⁱ	?	9
R65A ^e	NO	104	H388Q ^j	?	20
H167S ^b	NO	19	H388K ^k	YES	6
			H388A ^k	YES	13

^aWalker et al., 1992.

^bSherman et al., 1990.

^cFairbrother et al., 1989a.

^dSherman et al., 1992.

^eSherman et al., 1991.

^fBurgess & Pain, 1977.

^gMas et al., 1987, 1988;

^hJoao et al., 1992.

ⁱMinard et al., 1990.

^jGraham et al., 1991.

^kMas et al., 1988.

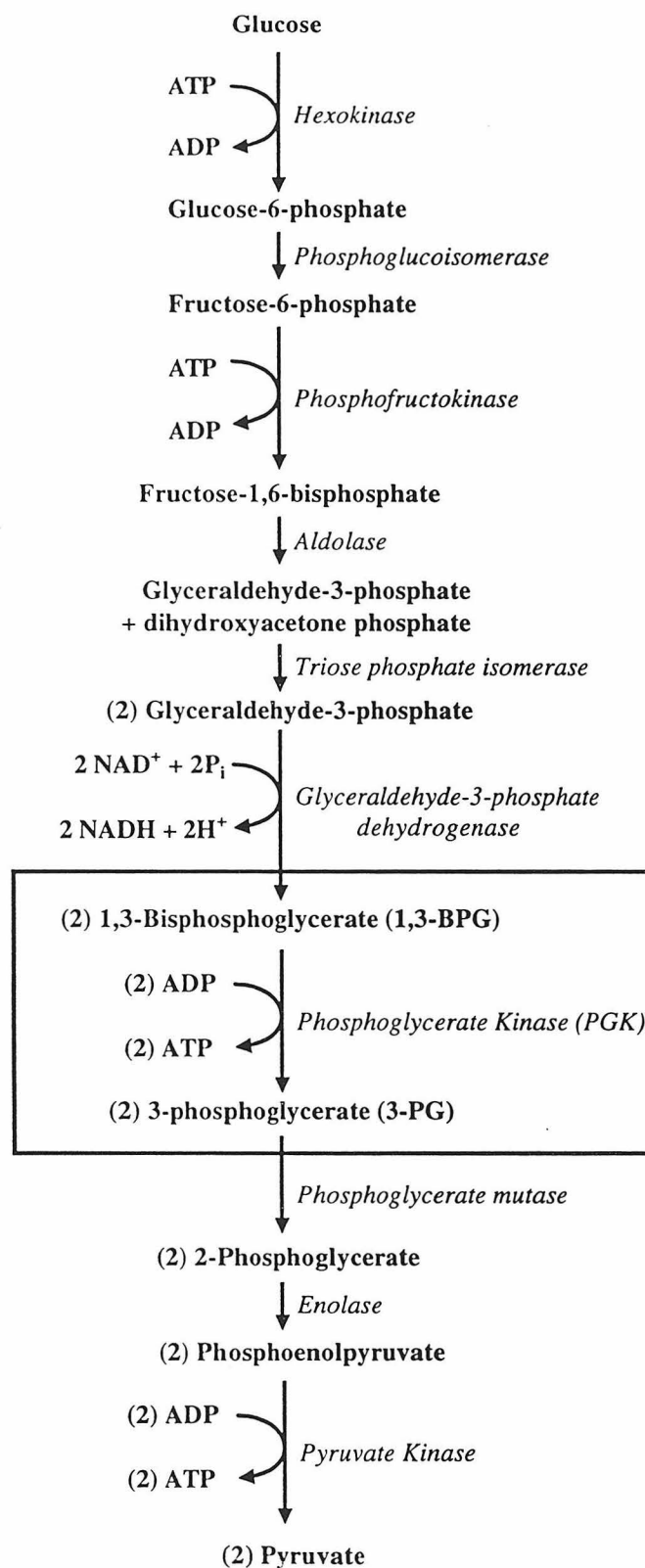


Figure 1.1 Role of PGK in glycolytic pathway. The reaction catalyzed by PGK is boxed.

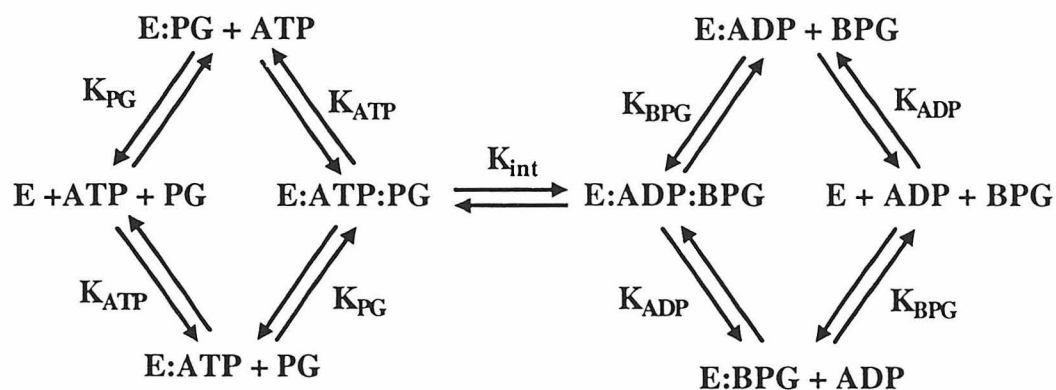


Figure 1.2 Substrate/product binding equilibria for PGK. K_{ATP} , K_{PG} , K_{ADP} , and K_{BPG} represent dissociation constants for PGK and ATP, 3-PG, ADP, and 1,3-BPG, respectively. K_{int} represents the internal equilibrium between reactant-bound and product-bound PGK. Table 1.1 lists values for each constant. From Schmidt et al., 1995.

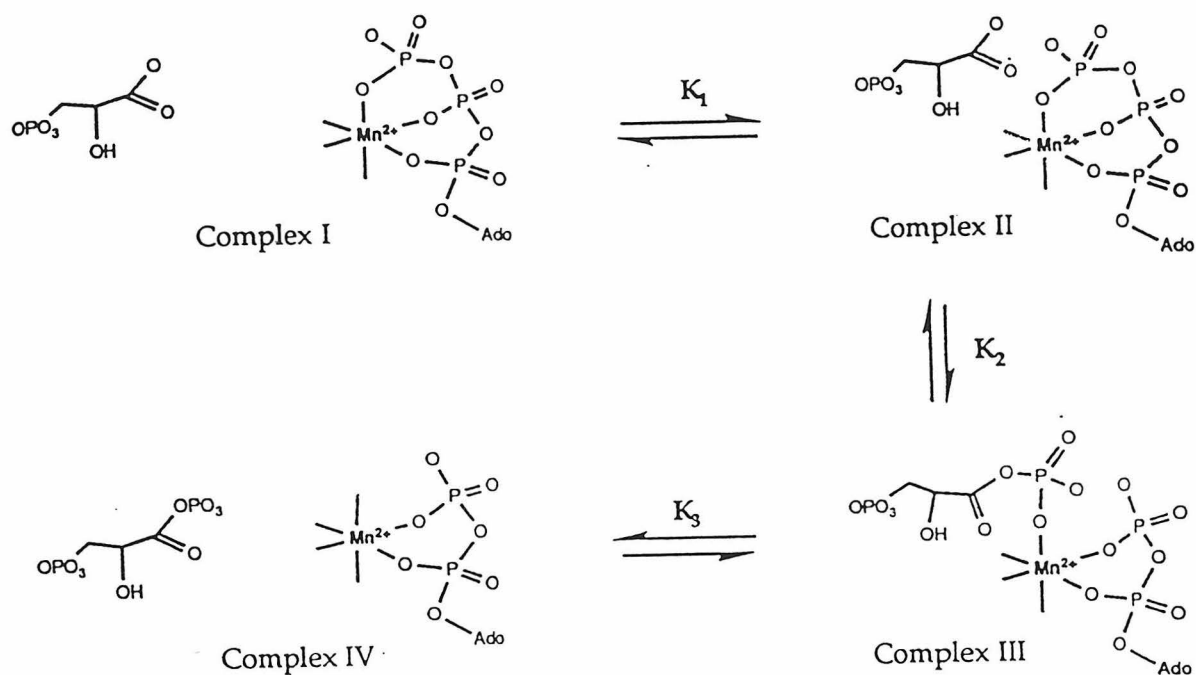


Figure 1.3 Possible mechanism of phosphoryl transfer in PGK emphasizing the role played by the divalent cation. Based on an EPR study of the internal equilibrium of PGK which reported relative concentrations of each complex shown. See Table 1.2 for relative concentrations.

1	11	21	31
YE S L S S K L S V Q D L D L K D <u>K R V F I</u> <u>R V D F N V P L D G</u> K K I T S N Q <u>R I V</u>			
HO S L S N K L T L D K L N V K G <u>K R V V M</u> <u>R V D F N V P M K N</u> N Q I T N N Q <u>R I K</u>			
BA - - M N K K T I R D V D V R G <u>K R V F C</u> <u>R V D F N V P M E Q</u> G A I T D D T <u>R I R</u>			
41	51	61	71
YE <u>A A L P</u> T I K Y V L E H H P R Y <u>V V L A</u> <u>S H L G R P</u> N G E R N - E K Y S <u>L A P V A</u>			
HO <u>A A V P</u> S I K F C L D N G A K S <u>V V L M</u> <u>S H L G R P</u> D V G P M P D K Y S <u>L Q P V A</u>			
BA <u>A A L P</u> T I R Y L I E H G A - K <u>V I L A</u> <u>S H L G R P</u> K G K V V - E E L R <u>L D A V A</u>			
81	91	101	111
YE K E <u>L Q S L L</u> G K D V T F L N D C <u>V G P</u> E V E A A V K A S A P G S V I <u>L L E N L</u>			
HO V E <u>L K S L L</u> G K D V L F L K D C <u>V G P</u> E V E K A C A D P A A G S V I <u>L L E N L</u>			
BA K R <u>L G E L L</u> E R P V A K T N E A <u>V G D</u> E V K A A V D R L N E G D V L <u>L L E N V</u>			
121	131	141	151
YE <u>R Y H I E E</u> E G S R K V - D G Q K V K A S K E - D V Q K F R H E L S S L A D V Y I N			
HO <u>R F H V E E</u> E G K G K D A S G N K V K A E P A - K I E T F R A S L S K L G D V Y V N			
BA <u>R F Y P G E</u> E K - - - - - - - - - - N D P E - L A K A F A E L A D L Y V N			
161	171	181	191
YE <u>D A F G T A H R A H</u> S S M V <u>G F - - D L P Q</u> R A A <u>G F L L E K E</u> L K Y F G K A <u>L E N</u>			
HO <u>D A F G T A H R A H</u> S S M V <u>G V - - N L P Q</u> K A G <u>G F L M K K E</u> L N Y F A K A <u>L E S</u>			
BA <u>D A F G A A H R A H</u> A S T E <u>G I A H Y L P -</u> A V A <u>G F L M E K E</u> L E V L G K A <u>L S N</u>			
201	211	221	231
YE <u>P T R P F L A I L G</u> <u>G A K V A D K I Q L</u> I D N L L D K V D S I I I G <u>G G M A F T</u>			
HO <u>P E R P F L A I L G</u> <u>G A K V A D K I Q L</u> I N N M L D K V N E M I I G <u>G G M A F T</u>			
BA <u>P D R P F T A I I G</u> <u>G A K V K D K I G V</u> I D N L L E K V D N L I I G <u>G G L A Y T</u>			
241	251	261	271
YE <u>F K K V L E N T E I</u> <u>G D S I F D K A G A</u> E I V P K L M E K A K A K G V E V V L P			
HO <u>F L K V L N N M E I</u> <u>G T S L F D E E G A</u> K I V K N L M S K A E K N G V K I T L P			
BA <u>F V K A L - G H D V</u> <u>G K S L L E E D K I</u> E L A K S F M E K A K E K G V R F Y M P			
281	291	301	311
YE <u>V D F I I A D A F S</u> A D A N T K T V T D K E G I P A G W Q G <u>L D N G P E S R K L</u>			
HO <u>V D F V T A D K F D</u> E H A K T G Q A T V A S G I P A G W M G <u>L D C G T E S S K K</u>			
BA <u>V D V V V A D R F A</u> N D A N T K V V P I D - A I P A D W S A <u>L D I G P K T R E L</u>			
321	331	341	351
YE F A A T V A K A K T I V <u>W N G P P G V F</u> <u>E F E K F A A G T K</u> A L L D E V V K S S			
HO Y A E A V A R A K Q I V <u>W N G P V G V F</u> <u>E W E A F A R G T K</u> A L M D E V V K A T			
BA Y R D V I R E S K L V V <u>W N G P M G V F</u> <u>E M D A F A H G T K</u> A I A E A L A E A -			
361	371	381	391
YE A A G N T V I I <u>G G</u> <u>G D T A T V A K K Y</u> G V T D K I S H V S <u>T G G G A S L E L L</u>			
HO S R G C I T I I <u>G G</u> <u>G D T A C C S G - W</u> N T E D K V S H V S <u>T G G G A S L E L L</u>			
BA - L D T Y S V I <u>G G</u> <u>G D S A A A V E K F</u> G L A D K M D H I S <u>T G G G A S L E F M</u>			
401	411		
YE <u>E G K E L P G V A F</u> <u>L S E K K</u>			
HO <u>E G K V L P G V D A</u> <u>L S N V</u>			
BA <u>E G K Q L P G V V A</u> <u>L E D K</u>			

Figure 1.4 Sequence alignment for PGK from yeast (YE), horse (HO), and *B. stearothermophilus* (BA). Bold letters represent residues conserved in these three species. Underlined letters represent residues conserved in 23 known species. Residues are numbered according to the yeast sequence.

Chapter 2

Structure Determination of the R65Q Mutant of Yeast

Phosphoglycerate Kinase Complexed with Adenylylimido-diphosphate and 3-Phosphoglycerate at 2.4 Å Resolution*

The structure of a ternary complex of the R65Q mutant of yeast 3-phosphoglycerate kinase (PGK) with magnesium 5'-adenylylimido-diphosphate (Mg-AMP-PNP) and 3-phospho-D-glycerate (3-PG) was determined by X-ray crystallography to 2.4 Å resolution. Crystals of the complex belong to the monoclinic space group C2 with unit cell parameters $a=96.0$ Å, $b=70.1$ Å, $c=82.3$ Å, and $\beta=122.2^\circ$. One native and one derivative data set collected from these crystals were used in solving the structure. Phasing information from single isomorphous replacement and anomalous scattering (SIRAS) effects was insufficient to solve the structure. The accuracy of the phases failed to improve after combination with model phases from molecular replacement. The search model used was the yeast 3PGK coordinates from the Protein Data Base. However, solvent flattening significantly improved the SIRAS phases, allowing an initial model containing 79% of the polypeptide chain to be built into the experimental map. Three cycles of model building, refinement and phase combination revealed the remaining portions of the chain. Ten additional rounds of model building and

* Parts of this chapter have been published in McPhillips, T.M., Hsu, B.T., Sherman, M.A., Mas, M.T.; Rees, D.C. (1996) Structure of the R65Q Mutant of Yeast 3-Phosphoglycerate Kinase Complexed with Mg-AMP-PNP and 3-Phospho-D-Glycerate. *Biochemistry* **35**:4118-4127.

refinement were required to complete the model. The final model consists of 414 protein residues, the AMP-PNP and 3-PG molecules, a single Mg^{2+} ion, and 82 water molecules. The final R-factor is 0.185 from 8.0 to 2.4 Å resolution, and the rms deviations from ideal bond distance and angles are 0.009 Å and 1.82° , respectively. The final refined coordinates have been deposited with the Protein Data Bank as entry number 1QPG.

A. Experimental Procedures

A.1. Protein Expression and Purification

The R65Q mutant of yeast 3-phosphoglycerate kinase was prepared in the laboratories of M. Mas at the City of Hope (Sherman et al., 1991). The mutant was overexpressed in the PGK⁺ *Saccharomyces cerevisiae* strain XSB44-35D (Lam and Marmur, 1977). The protein was isolated and purified according to the method described by Mas (1986). The purified protein was concentrated to approximately 20 mg/mL using Amicon Centricon-30 microconcentrators and then dialyzed against three changes of 10 mM HEPES buffer at pH 7.5 or 8.0. The concentrations of protein samples were determined spectrophotometrically using an extinction coefficient of 0.49 cm^2/mg at 280 nm (Scopes, 1978).

A.2. Crystallization

Crystals for X-ray analysis were prepared by M. Mas and M. Sherman at the City of Hope using the hanging-drop method. Linbro tissue culture plates and silanized glass coverslips were used. Well solutions consisted of 21% polyethylene glycol-8000, 20 mM magnesium acetate, 0.02% NaN_3 , and 50 mM HEPES at pH 7.5. The hanging drops were prepared by mixing well solution with an equal volume of a protein/substrate solution containing 18 mg/mL R65Q PGK; 10 mM 5'-adenylylimidodiphosphate, tetralithium salt (AMP-PNP); and 10 mM D(-) 3-phosphoglycerate, disodium salt (3-PG). The presence of both substrates was required for crystallization. Drops 10-20 μL in volume were equilibrated against well solution at 4°C. Microcrystals grown over a period of two weeks were transferred to fresh drops that had been pre-equilibrated against well-solution for two days. A reiteration of this macroseeding procedure resulted in the growth of crystals with dimensions as large as 2.0 x 0.5 x 0.3 mm. Figure 2.1 shows a typical crystal used for data collection.

A.3. Data Collection

B. Hsu in the Rees lab collected and processed all of the X-ray diffraction data for this structure determination. Data were collected at room temperature on crystals mounted in “quartz” (i.e., pure silicon dioxide) capillaries. X-rays were generated using a Siemens rotating anode with a Cu target operating at 4500 W (50kV x 90mA). A graphite monochromator was used to select the $\text{CuK}\alpha$ wavelength of 1.5418 Å. X-ray diffraction measurements were made using a Siemens X-1000 multiwire area detector.

All data were indexed, integrated, and scaled using the program XENGEN (Howard et al., 1987).

All data sets exhibited monoclinic symmetry and were characterized by unit cell parameters $a=96.0 \text{ \AA}$, $b=70.1 \text{ \AA}$, $c=82.3 \text{ \AA}$, and $\beta=122.2^\circ$. Systematic absences were observed for the condition $h + k = 2n + 1$, indicating a C-centered cell and the space group C2. Crystal density calculations indicated four 45 kD molecules per cell, or one per asymmetric unit, and a molecular volume coefficient of 2.6.

Two data sets were used in solving the structure, a native data set and a single heavy atom derivative data set. Each data set was collected on a single crystal. The derivative was a crystal soaked in 2-chloromercuri-4-nitrophenol (CMNP). Table 2.1 provides a summary of the data collection statistics for the native and derivative data sets. The statistics shown correspond to data processing runs in which Bijvoet pairs were merged. Although other heavy atom compounds were used in attempts to make additional derivatives, only the CMNP derivative was found to be useful for phasing.

B. Phasing Procedures

B.1. Phase Combination Using *Structure Factor Utilities* (SFU)

In the course of solving this structure, a program was written to manipulate structure factor files, combine phases from various sources, and compare phase sets. The advantage of this program, called SFU for Structure Factor Utilities, over other software packages available at the time was that it allowed both interactive and batch control.

Thus, multiple data sets could be read in, viewed interactively, scaled together, viewed again, and then written out. When an appropriate procedure had been worked out for solving a specific problem, a batch file could be written to perform the same functions noninteractively.

The program was used for all phase combination steps and was essential to solving this structure. Phase combination involves calculating the phase probability distribution for every reflection in each data set to be combined. The corresponding values in each distribution are multiplied together to arrive at the combined probability distributions. SFU represents phase probability distributions using Hendrickson-Lattman coefficients according to the equation,

$$P_{ABCD}(\alpha) \propto e^{A\cos\alpha+B\sin\alpha+C\cos2\alpha+D\sin2\alpha},$$

where $P_{ABCD}(\alpha)$ is the probability that the phase has the value α . Thus, the phase for each reflection in a data set is represented by four Hendrickson-Lattman coefficients, A , B , C , and D . Combining two probability distributions for a reflection is a simple matter of adding corresponding coefficients and normalizing (Bricogne, 1976).

Isomorphous replacement phasing programs typically output phase probabilities in the form of Hendrickson-Lattman coefficients, making it convenient to combine phase information from multiple runs of such programs. When phase information is derived from partial atomic models, however, these coefficients must be estimated prior to phase combination. The program SFU estimates the probability distributions for phases derived from partial models according to the method of Sim,

$$P_{sim}(\alpha) \propto e^{F_o * F_c * \cos(\alpha - \alpha_c) / \langle I_u \rangle},$$

where $P_{sim}(\alpha)$ is the probability that the phase of the reflection has the value α , α_c is the calculated model phase, F_o and F_c are the observed and calculated structure amplitudes, and $\langle I_u \rangle$ is the mean intensity contributed by the unknown part of the structure (Bricogne, 1976). SFU estimates $\langle I_u \rangle$ in resolution shells as the mean deviation between calculated and observed intensities,

$$\langle I_u \rangle = \langle |I_o - \theta * I_c| \rangle,$$

where θ is the fraction of the model contributing to I_c . SFU can take a value for θ and two reflection files containing observed and calculated structure factors, respectively, and calculate phase probability distributions for each model phase. These model phases may then be combined with other sources of phase information.

Subsequent to phase combination, the resulting phase probability distributions may be used to derive two parameters that describe the distribution (Blow & Crick, 1959). The ‘best’ phase (α_{BEST}) associated with the distribution is calculated by SFU according to the equation

$$\alpha_{BEST} = \tan^{-1} \left(\frac{\int \sin(\alpha) P_{ABCD}(\alpha) d\alpha}{\int \cos(\alpha) P_{ABCD}(\alpha) d\alpha} \right),$$

while the figure of merit (FOM) of the distribution is calculated using the equation

$$FOM = \sqrt{\left(\int \cos(\alpha) P_{ABCD}(\alpha) d\alpha \right)^2 + \left(\int \sin(\alpha) P_{ABCD}(\alpha) d\alpha \right)^2}.$$

The best phase does not necessarily occur at the maximum of the phase probability distribution. Rather, it is the centroid of the distribution. Similarly, the figure of merit

is the moment of the distribution about the best phase. It is a description of how tight the distribution is about the best phase and an indication of how much confidence should be placed in it. When the figures of merit are used to weight the associated structure factors in the Fourier calculations of electron density maps and α_{BEST} values are used for the phases, the result is an electron density map optimized for revealing the parts of a model that remain to be built. SFU can produce reflection files consisting of FOM-weighted structure amplitudes and ‘best’ phases to facilitate the calculation of such maps. All maps based on experimental or combined phases were calculated in this way during this structure determination.

In addition to performing all phase combination steps, SFU was used to compare the phases from various sources with phases calculated from the final yeast R65Q model, PDB entry 1QPG. The rms and average deviations between the various phase sets and the 1QPG model phases are summarized in Table 2.2 and graphed versus phase set in Figure 2.2. The rms deviations for some of these phase sets are graphed versus resolution in Figure 2.3. References to the rms deviations will be used in this discussion to illustrate the relative quality of the various phase sources. For comparison, it should be noted that the rms and average deviations between two completely uncorrelated phase sets are 105° and 90° , respectively.

B.2. SIRAS Phasing Using PHASIT

Difference-Patterson maps calculated using the native and CMNP derivative data sets revealed a single Hg(II) site. Figure 2.4 shows the $v=0$ Harker section of the

difference Patterson. The coordinates of the heavy atom within the x-z projection of the unit cell were estimated from this map using the relationships

$$\begin{aligned}x &= u / 2 \\z &= w / 2\end{aligned}$$

where x and z are the real-space fractional cell coordinates, and u and w are the corresponding Patterson-space fractional cell coordinates. The polar nature of the C2 space group precluded determination of a unique origin along the b-axis of the cell. The program PHASIT (Furey & Swaminathan, 1990) was used to refine the heavy-atom parameters, yielding fractional cell coordinates of $x=0.1164$ and $z=0.0411$, a site occupancy factor of 0.600, and an isotropic B-factor of $34.0 \text{ \AA}^2$. Phases were calculated to $2.75 \text{ \AA}$ resolution from both isomorphous difference and anomalous scattering effects. Table 2.3 summarizes the results of these calculations. The overall phasing power from single isomorphous replacement (SIR) effects was 2.72 for 9520 reflections. For anomalous scattering (AS), the overall phasing power was 2.47 for 8697 reflections. The overall figure of merit for the 9520 reflections phased was 0.665 at this stage.

The anomalous scattering effects of the mercury derivative were critical to this stage of phasing. The FOM was only 0.463 overall using SIR information alone, compared to 0.665 using both SIR and AS information. These statistics are consistent with the deviations between the experimental phases and the 1QPG phases. The rms deviation of the SIR phases from the 1QPG phases is 92.0° , while the corresponding value for the SIRAS phases is 76.5° . As shown in Figure 2.3, the phase error is about 15° smaller for the SIRAS phases throughout the entire resolution range of the calculations. The AS effects improved the electron density maps significantly, as shown below.

B.3. Electron Density Maps

The 26 panels of Figure 2.5 show regions of electron density calculated from the different phase sets used in solving the structure. Two different regions of electron density are shown for each phase set. The first contains a well-ordered segment of β -strand from the core of the N-domain of PGK and exemplifies the best parts of each map. The second contains the ATP analog AMP-PNP bound to the C-domain of PGK, and is representative of the less interpretable parts of the maps. Each region of density is shown twice for each map, once without any model, and once with the final 1QPG coordinates modeled in. The final four panels, Figures 2.5w-z, show $2F_o - F_c$ maps calculated from the 1QPG coordinate set. The differences between the various maps and the final $2F_o - F_c$ maps illustrate the phase deviations from the model phases listed in Table 2.2 and graphed in Figures 2.2 and 2.3.

Panels a-d of Figure 2.5 show electron density calculated using the SIR phases alone. Panels e-h show maps calculated using the SIRAS phases. The map calculated from the SIRAS phases for the N-domain view shows mainchain connectivity that is missing in the map calculated from the SIR phases. Similarly, the SIRAS map of the nucleotide shows distinct blobs of density for the adenine ring and sugar moiety, whereas the SIR map shows spurious connections to adjoining density. As discussed above, these differences between the SIR and SIRAS maps are consistent with the deviations between the final model phases and the SIR and SIRAS phases. Despite the

improvements over the SIR map, the SIRAS map contained too many ambiguities to trace the chain of PGK confidently. Thus, additional phase information was necessary.

B.4. Molecular Replacement Phasing

To improve the SIRAS phases provided from the single derivative data set, B. Hsu attempted molecular replacement using X-PLOR (Brünger, 1992). A successful molecular replacement would have provided additional phase information and improved the electron density maps prior to model building. The molecular replacement search model was the yeast PGK coordinate set 3PGK (Watson et al., 1982) from the Brookhaven protein database (Bernstein & Tasumi, 1977). All protein atoms from this file were included in the search model. Top solutions from the rotation function were refined against the Patterson-correlation (PC) function in X-PLOR. To account for possible differences in the relative orientations of the domains, the model used for PC-refinement was broken into three rigid bodies consisting of the residue ranges 1-183, 184-200, and 201-415, corresponding to the N-domain, the C-domain, and the interdomain junction, respectively. A translation search using the best results from PC-refinement gave an unambiguous solution with an R-factor of 0.528 against all data in the resolution range 7.0-4.0 Å. Rigid body refinement of this solution, using the three residue ranges used in PC-refinement above, reduced the R-factor of the solution to 0.522. Simulated annealing, positional refinement, and temperature factor refinement ultimately resulted in a model with an R-factor of 0.293 for all data in the resolution range 7.0-2.6 Å.

This molecular replacement solution appeared convincing for several reasons. First, the R-factor seemed reasonable for a system in which significant differences in interdomain orientation between the search model and the solution were expected. Second, symmetry-related molecules within the unit cell packed in a reasonable way. There were no significant collisions between molecules, nor were there large volumes of empty space in the cell. Finally, the location of the mercury heavy atom in the CMNP derivative was consistent with it being near the only cysteine in the structure. The fractional coordinates of the sulfur atom of cysteine-97 in the molecular replacement solution were (0.103, 0.057, 0.018) compared to (0.1164, y , 0.041) for the mercury atom. Had the sulfur and mercury atoms had the same y value, the distance between the two atoms would have been as low as 1.6 Å. Thus, the molecule appeared to be oriented correctly in the unit cell.

However, electron density maps calculated using the molecular replacement (MR) phases were disappointingly poor. Although much of the $2F_o - F_c$ map was interpretable and corresponded closely with the refined molecular replacement model, other parts were very difficult to interpret. Significant breaks in mainchain density were observed even in the core of the N-domain (see Figures 2.5i and 2.5j). Thus, the model appeared to require more rebuilding than an R-factor of 29% to 2.6 Å should indicate. Worse, electron density corresponding to the ATP analog AMP-PNP could not be found in either the $2F_o - F_c$ or the $F_o - F_c$ maps. Since the nucleotide coordinates had been left out of the search model, the phosphate and nucleoside groups of the AMP-PNP were expected to show up well in $F_o - F_c$ maps. The lack of such difference density meant that the

molecular replacement was probably wrong. At the very least, the phases were much poorer than the refinement statistics indicated.

B.5. Combination of MR and SIRAS Phases

Despite the apparent problems with the molecular replacement phases, an attempt was made to improve the SIRAS phases by combining them with the molecular replacement phases. Had the MR phases been roughly correct, combination with the SIRAS phases might have resolved the remaining phase ambiguities in the experimental phases arising from the lack of multiple derivatives. The combined phases did yield maps that seemed more interpretable than maps calculated using either the SIRAS or molecular replacement phases alone. Panels o-p of Figure 2.5 show the phase-combined map in the N-domain. In this section, the map is dramatically better than the map calculated from the molecular replacement phases (panels i-j), and considerably better than the SIRAS map (panels e-f). But the map still did not reveal the position of the nucleotide. Therefore, little confidence was put in the molecular replacement phases. In the end, no phase information from the molecular replacement solution was used in solving the structure.

It is interesting to note in retrospect that the molecular replacement phases were not completely wrong. Comparison of the MR phases with the 1QPG phases indicates that the MR phases were only slightly less accurate than the SIR phases, overall. The rms phase error in the MR phases was 95° , versus 92° for the SIR phases. It is also interesting to note from Figure 2.3 that the MR phases were actually somewhat more

accurate than the SIR phases at resolutions higher than 3.4 Å, but significantly less accurate at low resolution. Figure 2.3 also shows that combination of the molecular replacement phases with the SIRAS phases did not result in a significant increase in phase accuracy over the latter in any resolution bin. Over the entire resolution range, the combined phases deviated from the 1QPG phases by 77.6°, while the error in the SIRAS phases was slightly less at 76.5°. So, even though the MR phases were not completely wrong, they were not accurate enough to help solve the structure.

B.6. Why Molecular Replacement Failed

Once a correct model had been obtained, it became clear why the MR phases were relatively inaccurate. Figure 2.6 shows the relative positioning of the refined molecular replacement solution and the 1QPG model. Least-squares superposition of the molecular replacement solution on the 1QPG coordinates requires only a translation of 1.8 Å but a rotation of 8.4°. Thus, while the translation function was roughly correct, the rotation function solution was poor. More significant, the N-domain was placed more accurately than the C-domain. Superpositioning the individual domains of the molecular replacement solution on those from 1QPG requires only a 3.7° rotation and 1.7 Å translation for the N-domain, but a 14.3° rotation and 3.3 Å translation for the C-domain. Thus, the overall placement of the molecule was strongly biased toward accurately positioning the N-domain. This bias may have arisen from the greater similarity between the N-domains over that of the C-domains. Superposition of the

individual domains of the 3PGK structure on those from 1QPG results in an rms deviation in C_{α} coordinates of 2.9 Å for the N-domain, but 3.2 Å for the C-domain.

The poor placement of the C-domain in the molecular replacement solution explains why density for the nucleotide did not appear in MR-phased F_o-F_c maps. Panes k, l, q, and r of Figure 2.5 show the MR-phased $2F_o-F_c$ map and the combined SIRAS*MR phased FOM* F_o map at the nucleotide binding site. The density does not appear unusually poor in the MR phased map, and is actually quite good in the combined map. Unfortunately, the reason there is density at this site is because protein atoms in the MR model occupy the site. Figure 2.5n shows the MR model in the MR-phased map. Protein sidechains occupy the density where the nucleotide should be built. Thus, no extra density for nucleotide was observed in $2F_o-F_c$ maps and no difference density was evident in F_o-F_c maps. The misplaced protein atoms masked the presence of the AMP-PNP.

B.7. Improvement of SIRAS Phases by Solvent Flattening

Once it was clear that little confidence could be placed in the molecular replacement phases, solvent flattening of the SIRAS phases was performed using the PHASES package. The procedure substantially improved the experimental phases. Figure 2.7 is a flowchart illustrating the solvent flattening procedure used.

The overall procedure for solvent flattening using PHASES is shown in Figure 2.7a. It is a sequence of two operations performed several times in alternation: first,

determination of the solvent volume mask and, second, density modification using this mask. Figure 2.7b outlines the procedure used to calculate a solvent mask given a set of phases. First an electron density map is calculated from the phases. All negative density values within the map are then truncated to 0. The truncated map is inverted and the resulting structure factors multiplied by the transform of a real-space smearing function. The parameter RAD is used to control the radius of the smearing. Fourier transformation of the resulting structure factors yields a map in which the density at each grid point in the electron density map is equal to a weighted average of all the grid points within radius RAD of the point in the unsmeared map. This procedure is the reciprocal space equivalent of the real space, Wang algorithm for mask determination. A histogram of the electron density values within the smeared map is computed, and the estimated solvent volume fraction for the unit cell is used to determine a solvent density cutoff. A mask is then generated that discriminates the solvent and protein regions.

This mask is used in the solvent flattening operations in Figure 2.7a as outlined in Figure 2.7c. An electron density map is calculated from the current phases. From this map, the value of F_{000}/V is estimated using the equation

$$\frac{\langle \rho_{Solvent} \rangle + F_{000} / V}{\rho_{max} + F_{000} / V} = SVAL,$$

where $\langle \rho_{Solvent} \rangle$ is the average density value for all grid points in the solvent region of the map, ρ_{max} is the maximum density value in the entire map, and $SVAL$ is a constant supplied by the user. The value F_{000}/V is then used in the following equation to modify each grid point in the map. For each grid point in the solvent volume, the density values are modified according to the equation

$$\rho_{new} = \rho_{Solvent} + F_{000} / V ,$$

For the grid points in the protein volume of the map, the density values are modified according to

$$\rho_{new} = \text{Max}(\rho_{old} + F_{000} / V, 0) .$$

In other words, ρ_{new} is either $\rho_{old} + F_{000}$ or 0, whichever is greater. In this way, the solvent regions are flattened to a single electron density value, and the protein regions are truncated to positive values. The resulting map is then inverted to yield phases that are combined with the original experimental phases and a new map is calculated. The flattening procedure is repeated either four or eight times using the same mask.

Referring back to Figure 2.7a, a total of three successive solvent masks were used over the course of 16 cycles of solvent flattening. Four cycles of flattening were performed with each of the first and second masks, and eight cycles following the generation of the third mask. A solvent fraction of 30% and a sphere radius of 8 Å were used for calculating the solvent masks. A value of 0.060 was used for SVAL as recommended by the PHASES manual for 3 Å data. No phase extension was performed. The R-factor between observed structure factors and those calculated from the final solvent-flattened map was 18.5% to 2.75 Å. The final overall figure of merit for the 9520 phased reflections was 0.863, compared to 0.665 for the SIRAS phases. It should be noted that the phases resulting from this solvent flattening procedure are the combination of the phases derived from the last solvent flattened map and the SIRAS phases. From here on, the term ‘solvent flattened phases’ (and the mnemonic, FLAT)

will refer to these combined phases, while ‘solvent flattened map’ will refer to a map calculated from the combined phases.

An electron density map calculated using the solvent flattened phases was interpretable in most of the unflattened region. Panes s-v of Figure 2.5 show representative regions of the solvent-flattened map. The solvent-flattened map of the N-domain (Figure 2.5t) shows much clearer density for the mainchain and more complete density for the sidechains than does the analogous SIRAS map (Figure 2.5f). The solvent-flattened map of the C-domain (Figure 2.5v) shows much more complete density for the nucleoside moiety than does the SIRAS map (Figure 2.5h). The rms phase error in the FLAT phases was 68.1° overall, an 8.4° improvement over the SIRAS phases. Figure 2.3 shows that the solvent flattened phases are more accurate than the SIRAS phases throughout the entire resolution range.

C. Model Building and Refinement

C.1. Initial Model Building

The initial model of the R65Q mutant of PGK was built into the solvent-flattened map described above. All model building was performed with TOM/FRODO (Jones, 1985), modified by A. Chirino to run on a Silicon Graphics workstation. A rough starting model was generated by superimposing the unrefined 3PGK coordinates on the refined molecular replacement solution.

Since much of the N-domain (residues 1-185) of this starting model corresponded well with the observed electron density, building the N-domain was fairly straightforward. Whole segments of secondary structure from the superimposed 3PGK model could be fitted into observed density and then manually adjusted. However, residues 66-72 had to be built from scratch. Residues 85-151 could not be built into the solvent-flattened map at all for lack of clear density. The 15 N-terminal and 5 C-terminal residues of the PGK sequence were also left out of the initial model for the same reason.

Unlike the N-domain, nearly the entire C-domain had to be built from scratch. Even the core β -sheet structure of the 3PGK C-domain did not fit the solvent-flattened map. The β -strands of this sheet in the 1QPG model turned out to be curved very differently than in the 3PGK model (see Figure 2.8). Most of the other differences between the C-domain built into the solvent-flattened map and the 3PGK model are a result of this difference. One other dramatic difference between the 3PGK and 1QPG models is shown in Figure 2.9. The figure illustrates how helix 9 in the 1QPG model replaces helices VII and VIII of the 3PGK C-domain. Despite the significant amount of building necessary, the solvent-flattened map was clear enough to build the entire domain on the first pass.

A total of 328 residues were included in the first model, accounting for 79% of the polypeptide backbone. In addition to the three stretches of omitted backbone, the sidechains of 49 non-glycine, non-alanine residues were modeled as single C_β atoms for lack of clear sidechain density. Thus, only about 73% of the total non-hydrogen atoms

could be built into the solvent-flattened map. The PNP-AMP molecule was evident in the solvent-flattened map and was modeled in its entirety.

C.2. Crystallographic Refinement

A total of 13 cycles of model building and refinement were required to complete the structure. Figure 2.10 summarizes the course of the entire structure determination, while Table 2.4 lists the residue ranges, the total number of residues, the number of omitted sidechains, the total number of atoms, and the R-factor associated with each model. These values are graphed as a function of model number in Figure 2.11.

As shown in Figure 2.10, two different refinement packages were used to refine the structure. The first six models were refined using TNT version 5A (Tronrud et al., 1987) run under VAX/VMS on a VAXstation 4000. Appendix 1 includes example command files showing a typical refinement protocol used with TNT as well as a file defining the geometry parameters for the AMP-PNP molecule. The general procedure used was to perform multiple long cycles of positional refinement at successively higher resolution ranges, starting at a significantly lower resolution than that of the map into which the most recent model building had been performed. As an example, Table 2.5 shows in detail how the first model was refined. For the first three models, an overall temperature factor was refined while the atomic isotropic B-factors were fixed at 20.0 for all atoms. Isotropic B-factors were refined for individual atom starting with model 4.

The final seven models of the structure were refined using X-PLOR version 3.1 (Brunger, 1992) running on UNIX computers including Silicon Graphics Indigo and

Indigo2 workstations running IRIX; IBM RS/6000 POWERservers running AIX, and an AlphaStation 600 running Digital UNIX. Appendix 2 includes example X-PLOR scripts used for positional refinements, temperature factor refinements, and map calculations. Geometry definitions for the AMP-PNP and 3-PG molecules are also included. Positional and temperature factor refinements were performed alternately starting at low resolution and proceeding to higher resolution as with the TNT refinements. No simulated annealing refinements were performed in this structure refinement. Table 2.6 shows the refinement protocol carried out for the X-PLOR refinement of model 7.

C.3. Phase Combination with Partial Model Phases

Since less than 80% of the polypeptide chain could be built into the solvent-flattened map, phase combination with partial model phases was required to reveal the missing parts of the protein. For each of the first three models, TNT was used to calculate model phases from the refined coordinates. SFU was used to estimate the phase probability distribution for each model phase. The model phases were then combined with the phases resulting from solvent flattening using SFU, yielding a 'best' phase and figure of merit for each reflection. A 'best' map was then calculated with TNT using the 'best' phases and FOM-weighted F_{obs} . Model building was resumed using the phase-combined map. Three iterations of refinement, phase combination, and model building revealed the remaining residues, reducing the R-factor at 2.9 Å

resolution from 0.390 for the first refined model, to 0.251 for the fourth refined model (see Figure 2.11).

Figure 2.2 shows why this technique worked. The rms and average phase errors in the partial models are shown in this figure as open squares and circles, respectively, connected by dashed lines. The filled squares and circles below the open shapes represent the errors in the phases after combination with the solvent flattened phases. In all cases, the combined phases are at least 8° more accurate than either the solvent flattened phases or the model phases. Figure 2.3 shows that the phases resulting from combination of the model 1 and solvent flattened phases (Model_1*FLAT) are more accurate than the FLAT phases throughout the entire resolution range, and as much as 10° more accurate at high resolution.

Figure 2.12 demonstrates the effect of the improving phases on the interpretability of an unclear region of the solvent flattened map. Panes a and b of the figure show the solvent flattened map for a volume of the cell containing residues from both the peptide segment 95-103 which was omitted in models 1-2, and the segment 118-133 which was omitted in models 1-3. The electron density shows false connections between these segments and is not clear enough to distinguish mainchain from sidechain. Panels c-d, e-f, and g-h of Figure 2.12 show density calculated from combined phases based on models 1, 2, and 3, respectively. For comparison, panels i-j show the map calculated using the 1QPG model phases. The false connections fade with each successive map, revealing the true course of the polypeptide chain. Thus, phase combination of partial models improved the density of regions for which no model had yet been built.

C.4. Determination of Water Structure

Water molecules were added after the refinement of the ninth model, at which stage the R-factor from X-PLOR was 19.8 % to 2.4 Å and the rms bond and angle deviations from ideality were 0.017 Å and 2.1°, respectively. Water molecules were added only where $F_o - F_c$ positive difference peaks of at least 3.0 σ height occurred in chemically reasonable positions. No waters were added in the vicinity of zero-occupancy atoms. Occupancies for all waters were fixed at 1.0 and the temperature factors were left unrestrained during refinement. Refined water molecules were deleted from the structure if they did not appear as 1.0 σ or higher peaks in $2F_o - F_c$ maps or if they did not have refined B-factors less than 80 Å². The tenth major cycle of refinement included five cycles of adding water molecules, refining the model, deleting poor waters, and looking for additional waters. Few significant changes were made to the water structure following model ten.

C.5. Refinement of the 3-Phosphoglycerate Molecule

Electron density corresponding in shape and size to a 3-PG molecule was observed in maps calculated in the final stages of model building and refinement. The molecule was included in refinement starting with model ten. Refinement at full occupancy resulted in high B-factors for the 3-PG atoms, indicating partial occupancy of the site or disordering of the molecule. Setting the occupancy of the 3-PG atoms to 0.5 resulted in minimum, average, and maximum temperature factors for the 3-PG atoms of 50, 61, and

75 Å², respectively for the final refinement. The shape of the electron density (see Figure 2.13) indicates that a significant population of 3-PG may be in an alternative orientation with the phosphate and carboxyl groups reversed.

C.6. Treatment of Disordered Regions

Throughout the model building process, seven short segments of the chain, containing residues 29-32, 132-136, 287-297, 301-307, 340-348, 360-364, and 377-385, appeared to be either somewhat disordered or thermally mobile about the average, modeled position. Ambiguities in these sections of the structure were resolved by best fit to the available density and by use of geometric constraints imposed by distinct density in adjoining regions. In addition, negligible density was observed in the final stages of refinement for the side chains of 15 non-glycine, non-alanine residues: D29, K107, D133, K141, N200, F289, K319, K327, K329, K344, K358, S360, K378, K384 and K385. The occupancies of the side chain atoms in these residues were fixed at zero for all structure factor calculations. Occupancies of all atoms in residue 415 were treated similarly because they lacked density.

C.7. Quality of the Final Model

The final model consists of 414 protein residues, the AMP-PNP and 3-PG molecules, a single Mg²⁺ ion, and 82 water molecules. Table 2.7 summarizes the refinement statistics for the 1QPG model. The final R-factor is 0.185, calculated using all positive structure factors in the data set above 8.0 Å resolution, without application

of a sigma cutoff or bulk solvent correction. The restraint weight used in the final refinement was that calculated by the CHECK routine of X-PLOR yielding rms bond and angle deviations from ideality of 0.009 Å and 1.78°, respectively. The Ramachandran plot from PROCHECK (Laskowski et al., 1993) in Figure 2.14 shows that over 88% of the non-glycine, non-proline residues are found in the most favored regions, with no residues in disallowed regions. Analysis of side chain geometry by χ_1 - χ_2 plots does not indicate any unfavorable conformations. Overall, the average B-factor in the N-domain, 31 Å², is lower than in the C-domain, 39 Å², reflecting a clearer electron density map of the N-domain than of the C-domain.

References

- Bernstein, F.C., Tasumi, M. (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Appl. Cryst.* **26**: 282-291.
- Blow, D.M., Crick, F.H.C. (1959) The treatment of errors in the isomorphous replacement method. *Acta Cryst.* **12**:794-802.
- Bricogne, G. (1976) Methods and programs for direct-space exploitation of geometric redundancies. *Acta Cryst.* **A32**:832-847.
- Brunger, A.T. (1992) X-PLOR. Version 3.1. A system for X-ray crystallography and NMR. Yale University Press, New Haven and London.
- Furey, W., Swaminathan, S. (1990) PHASES--A program package for the processing and analysis of diffraction data from macromolecules. *American Cryst. Assoc. Meet. Abstracts Series 2* **18**:73.
- Howard, A.J., Gilliland, G.L., Finzel, B.C., Poulos, T.L., Ohlendorf, D.H., Salemme, F.R. (1987) The use of an imaging proportional counter in macromolecular crystallography. *J. Appl. Cryst.* **20**: 383-387.

- Jones, T.A. (1985) Interactive computer graphics: FRODO. *Methods Enzymol.* **115**:157-171.
- Lam, K.-B., Marmur, J. (1977) Isolation and characterization of *Saccharomyces cerevisiae* glycolytic pathway mutants. *J. Bacteriol.* **130**:746-749.
- Laskowski, R.A., McArthur, M.W., Moss, D.S., Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**: 282-291.
- Mas, M.T., Chen, C.Y., Hitzeman, R.A., Riggs, A.D. (1986) Active human-yeast chimeric phosphoglycerate kinases engineered by domain interchange. *Science* **233**:788-790.
- Scopes, R.K. (1978) Binding of substrates and other anions to yeast phosphoglycerate kinase. *Eur. J. Biochem.* **91**:119-121.
- Sherman, M.A., Dean, S.A., Mathiowetz, A.M., Mas, M.T. (1991) Site-directed mutations of arginine 65 at the periphery of the active site cleft of yeast 3-phosphoglycerate kinase enhance the catalytic activity and eliminate anion-dependent activation. *Protein Eng.* **4**:935-940.
- Tronrud, D.E., Ten Eyck, L.F., Matthews, B.W. (1987) An efficient general purpose least-squares refinement program for macromolecular structures. *Acta. Cryst.* **A43**: 489-501.
- Watson, H.C., Walker, N.P.C., Shaw, P.J., Bryant, T.N., Wendell, P.L., Fothergill, L.A., Perkins, R.E., Conroy, S.C., Dobson, M.J., Tuite, M.F., Kingsman, A.J., Kingsman, S.M. (1982) Sequence and structure of yeast phosphoglycerate kinase. *EMBO J.* **1**:1635-1640.

Table 2.1 *Data collection statistics for yeast R65Q PGK complexed with AMP-PNP and 3-PG. Statistics are based on processing runs in which Bijvoet pairs were merged.*

Data Collection Statistics		
	native	CMNP derivative
Resolution limit (Å)	2.43	2.75
Number of measured reflections	100828	52248
Number of independent reflections	16281	10987
Data completion	92.4%	90.4%
R _{sym} ^a	7.65%	7.36%

^aUnweighted, absolute-value R-factor on intensity.

Table 2.2 *RMS and average deviations between the various phase sets and the 1QPG model phases. R-factors and model completeness are reported for phase sets based on atomic models. Phase errors were calculated in the range 8.0 - 2.9 Å.*

Phase Set	Fraction of Model Built	R-factor for Model	RMS Phase Error	Average Phase Error
SIR ^a	-	-	92.0°	75.0°
SIRAS ^b	-	-	76.5°	58.6°
MR ^c	-	0.293 (7.0 - 2.6 Å)	95.0°	77.8°
SIRAS*MR ^d	-	-	77.6°	59.4
FLAT ^e	-	-	68.1°	50.1°
Model_1 ^f	72.8 %	0.390 (20.0 - 2.9 Å)	71.0°	52.1°
Model_1*FLAT ^g	-	-	59.5°	42.1°
Model_2 ^f	80.1 %	0.334 (20.0 - 2.9 Å)	66.3°	47.5°
Model_2*FLAT ^g	-	-	56.9°	39.9°
Model_3 ^f	83.4 %	0.309 (20.0 - 2.9 Å)	65.0°	46.1°
Model_3*FLAT ^g	-	-	56.4°	39.5°
Model_4 ^f	91.6 %	0.251 (20.0 - 2.9 Å)	53.8°	35.8°
Model_5 ^f	89.1 %	0.263 (20.0 - 2.7 Å)	51.5°	34.0°
Model_6 ^f	89.7 %	0.246 (20.0 - 2.7 Å)	49.0°	31.9°
Model_7 ^f	93.8 %	0.202 (8.0 - 2.4 Å)	39.9°	24.2°
Model_8 ^f	99.3 %	0.199 (8.0 - 2.4 Å)	36.4°	21.4°
Model_9 ^f	97.3 %	0.198 (8.0 - 2.4 Å)	30.5°	17.4°
Model_10 ^f	99.6 %	0.183 (8.0 - 2.4 Å)	12.5°	5.6°
Model_11 ^f	99.9 %	0.187 (8.0 - 2.4 Å)	6.6°	2.2°
Model_12 ^f	100.0 %	0.185 (8.0 - 2.4 Å)	9.2°	3.6°
Model_13 ^f	100.0 %	0.185 (8.0 - 2.4 Å)	0.0°	0.0°

^aPhases from PHASIT run using single isomorphous replacement effects only.

^bPhases from PHASIT run using both single isomorphous replacement and anomalous scattering effects.

^cPhases calculated from the refined molecular replacement solution based on the 3PGK model.

^dPhases resulting from combining the MR and SIRAS phase sets listed above.

^ePhases resulting from solvent flattening of the SIRAS phases using the PHASES package.

^fPhases calculated solely from the indicated model.

^gPhases resulting from combining the indicated model phases with the FLAT phases.

Table 2.3 *Single isomorphous replacement (SIR) and anomalous scattering (AS) phasing statistics for the CMNP derivative using the program PHASIT.*

Resolution limit (Å)	2.75
Unique reflections	9520
Centric reflections	479
Acentric reflections	
Bijvoets merged for isomorphous replacement	9041
Bijvoets unmerged for anomalous scattering	17394
Phase refinement passes	10
Parameter refinement cycles per pass	15
Isomorphous replacement (SIR) statistics	
Cullis R (centric reflections)	0.469
Kraut R (acentric reflections)	0.076
Mean FOM (centric reflections)	0.748
Mean FOM (acentric reflections)	0.448
Mean FOM (all reflections)	0.463
Phasing power	2.72
Anomalous scattering (AS) statistics	
Kraut R (acentric reflections)	0.279
Mean FOM (all reflections)	0.310
Phasing power	2.47
Combined SIR and AS statistics	
Mean FOM (all reflections)	0.665

Table 2.4 *Characteristics of models refined during model-building and refinement stage of structure solution. R-factors listed are those resulting from refinement of the respective model.*

Model	Sequence Ranges	Residues	Missing Sidechains	Substrates Included	Waters	Atoms	R-factor (resolution)
1	16 - 84, 152 - 410	328	55	AMP-PNP	0	2322	0.390 (20.0 - 2.9 Å)
2	16 - 94, 104 - 117, 143 - 409	360	88	AMP-PNP	0	2557	0.334 (20.0 - 2.9 Å)
3	5 - 117, 134 - 415	395	107	AMP-PNP	0	2662	0.309 (20.0 - 2.9 Å)
4	1 - 415	415	82	AMP-PNP	0	2923	0.251 (20.0 - 2.9 Å)
5	1 - 337, 345 - 415	408	86	AMP-PNP	0	2842	0.263 (20.0 - 2.7 Å)
6	1 - 415	415	91	AMP-PNP	0	2864	0.246 (20.0 - 2.7 Å)
7	1 - 415	415	50	AMP-PNP, Mg ²⁺	1	2992	0.202 (8.0 - 2.4 Å)
8	1 - 415	415	4	AMP-PNP, Mg ²⁺	0	3169	0.199 (8.0 - 2.4 Å)
9	2 - 414	413	19	AMP-PNP, Mg ²⁺	0	3104	0.198 (8.0 - 2.4 Å)
10	1 - 414	414	16	AMP-PNP, Mg ²⁺ , 3-PG	81	3177	0.183 (8.0 - 2.4 Å)
11	1 - 414	414	16	AMP-PNP Mg ²⁺ , 3-PG	81	3189	0.187 (8.0 - 2.4 Å)
12	1 - 414	414	20	AMP-PNP Mg ²⁺ , 3-PG	82	3195	0.185 (8.0 - 2.4 Å)
13	1 - 414	414	15	AMP-PNP Mg ²⁺ , 3-PG	82	3191	0.185 (8.0 - 2.4 Å)

Table 2.5 *TNT refinement scheme for the first model. Refinement was performed in nine stages at successively higher resolutions up to the resolution at which model building had been performed. Multiple long cycles of refinement were performed as indicated at each stage. Low resolution data were included during the TNT refinements since simple bulk solvent corrections are standard in this package.*

Refinement Stage	Long Cycles Performed	Resolution Range (Å)	Reflections Used	Final R-factor
1	4	20 - 3.9	4629	0.406
2	4	20 - 3.8	4991	0.388
3	5	20 - 3.7	5428	0.367
4	5	20 - 3.6	5889	0.359
5	5	20 - 3.5	6440	0.363
6	4	20 - 3.4	7022	0.356
7	4	20 - 3.3	7713	0.354
8	5	20 - 3.2	8471	0.358
9	9	20 - 2.9	10321	0.390

Table 2.6 *X-PLOR refinement scheme for model 7. Prior to this refinement, TNT had been used exclusively and no data beyond 2.7 Å had been used. The X-PLOR refinement was carried out in three steps. First, the structure was refined at 2.7 Å resolution using a 2.0 σ cutoff. After this refinement converged, the sigma cutoff was eliminated and the refinement continued. Finally, all the data to 2.4 Å was included in the refinement. In all three steps, atomic positions (xyz) and temperature factors (B) were refined alternately. The WA value used was 800,000 in all xyz refinements of model 7, or 2.5 times the value recommended by the CHECK procedure.*

Stage	Refinement Type	Number of Cycles	Sigma Cutoff	Resolution Range (Å)	Reflections Used	Final R-factor
1	xyz	600	2.0	8.0 - 2.7	10379	0.222
2	B	20	2.0	8.0 - 2.7	10379	0.195
3	xyz	135	2.0	8.0 - 2.7	10379	0.194
4	B	20	2.0	8.0 - 2.7	10379	0.185
5	xyz	154	2.0	8.0 - 2.7	10379	0.185
6	B	20	2.0	8.0 - 2.7	10379	0.180
7	xyz	164	2.0	8.0 - 2.7	10379	0.181
8	B	20	2.0	8.0 - 2.7	10379	0.178
9	B	20	-	8.0 - 2.7	10972	0.191
10	xyz	208	-	8.0 - 2.7	10972	0.189
11	B	20	-	8.0 - 2.7	10972	0.187
12	xyz	222	-	8.0 - 2.4	13387	0.206
13	B	20	-	8.0 - 2.4	13387	0.202

Table 2.7 *Final refinement statistics for model 1QPG.*

Resolution range (Å)	8.0 - 2.4
Number of reflections	13387 (78.3% complete)
Number of scatterers	3271
Water molecules	82
WA weight used	3.6×10^5
rms ΔB values for adjacent atoms	2.0
Crystallographic R-factor	0.185
rms deviations from ideal values	
Bonds (Å)	0.009
Angles (°)	1.78
Dihedrals(°)	21.7
Impropers(°)	1.26

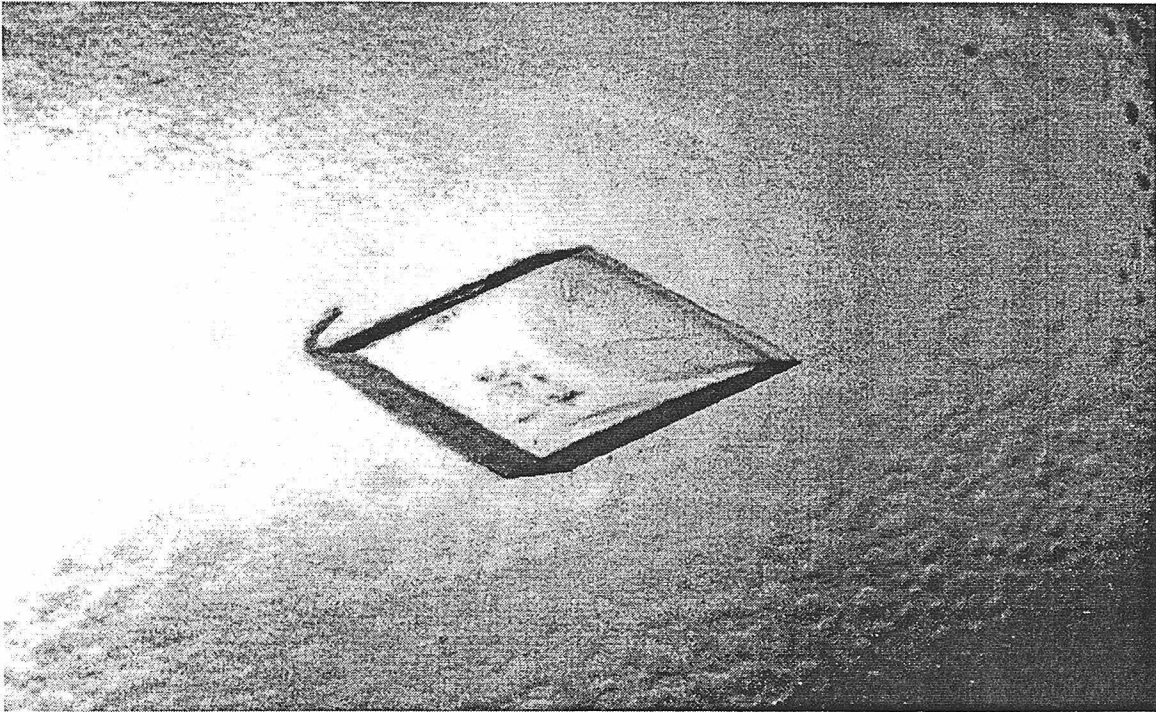


Figure 2.1 *Crystal of yeast R65Q PGK complexed with AMP-PNP and 3-PG.*

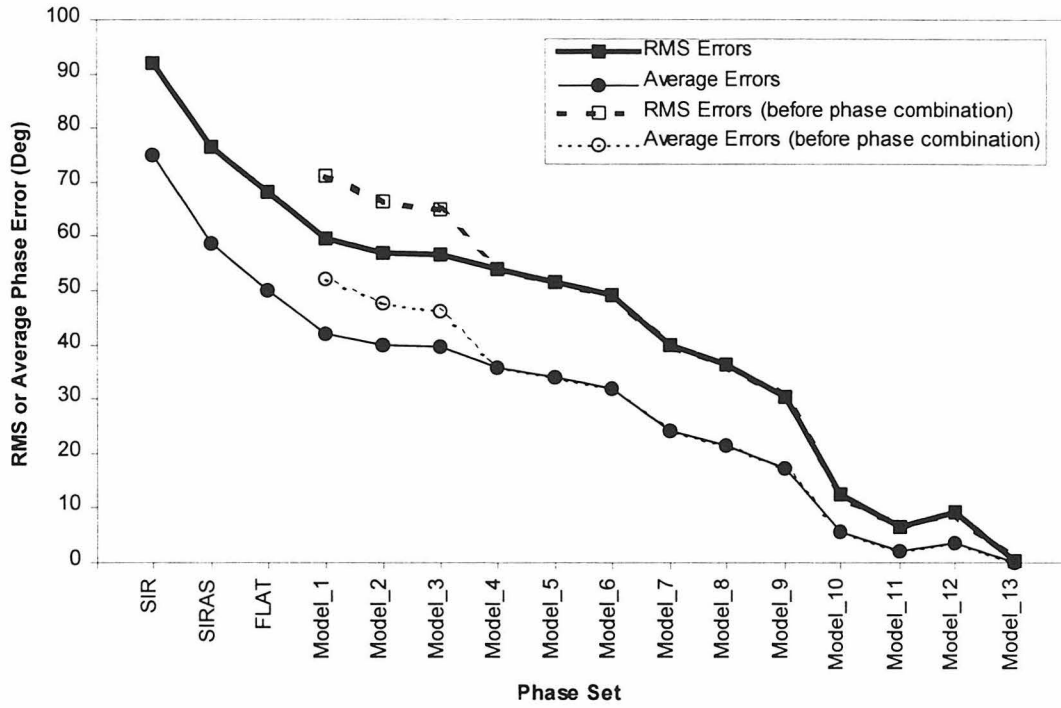


Figure 2.2 RMS and average errors in the phase sets used in solving the structure. The phase sets are ordered chronologically to show the improvement in phases during the course of the structure solution. The filled shapes connected by solid lines indicate rms and average errors in the phase sets from which the maps were calculated. The empty shapes connected by dashed lines correspond to model phase sets that were combined with the FLAT phases to yield the phase sets indicated by the filled shapes directly below.

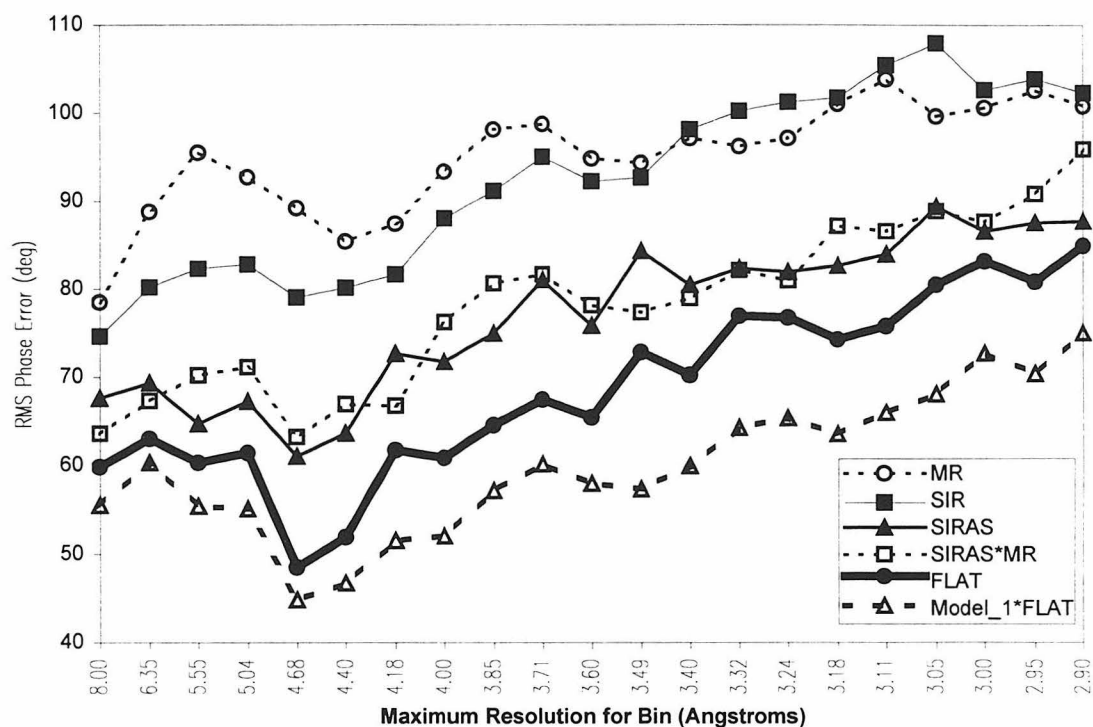


Figure 2.3 RMS phase error versus resolution for a number of phase sets used in the initial structure solution. The heavy lines indicate the solvent flattened phases (FLAT) into which the initial model was built. The MR, SIR, SIRAS, and SIRAS*MR phase sets are clearly less accurate than the FLAT phase set. The combination of the Model_1 and FLAT phases are more accurate than the FLAT phases.

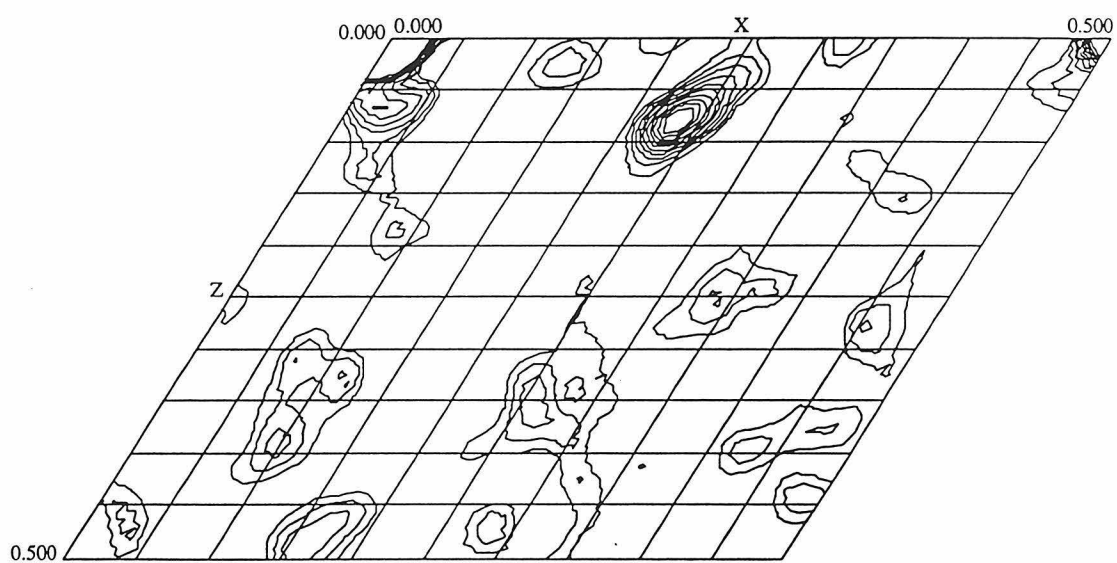
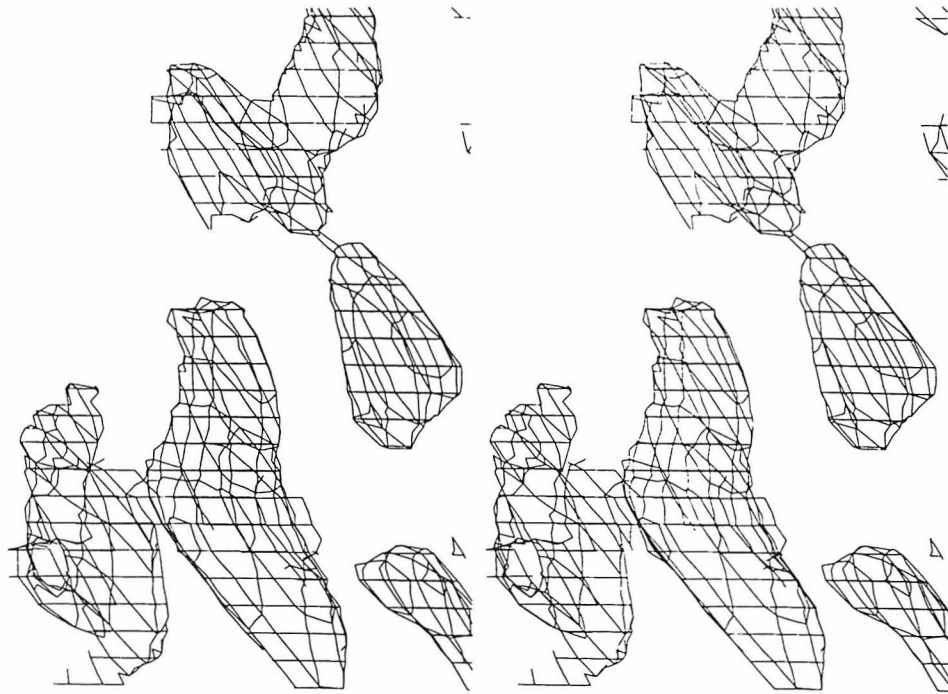


Figure 2.4 *The $v=0$ Harker section of the difference Patterson used in locating the Hg(II) heavy atom in the CMNP derivative data set.*

Figure 2.5 *The 26 panels on the next 13 pages show electron density calculated from six phase sets defined in Table 2.2: SIR (panels a-d), SIRAS (panels e-h), MR (panels i-n), SIRAS*MR (panels o-r), FLAT (panels s-v), and Model_13 (panels w-z). Two regions of the map are shown for each phase set. Region 1 is centered on phenylalanine-19 in the core of the N-domain. Region 2 is centered on the AMP-PNP molecule bound to the C-domain. Each view of the density is shown twice per page. The upper panel shows the density without any model built in. The lower panel shows the same density with the final 1QPG coordinates. Panels m and n show the MR-phased map with the MR model. All maps were calculated using structure factors in the 8.0-2.9 Å resolution range.*

a.



b.

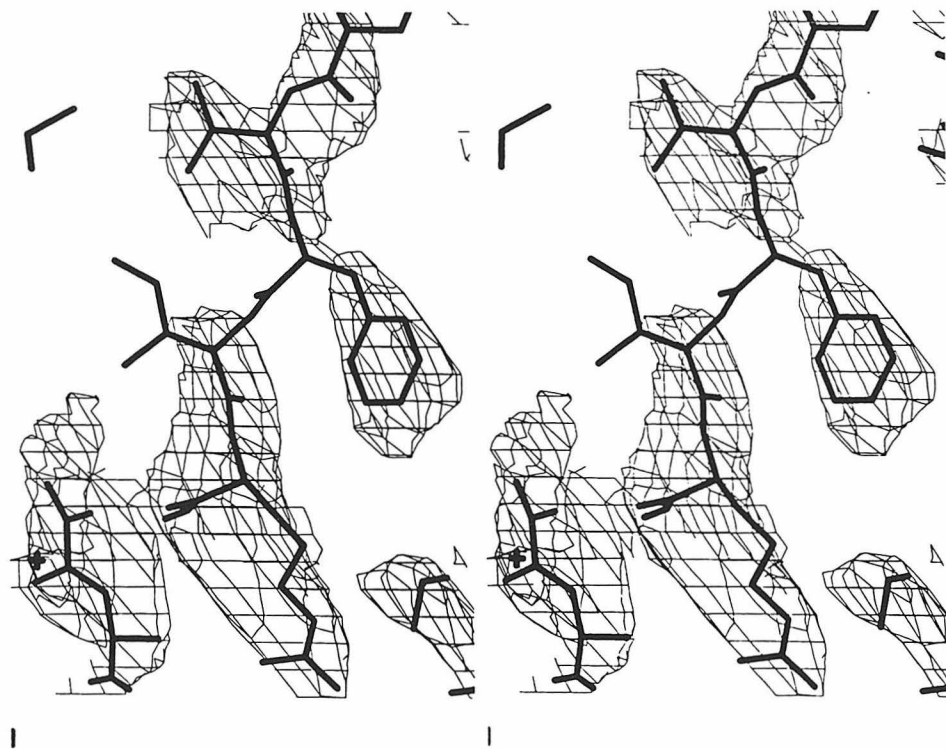
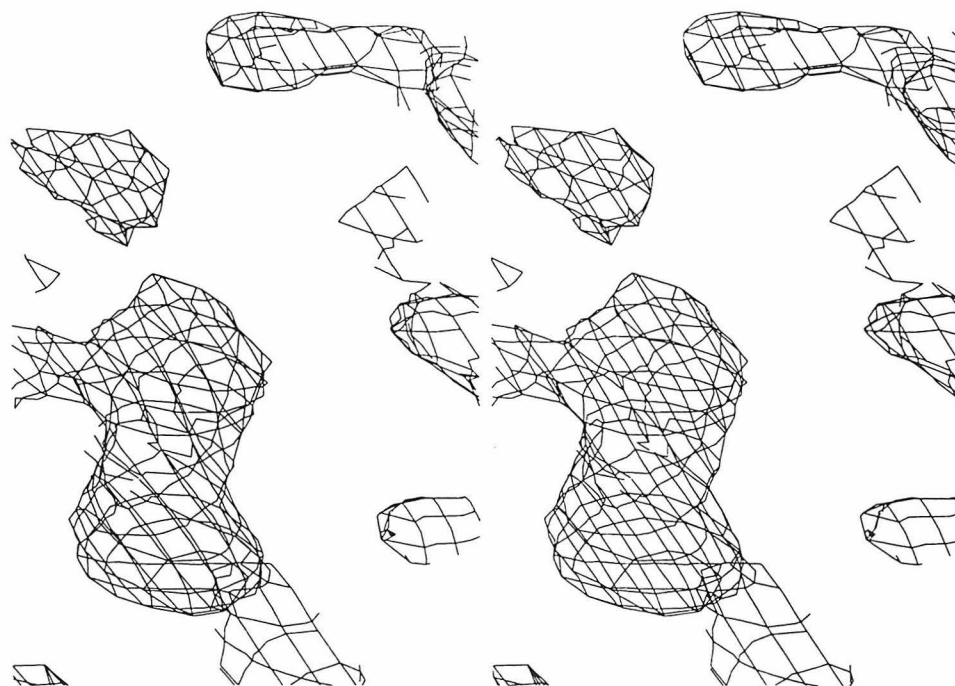


Figure 2.5 a-b *Region 1 of SIR-phased map.*

c.



d.

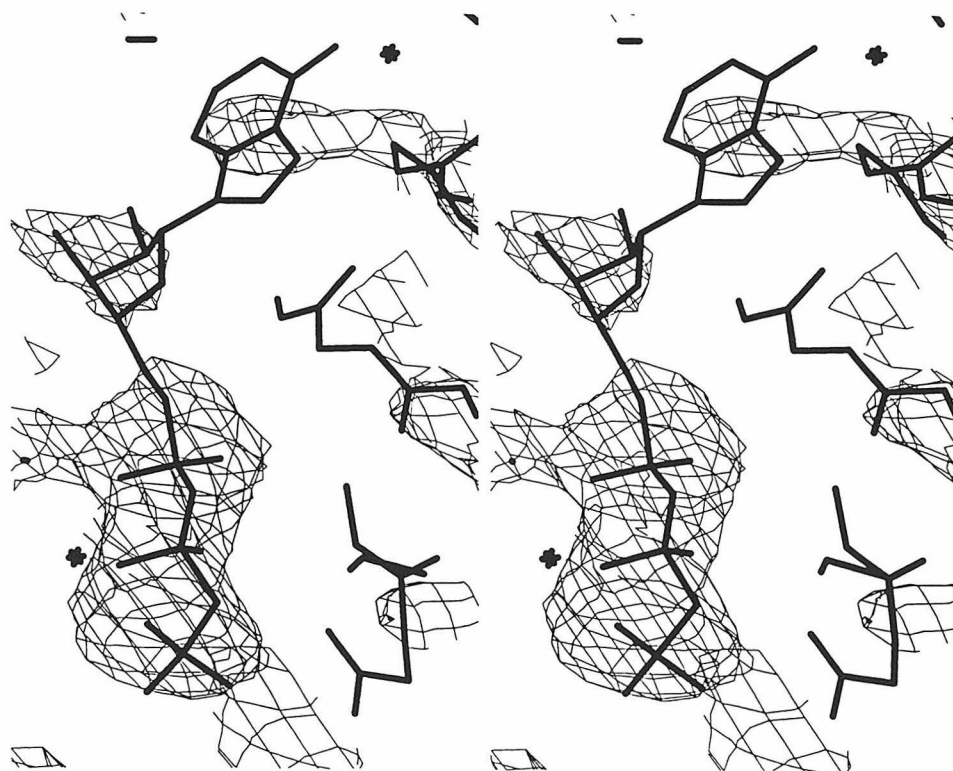
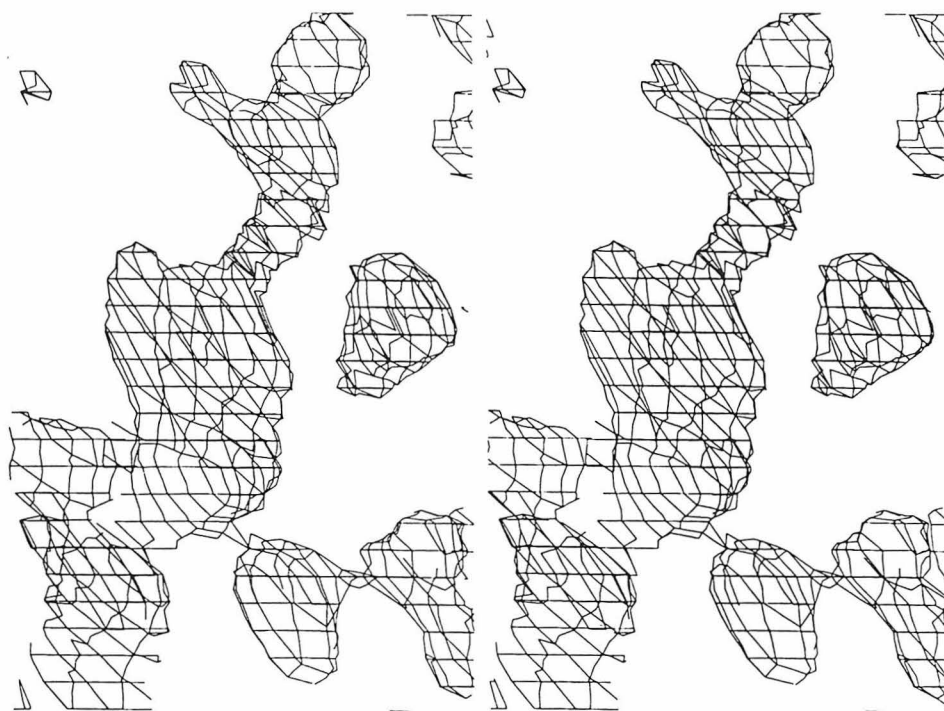


Figure 2.5 c-d *Region 2 of SIR-phased map.*

e.



f.

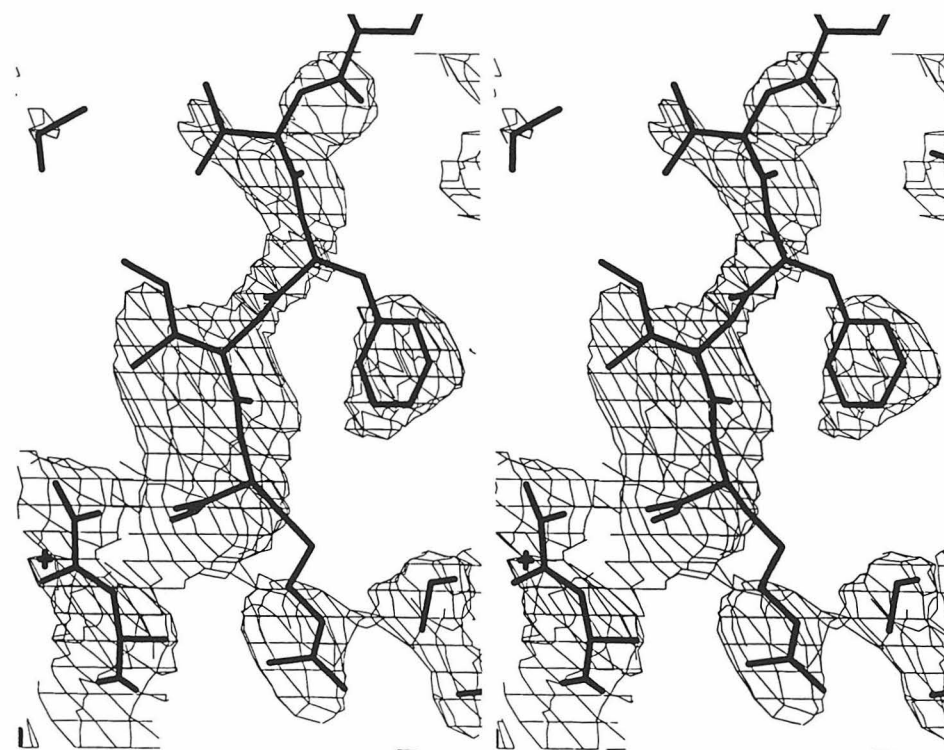
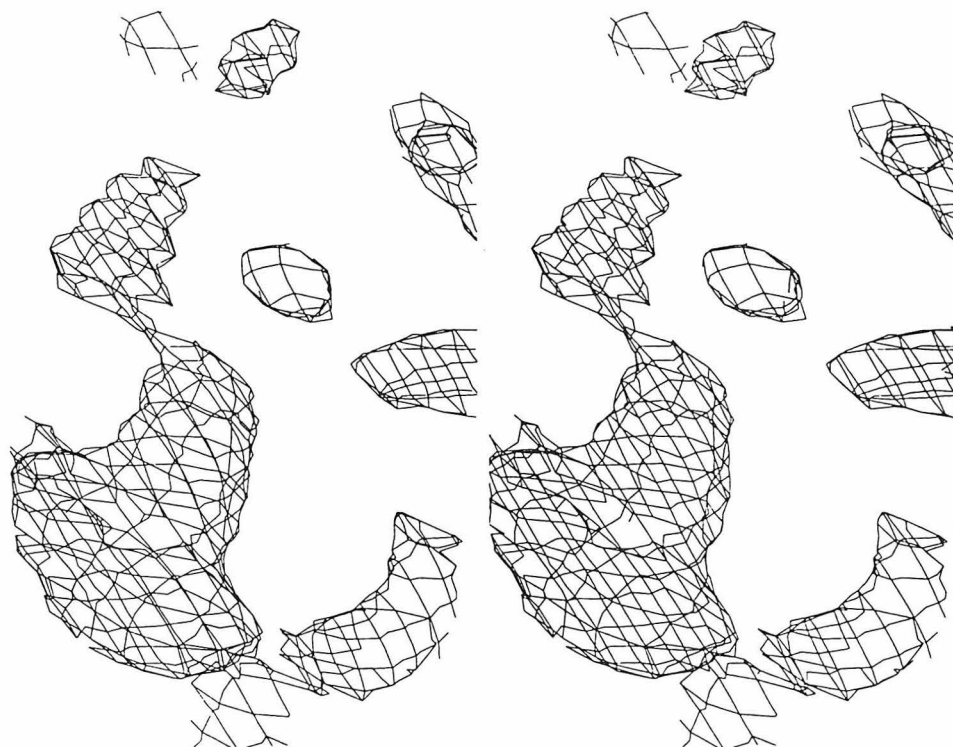


Figure 2.5 e-f *Region 1 of SIRAS-phased map.*

g.



h.

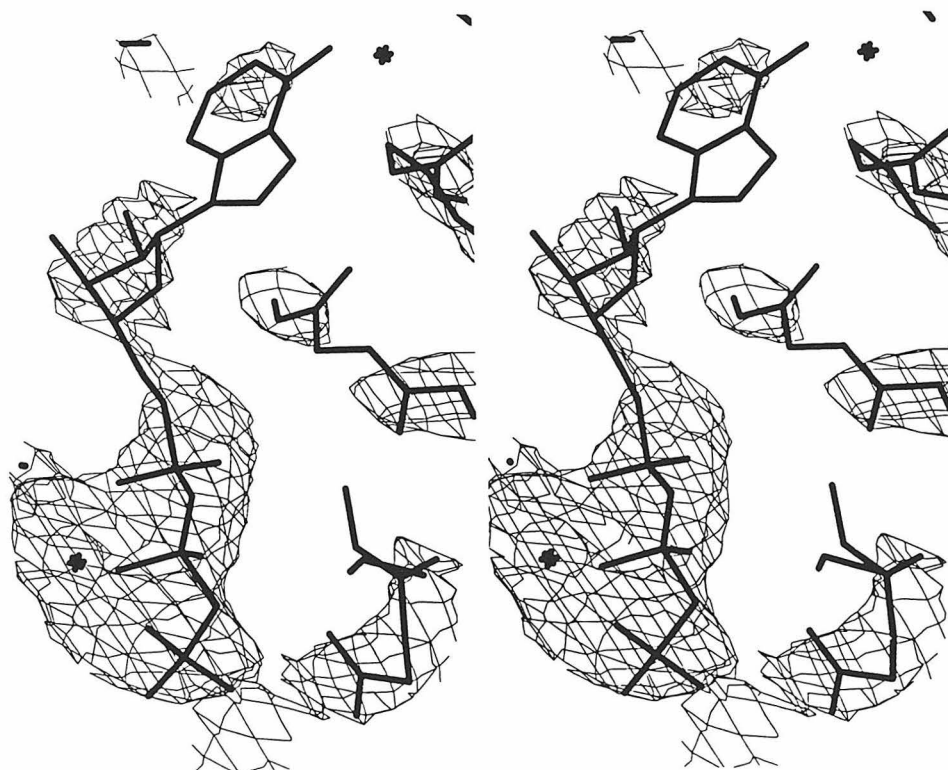
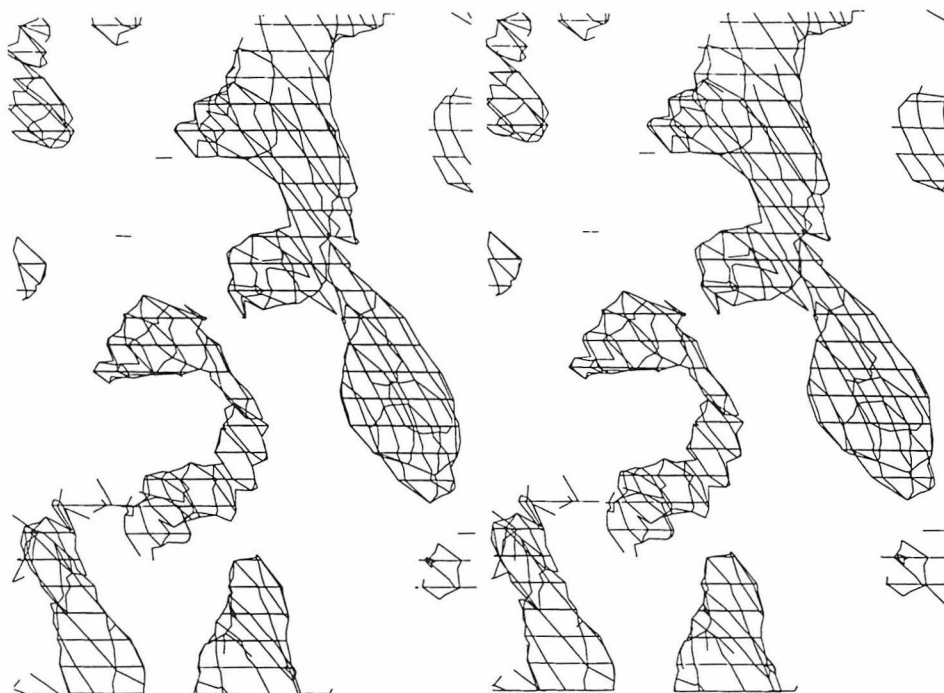


Figure 2.5 g-h *Region 2 of SIRAS-phased map.*

i.



j.

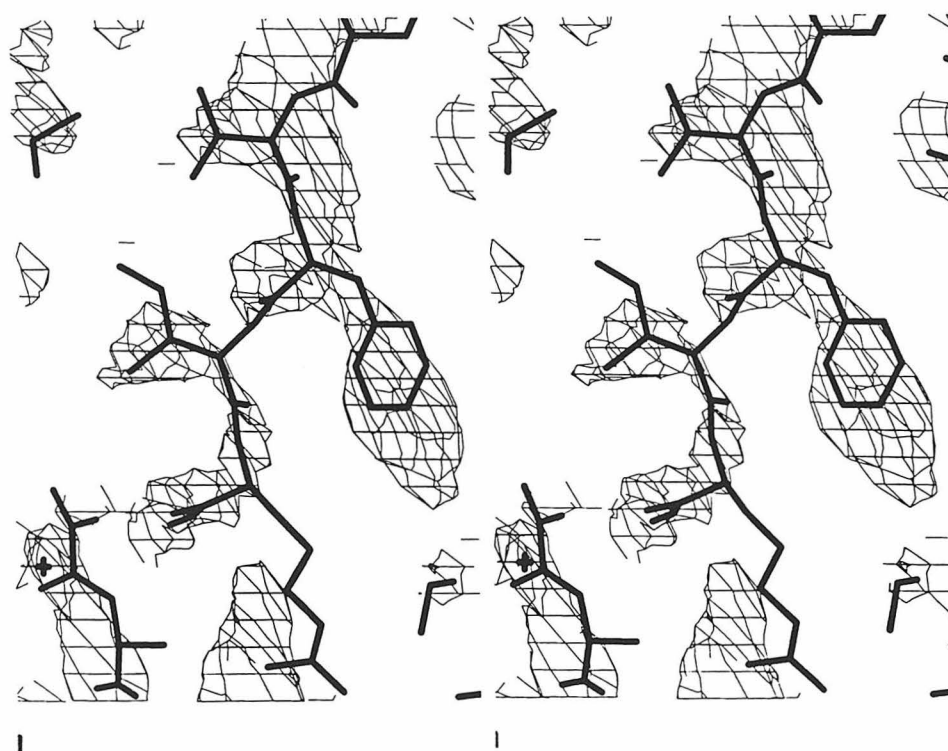
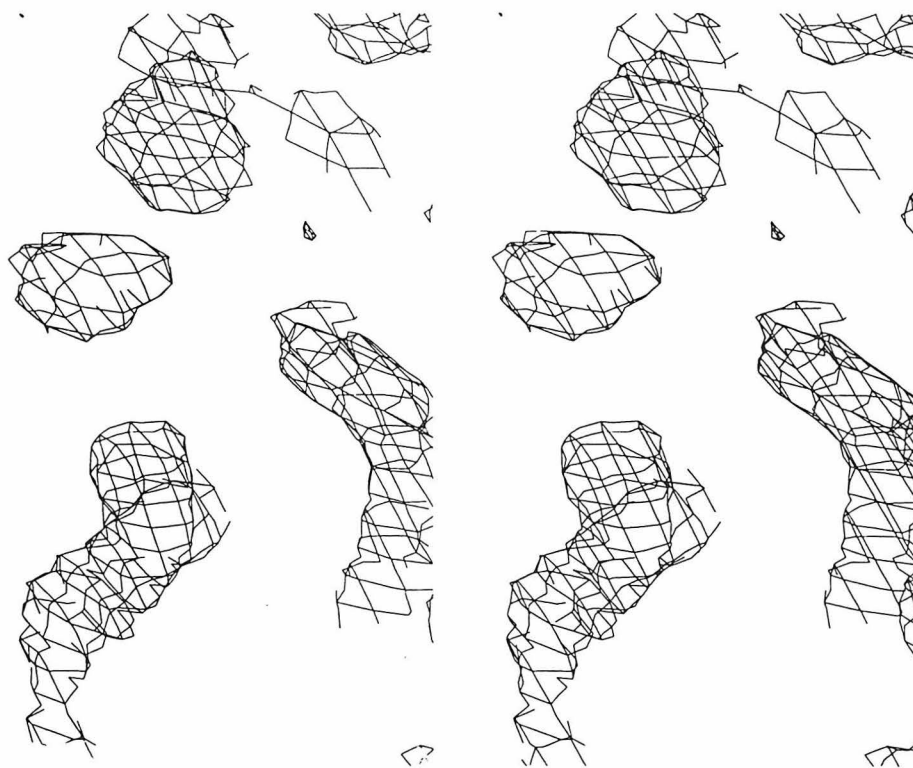


Figure 2.5 i-j *Region 1 of MR-phased map.*

k.



l.

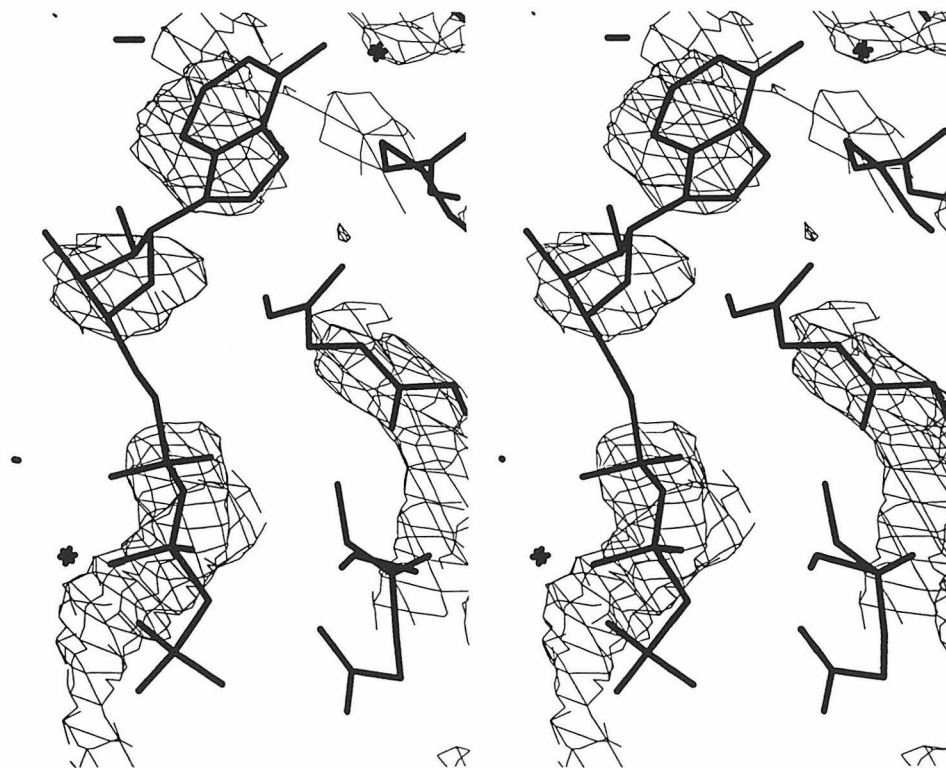
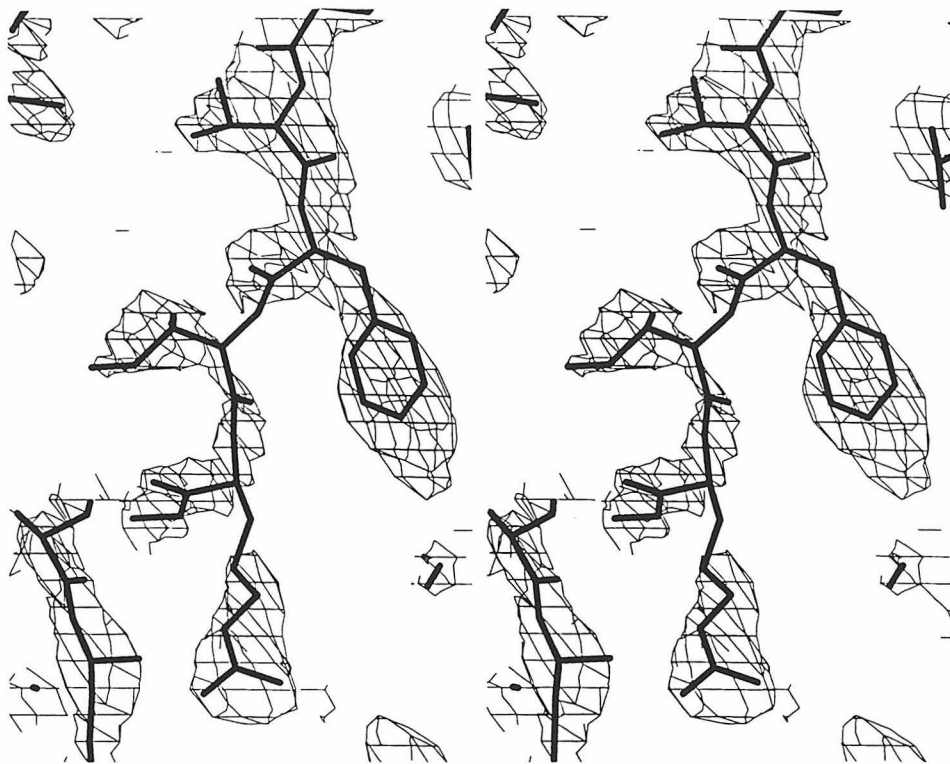


Figure 2.5 k-l *Region 2 of MR-phased map.*

m.



n.

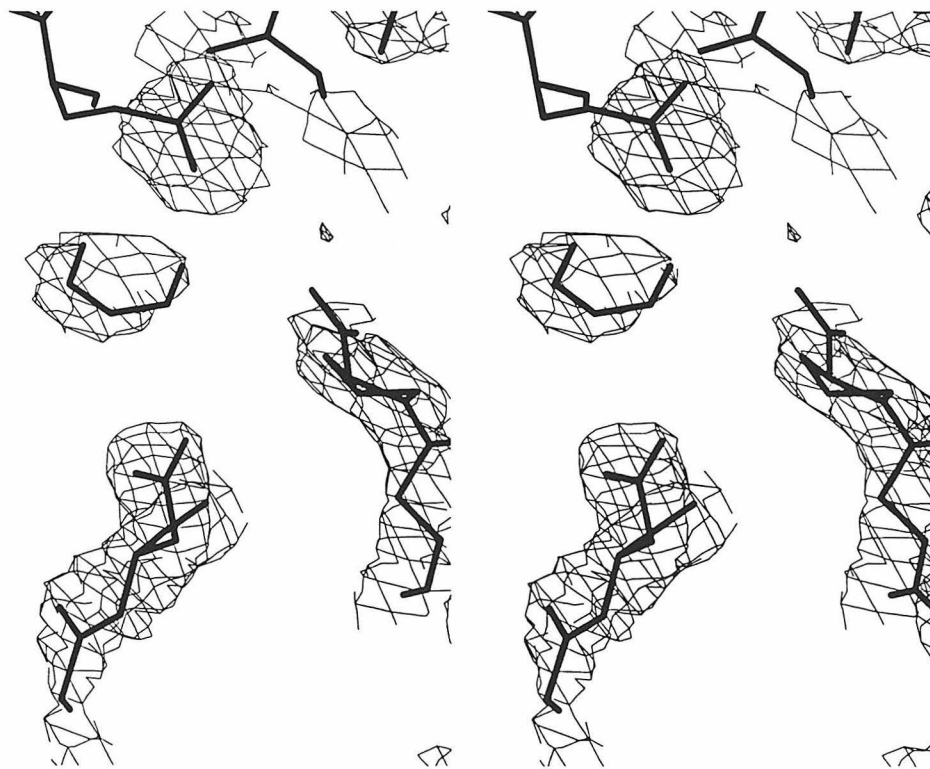
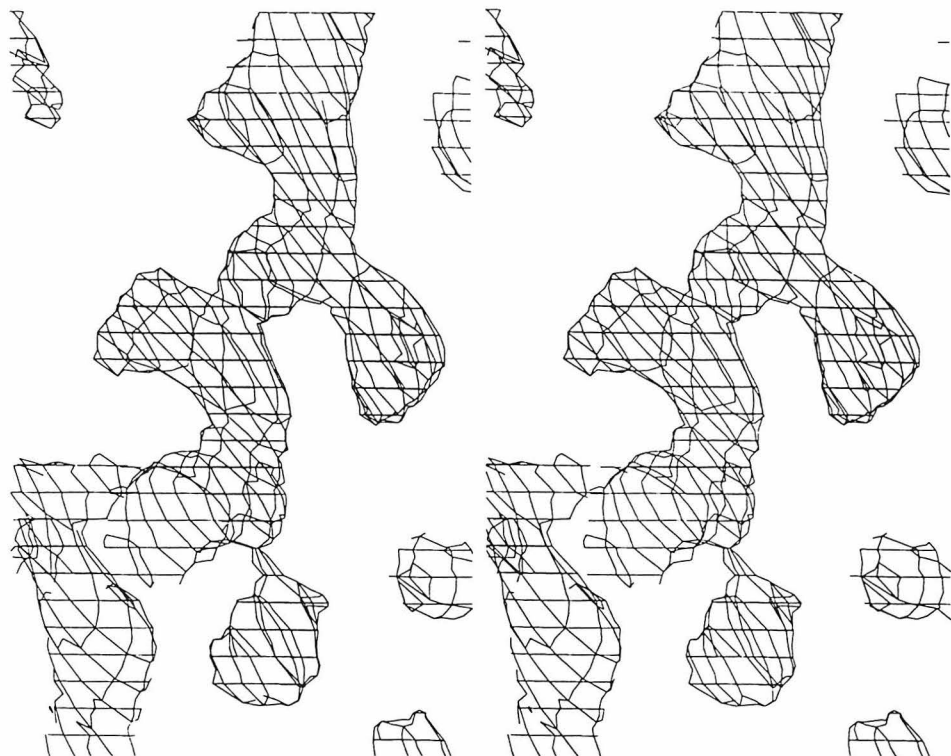


Figure 2.5 m-n *Regions 1 and 2 of MR-phased map with MR model.*

o.



p.

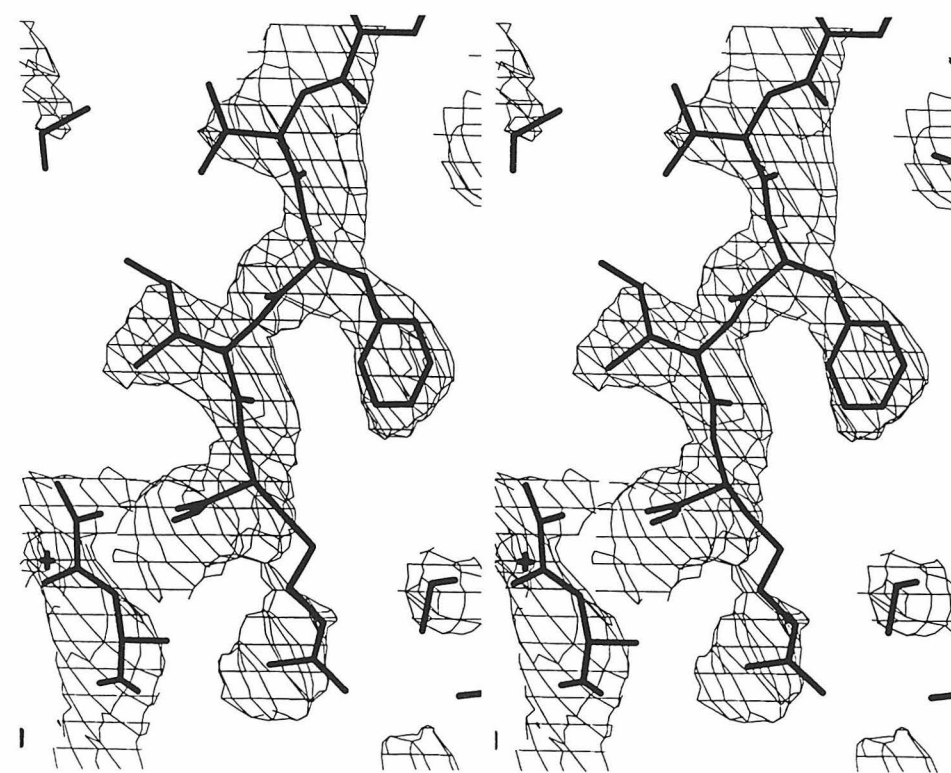
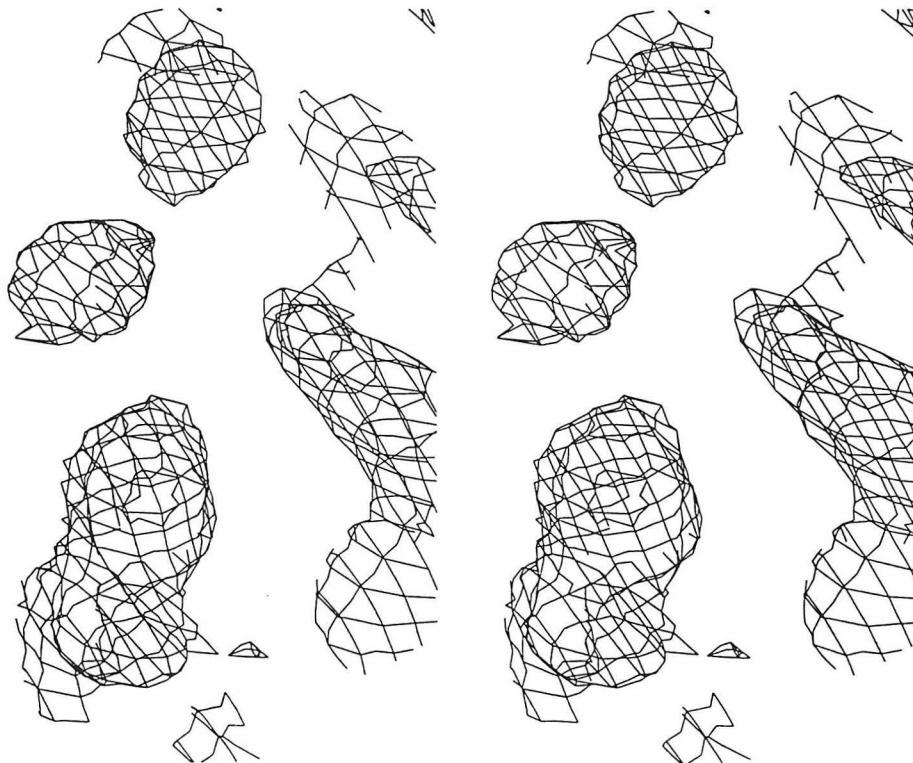


Figure 2.5 o-p *Region 1 of SIRAS*MR-phased map.*

q.



r.

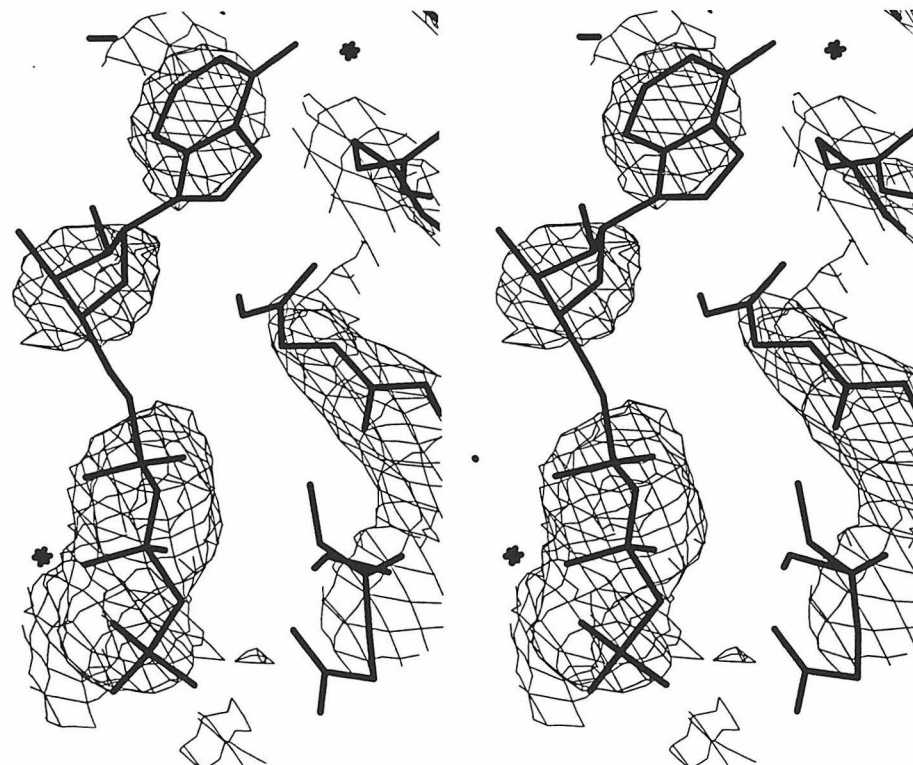
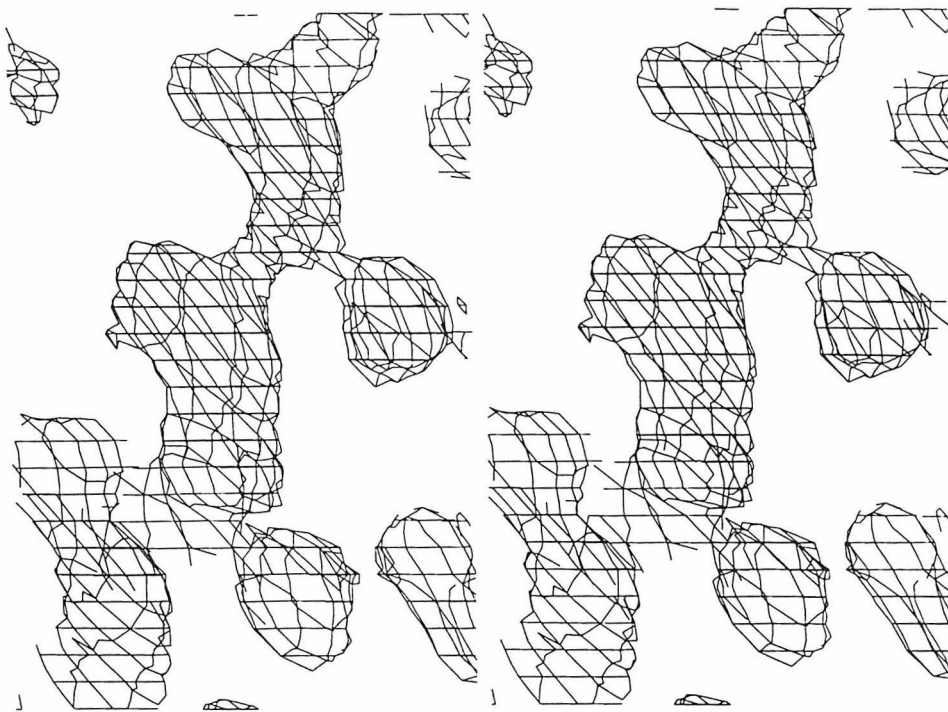


Figure 2.5 q-r Region 2 of SIRAS*MR-phased map.

s.



t.

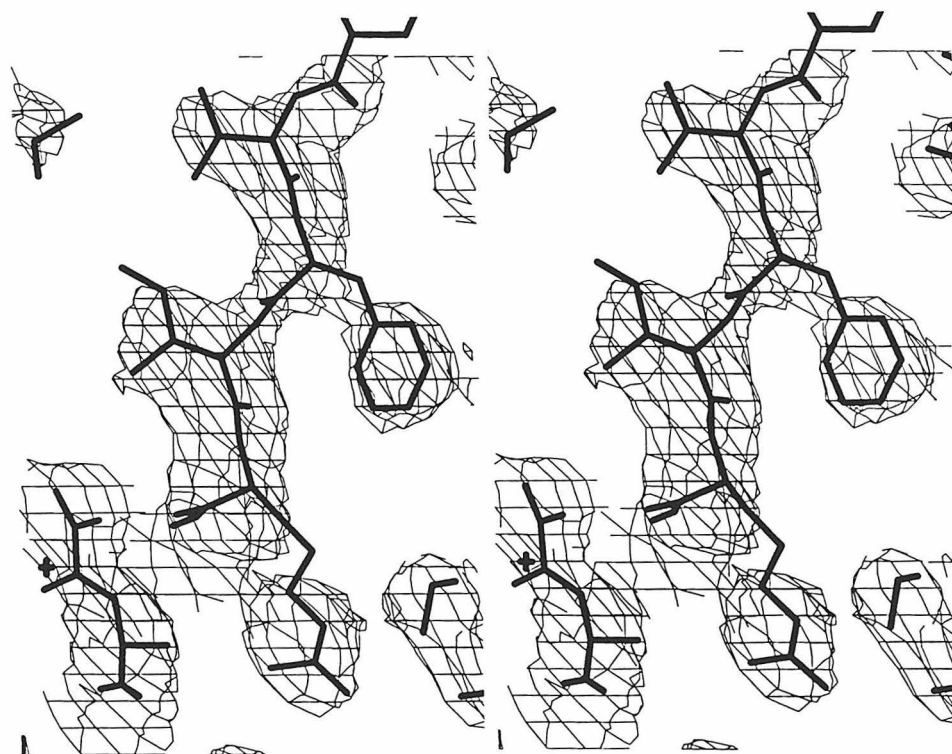
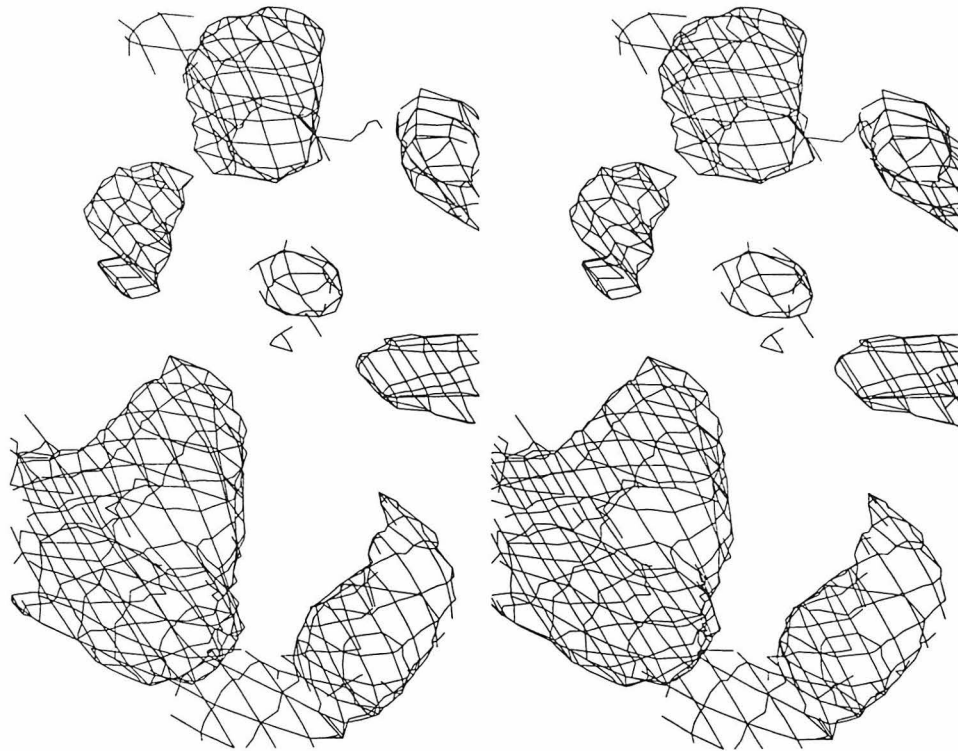


Figure 2.5 s-t *Region 1 of FLAT-phased map.*

u.



v.

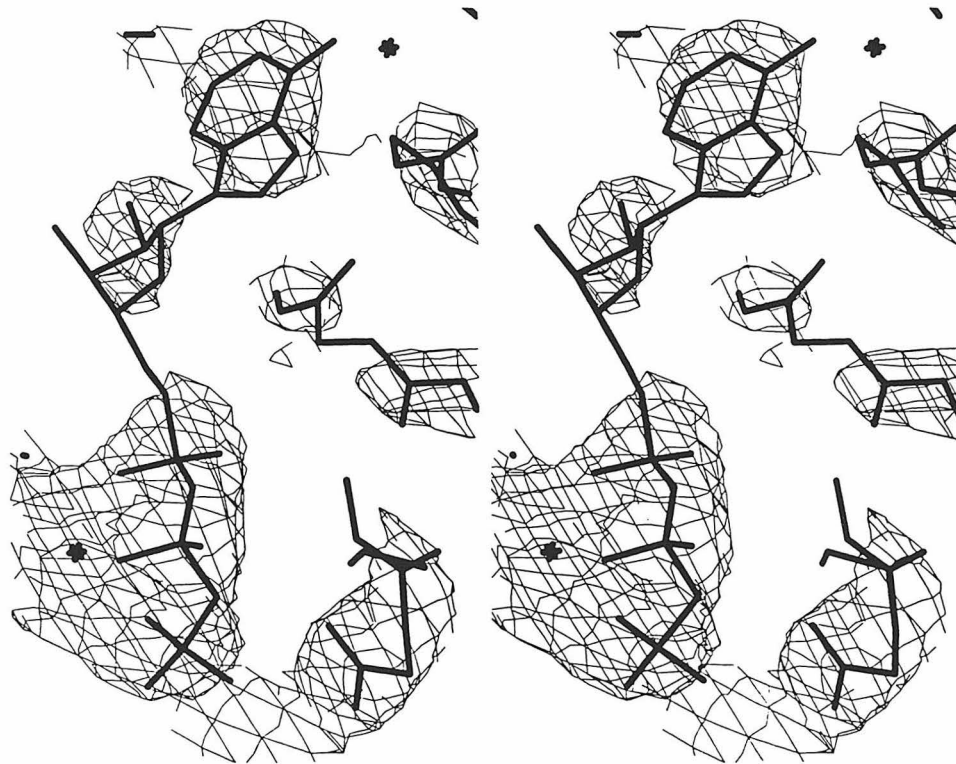
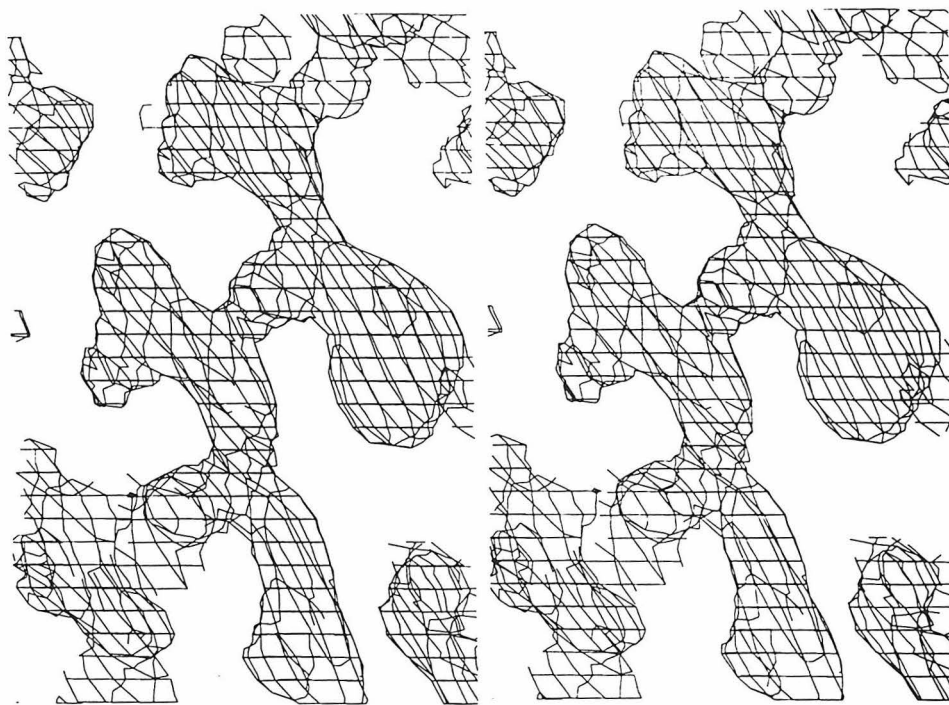


Figure 2.5 u-v Region 2 of FLAT-phased map.

w.



x.

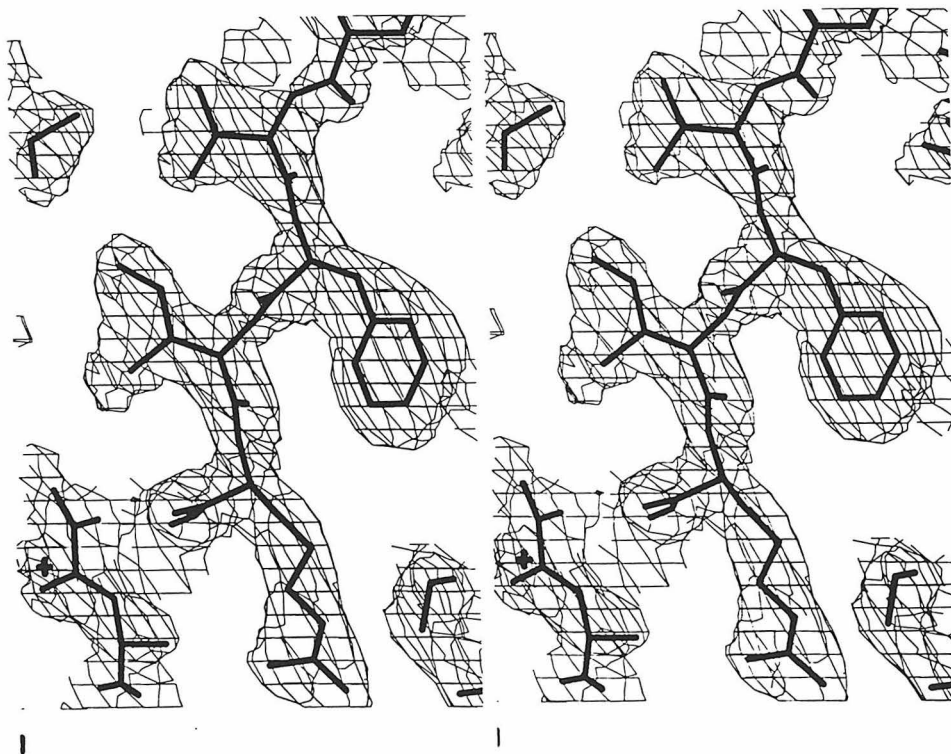
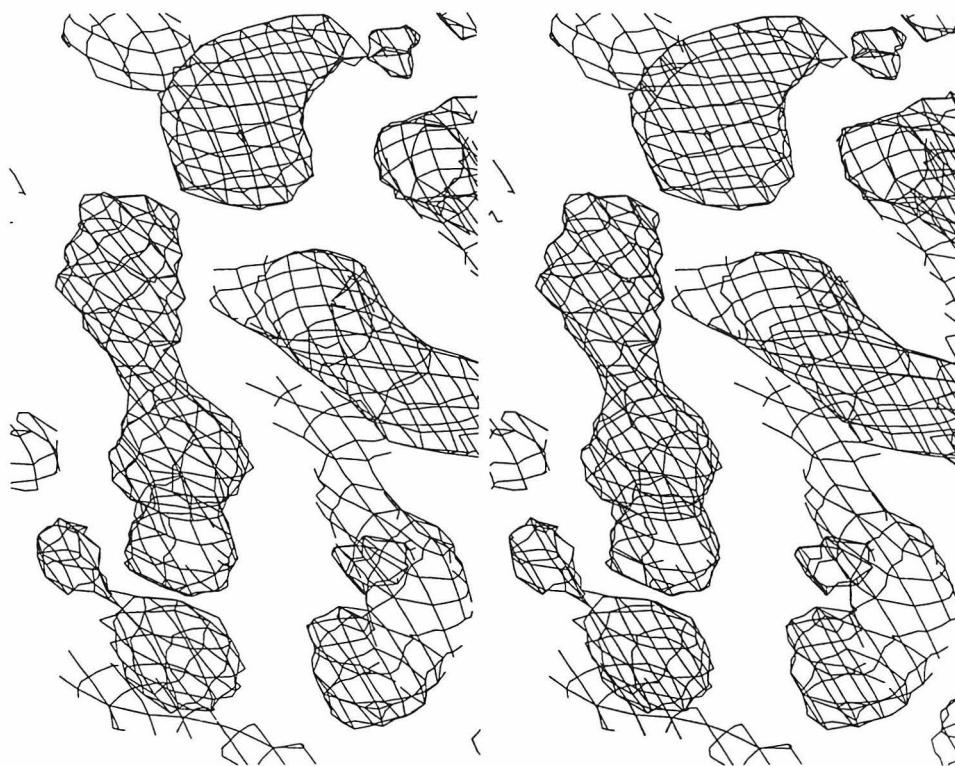


Figure 2.5 w-x Region 1 of Model_13-phased map.

y.



z.

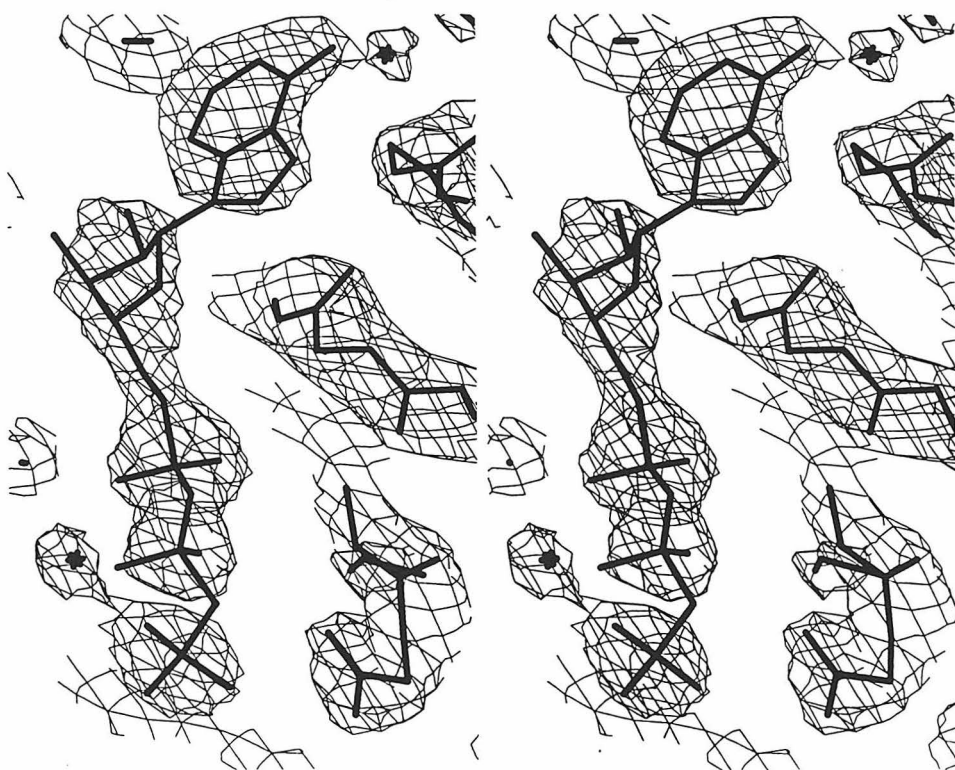


Figure 2.5 y-z Region 2 of Model_13-phased map.



Figure 2.6 Comparison of the MR model (the refined molecular replacement solution based on the 3PGK coordinates) with the final 1QPG coordinates. The MR model is shown in thick lines, the 1QPG model in thin lines. Note that while the N-domains of the two models (bottom half of the figure) superimpose fairly well, the C-domains (top half of the figure) do not.

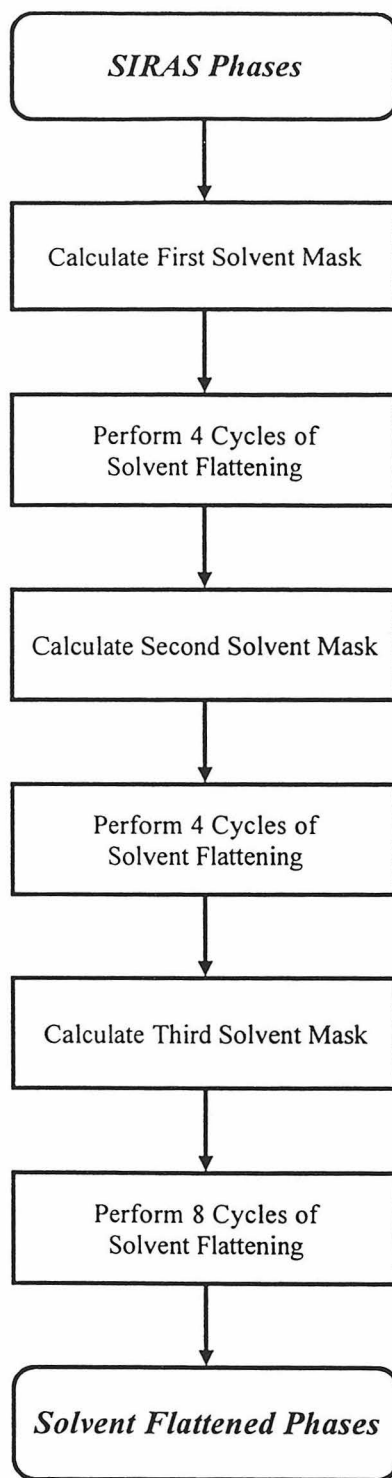


Figure 2.7 Flowcharts illustrating solvent flattening procedure using PHASES. (a) Overview of the procedure showing how solvent mask calculation and solvent flattening operations are performed alternately to arrive at the final phase set.

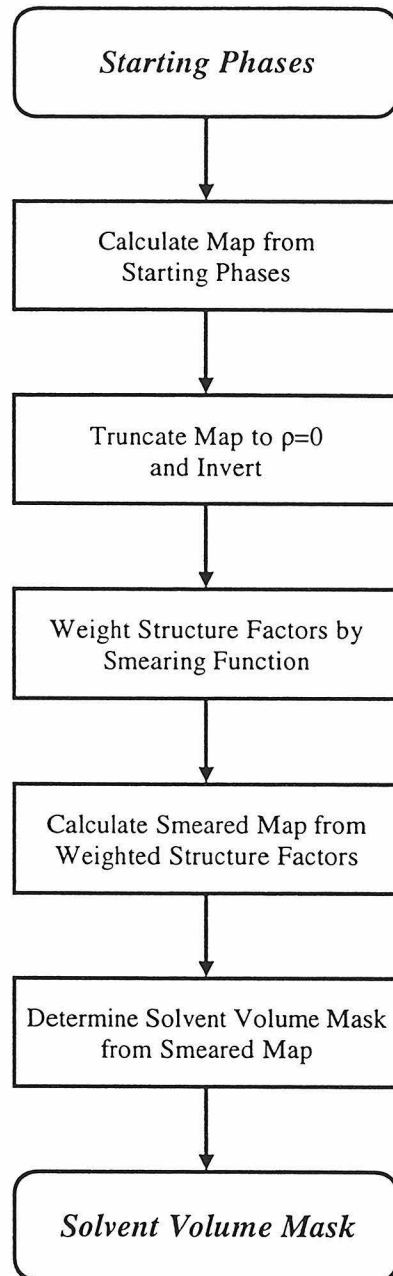


Figure 2.7 (b) *Solvent mask calculation procedure.*

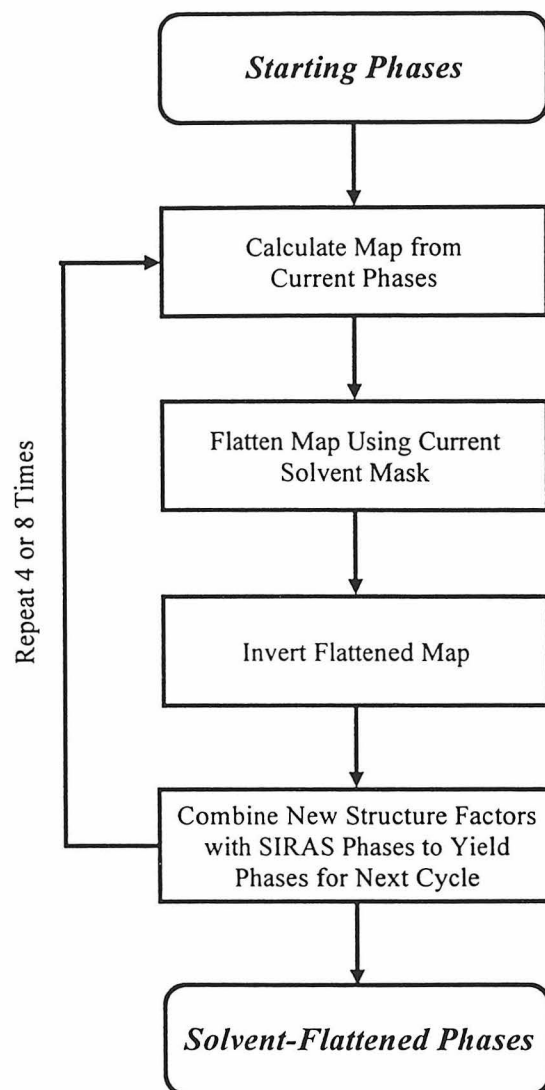


Figure 2.7 (c) *Solvent flattening procedure used for each solvent mask.*

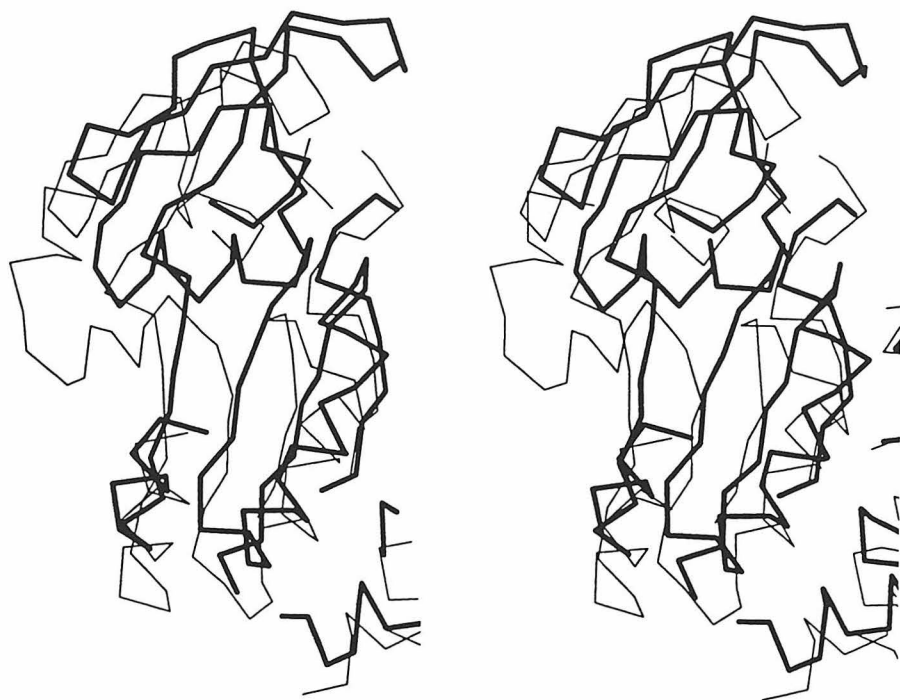


Figure 2.8 Superposition of C_{α} traces of C-domains in the 1QPG (heavy lines) and 3PGK models (thin lines). The 3PGK model was positioned in the experimental unit cell by performing a least-squares superposition with the refined molecular replacement model. Note that while the β -strands of the two models curve in opposite directions such that while the lower halves of the two structures superimpose well, the top halves deviate significantly.

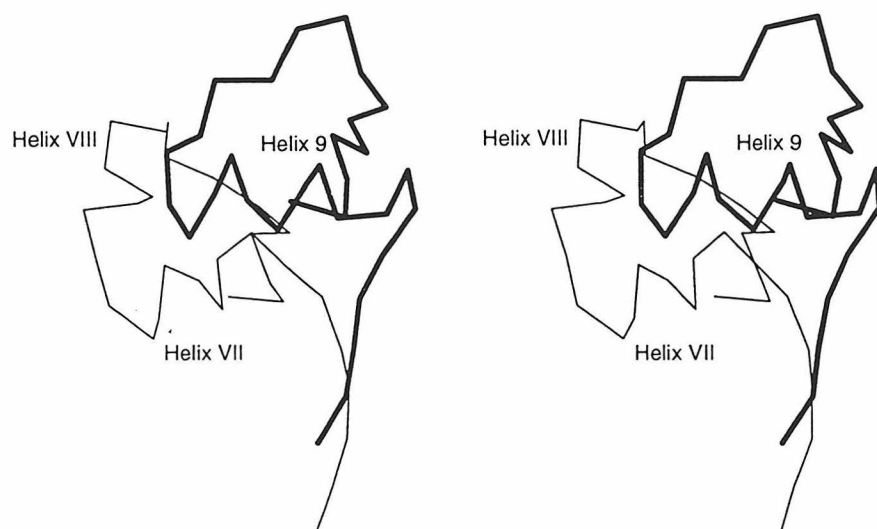


Figure 2.9 A detailed view of Figure 2.8 showing residues 230-260 of the C-domain in the 1QPG (heavy lines) and 3PGK models (thin lines). The 1QPG model has only one helix in this region, helix 9 which extends from residue 235 to 246. In contrast, the 3PGK model has two helices in the same region, helix VII from residue 236-244 which corresponds well with helix 9 of the 1QPG model, and helix VIII from residue 245-252 which corresponds to extended random coil in 1QPG.

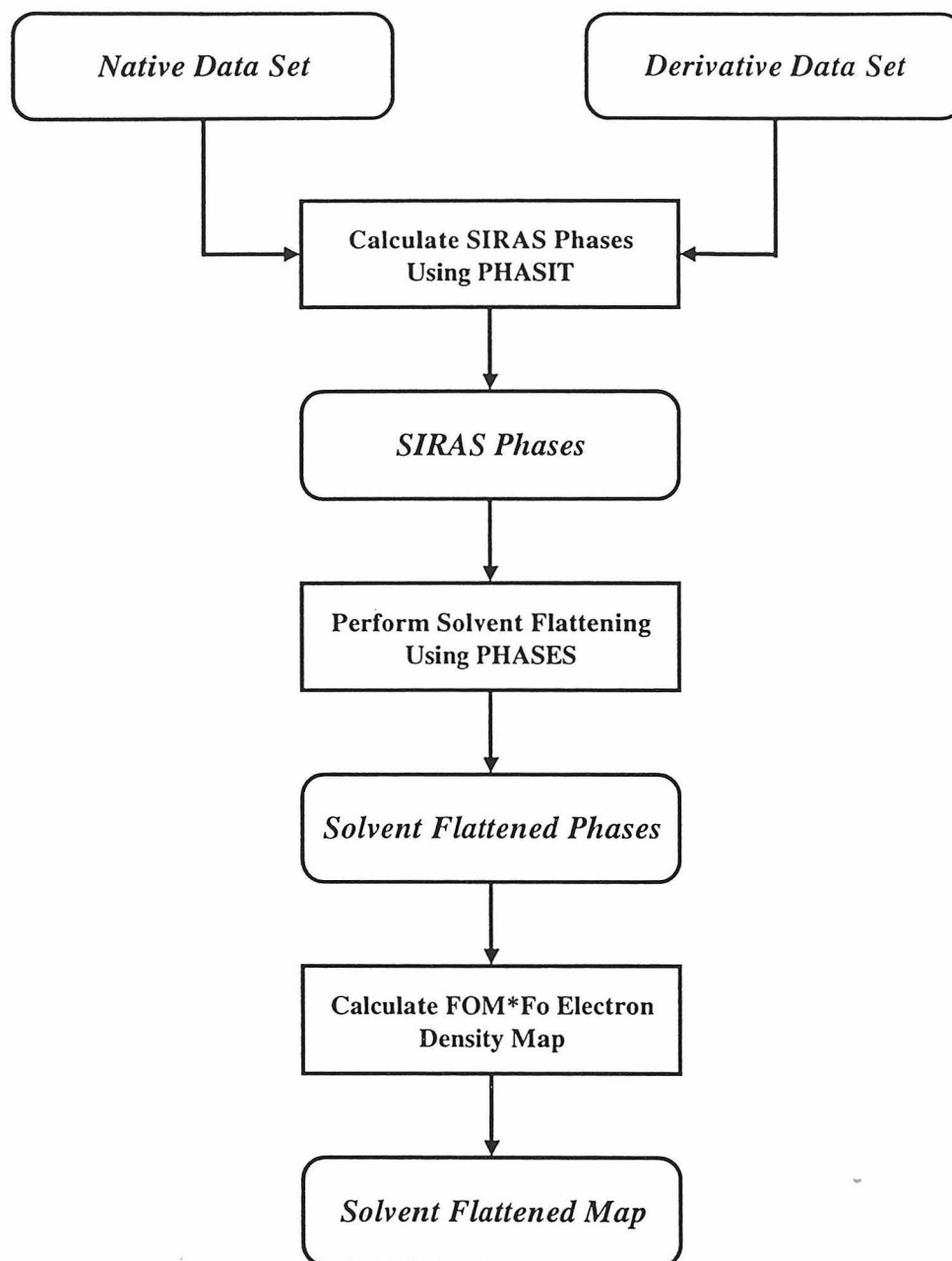


Figure 2.10 Flowcharts illustrating methods used in solving the structure. (a) How the solvent flattened map (into which the initial model was built) was calculated starting from the native and derivative data sets.

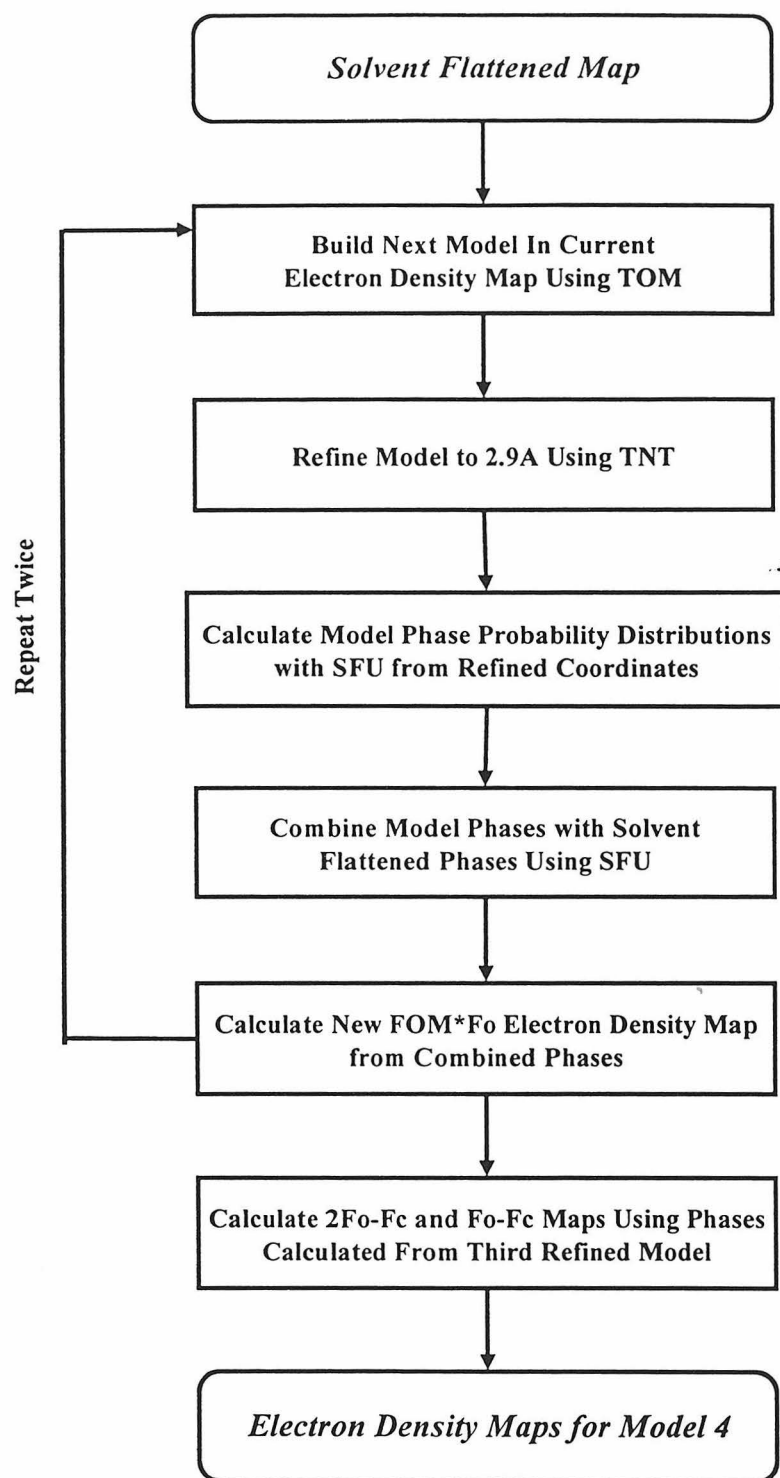


Figure 2.10 (b) *How the first three models were built and refined starting from the solvent flattened map and resulting in the electron density map into which model 4 was built.*

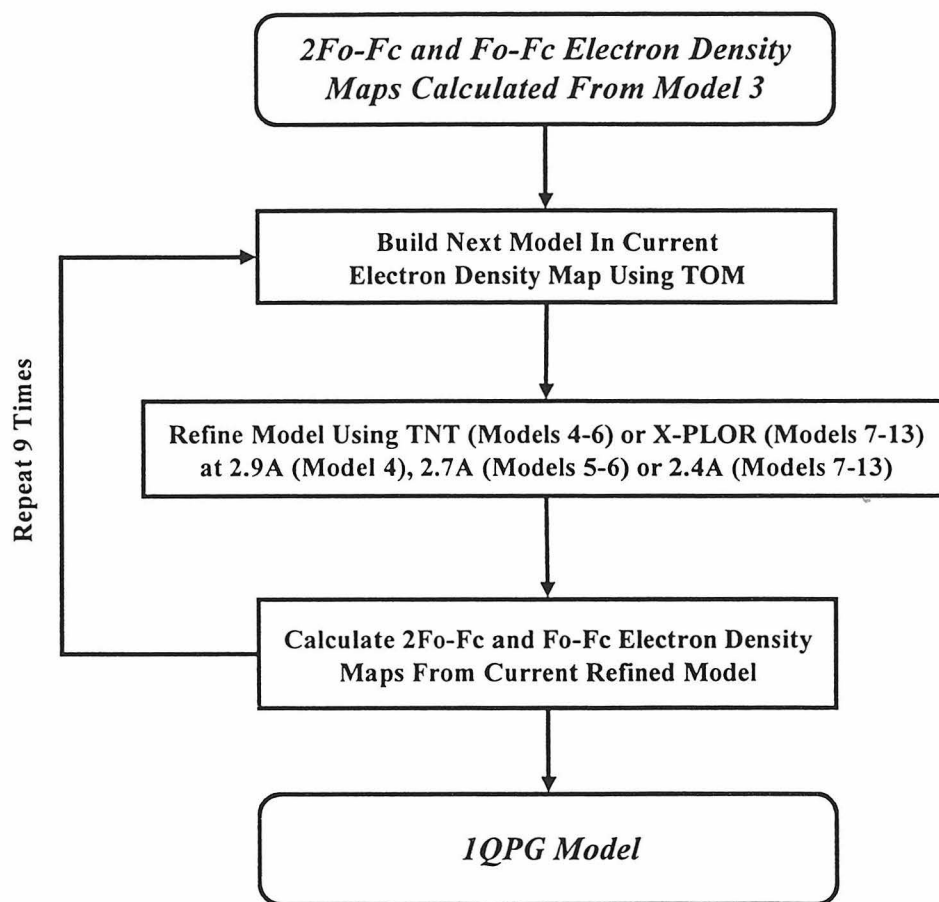


Figure 2.10 (c) How model building and refinement proceeded from the map into which model 4 was built up to the final 1QPG coordinates.

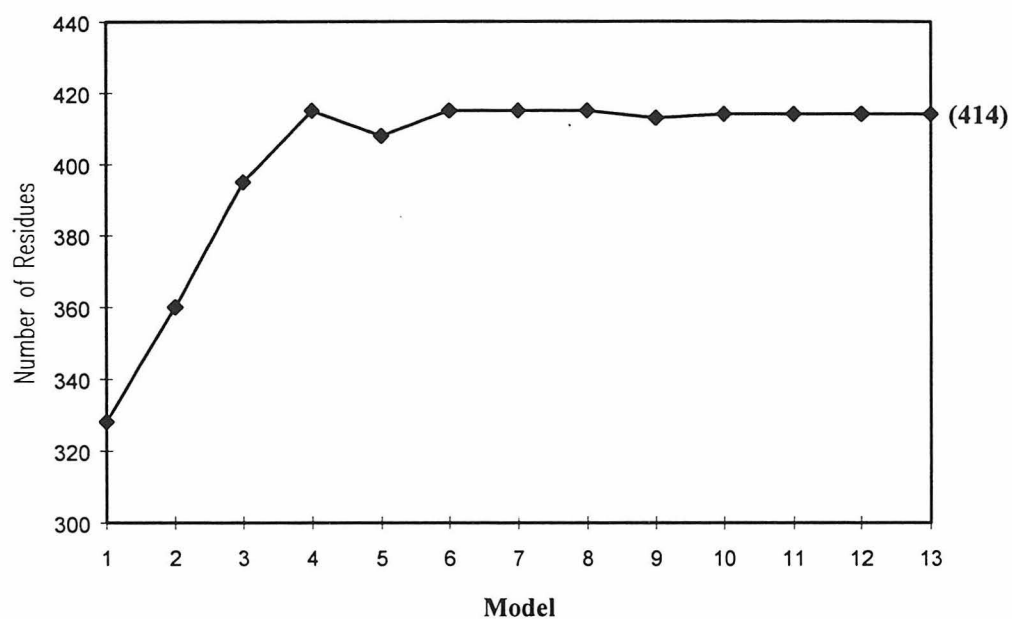


Figure 2.11 *Characteristics of the various models. (a) Residues included in each model. The entire sequence of yeast PGK consists of 415 residues.*

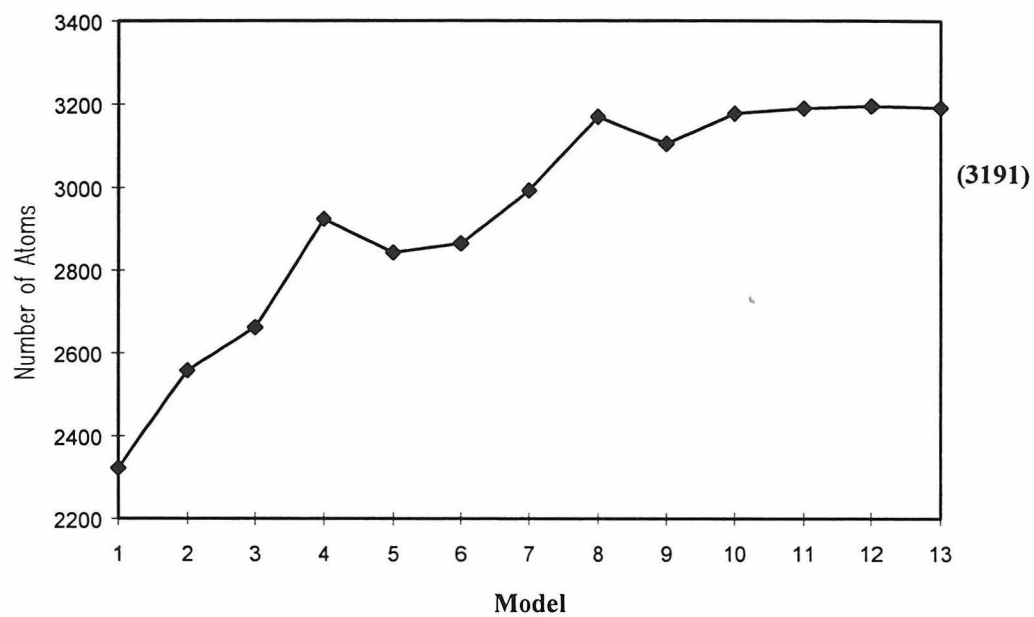


Figure 2.11 (b) *Number of atoms included in each model.*

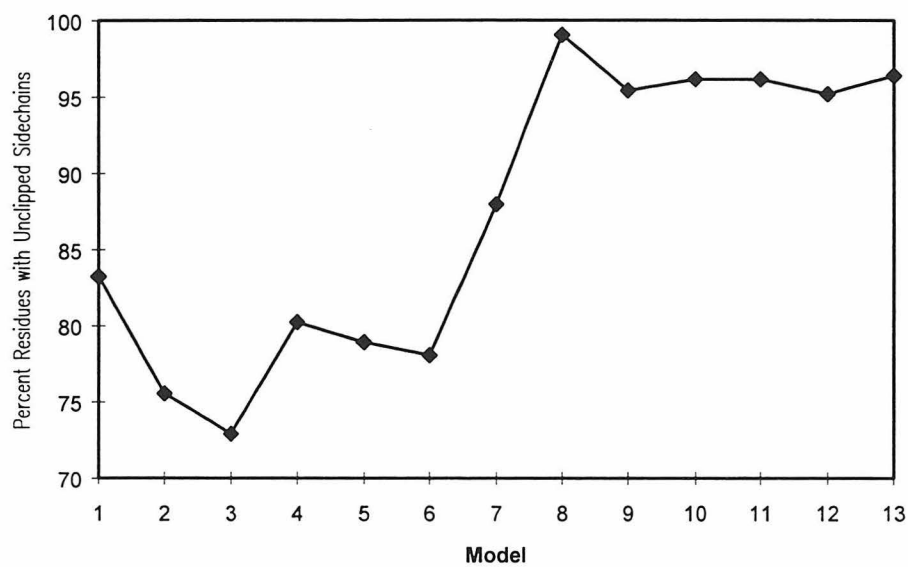


Figure 2.11 (c) *Percentage of non-alanine, non-glycine residues for which all side-chain atoms were included.*

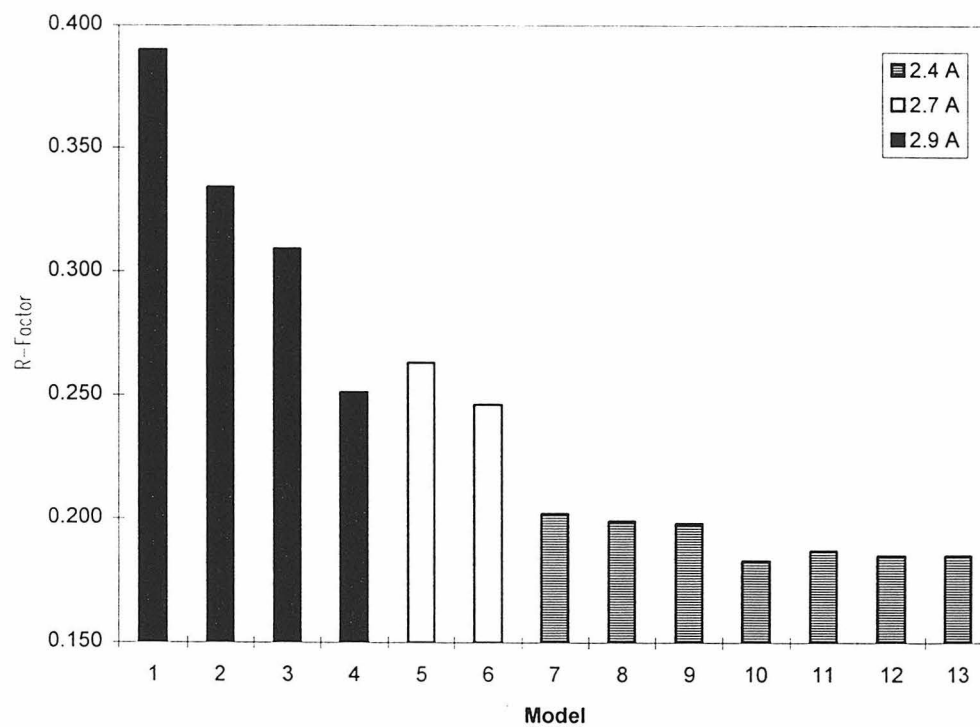
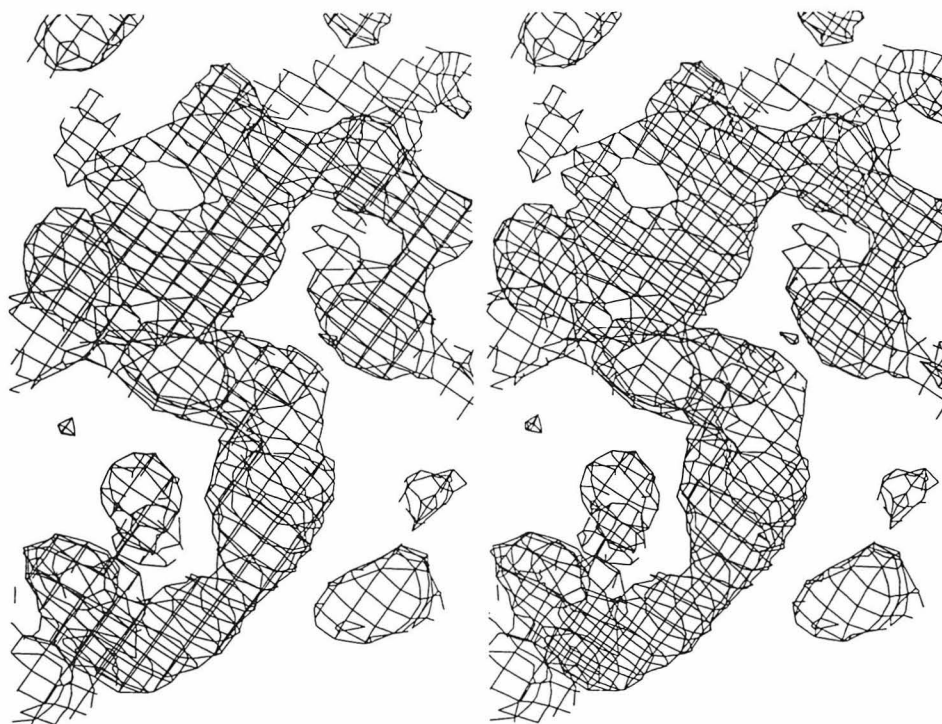


Figure 2.11 (d) *R-factor resulting from refinement of each model. Shading of the bars indicates maximum resolution for refinement.*

Figure 2.12 *The 10 panels on the next 5 pages show electron density calculated from 5 phase sets defined in Table 2.2: FLAT (panels a-b), Model_1*Flat (panels c-d), Model_2*FLAT (panels e-f), Model_3*FLAT (panels g-h), and Model_13 (panels i-j). The same region of the map is shown for each phase set. Each view of the density is shown twice per page. The upper panel shows the density without any model built in. The lower panel shows the same density with the final 1QPG coordinates included. Note that residues 95-103 were omitted in models 1 and 2, while residues 118-133 were omitted in models 1,2, and 3. Thus, the improvements in the density in successive figures are due to improvements in the rest of the model and demonstrate the effect of phase combination in the early stages of model building and refinement.*

a.



b.

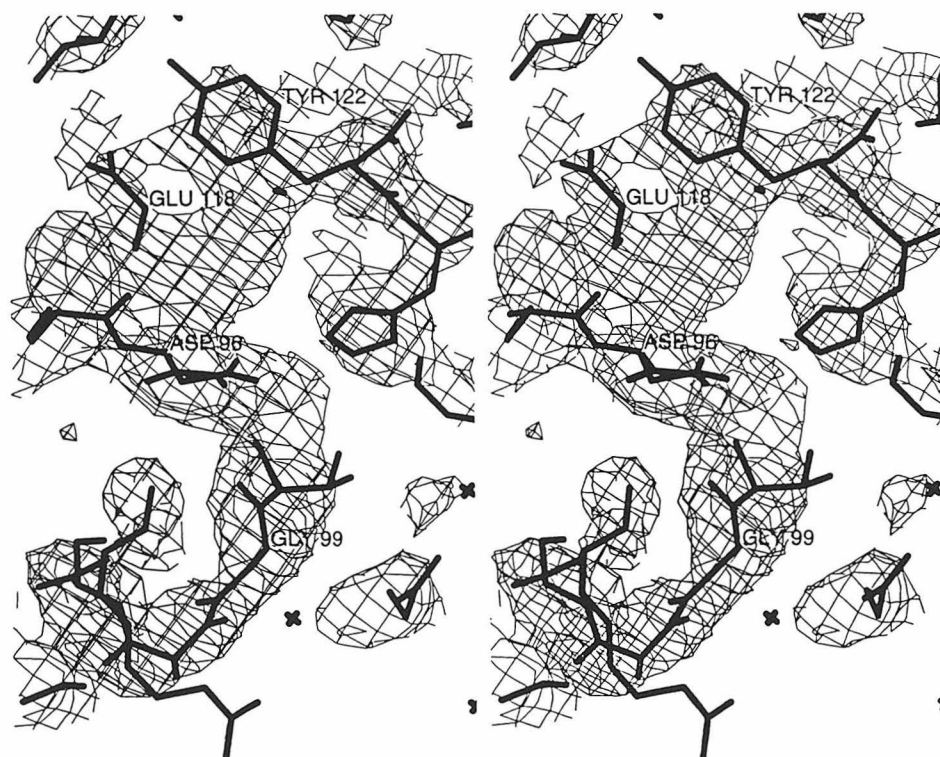
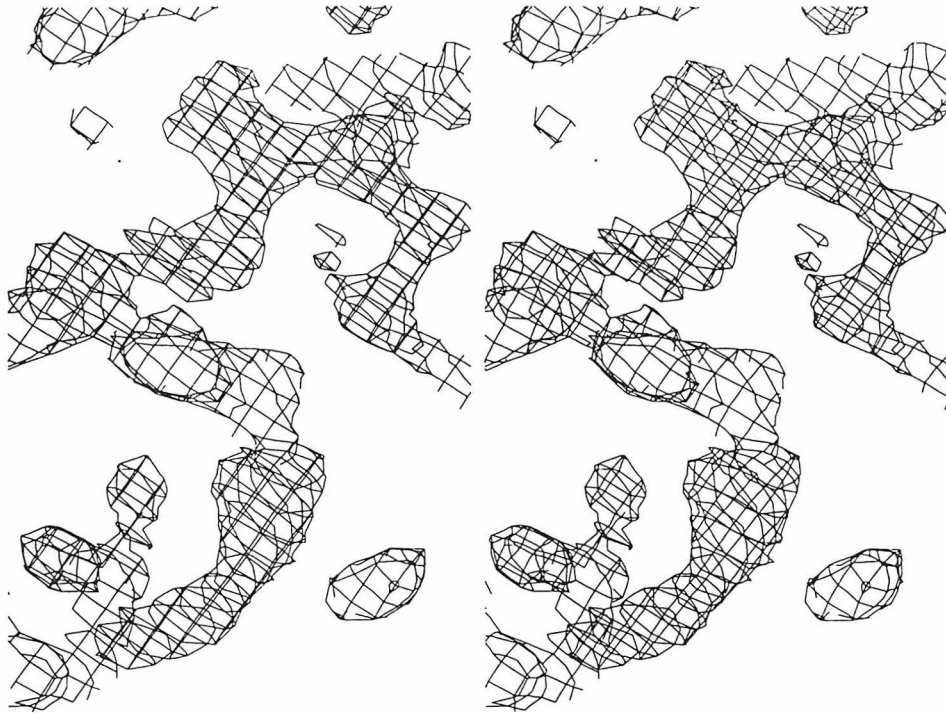


Figure 2.12 a-b *FLAT-phased map. No model phases were included in the structure factor calculations in this case.*

c.



d.

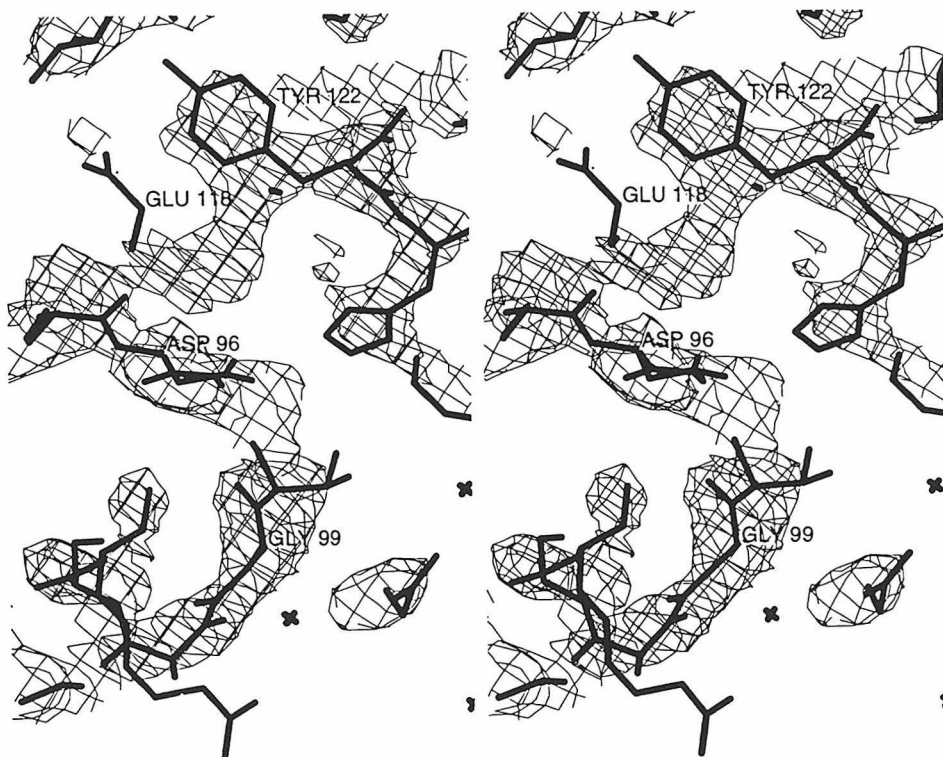


Figure 2.12 c-d *Model_1*FLAT-phased map.* Since none of the labeled residues were included in the model, improvements in the map over the FLAT-phased map (panels a-b) are due to phase combination of the partial model with the FLAT phases.

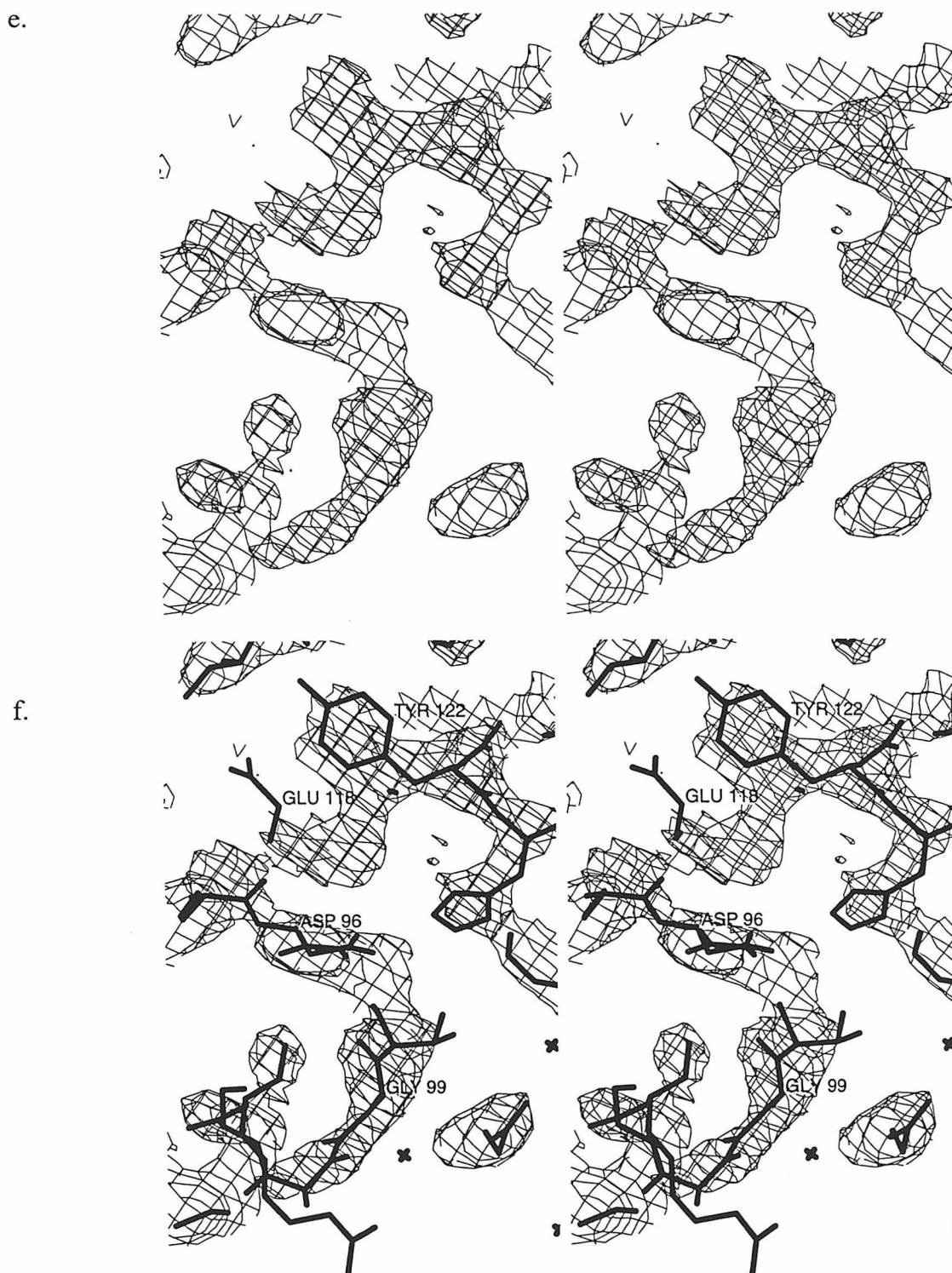
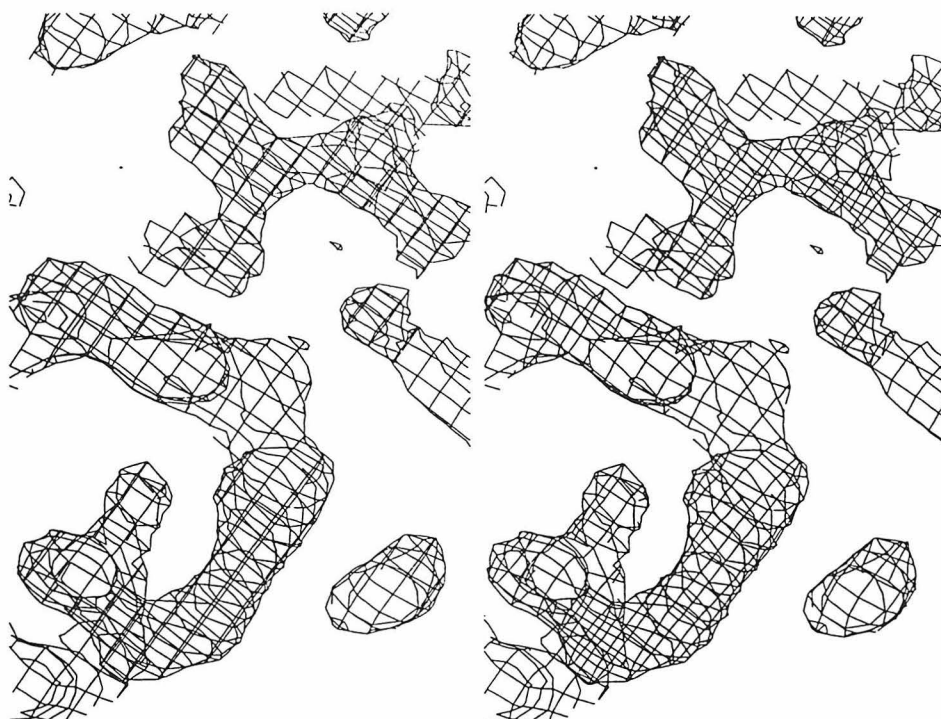


Figure 2.12 e-f *Model_2*FLAT-phased map.* Since none of the labeled residues were included in the model, improvements in the map over the FLAT-phased map (panels a-b) are due to phase combination of the partial model with the FLAT phases. Improvements over the *Model_1*FLAT-phased map* (panels c-d) are due to improvements elsewhere in the model.

g.



h.

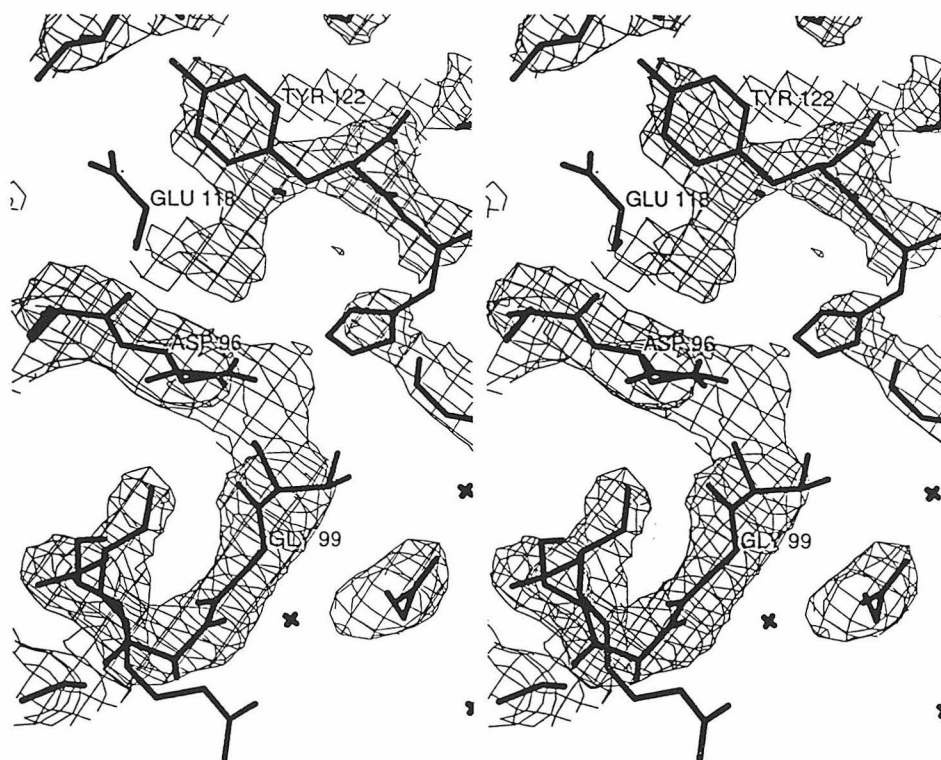
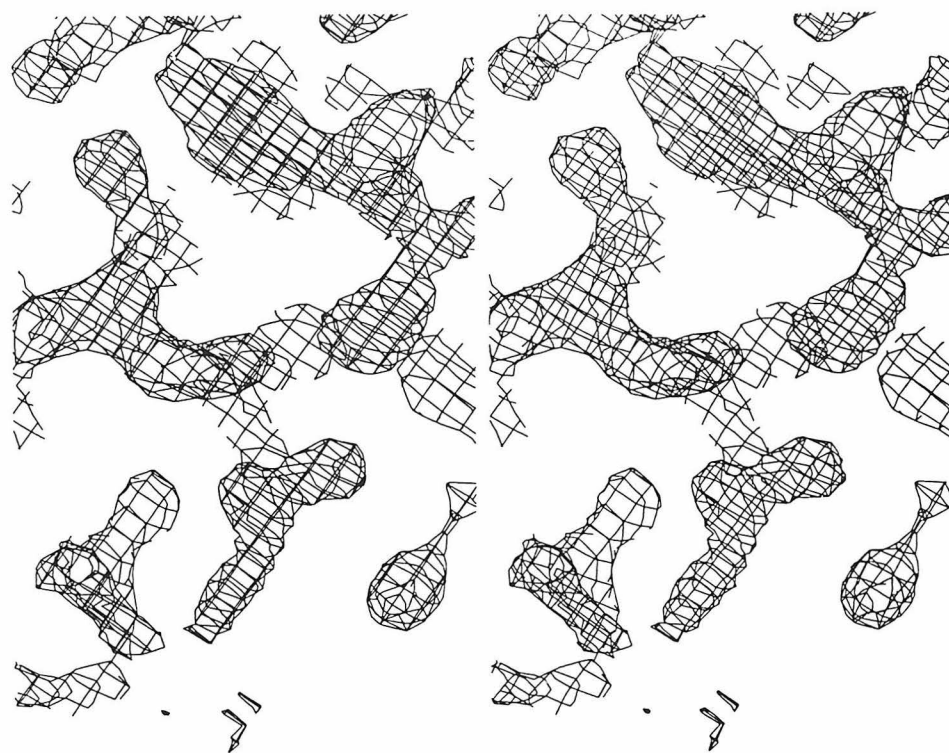


Figure 2.12 g-h *Model_3*FLAT-phased map.* Since residues 118-133 were not included in the model, improvements in the top half of the map over the FLAT-phased map (panels a-b) are due to phase combination of the partial model with the FLAT phases. Improvements over the Model_2*FLAT-phased map (panels e-f) are due to improvements elsewhere in the model.

i.



j.

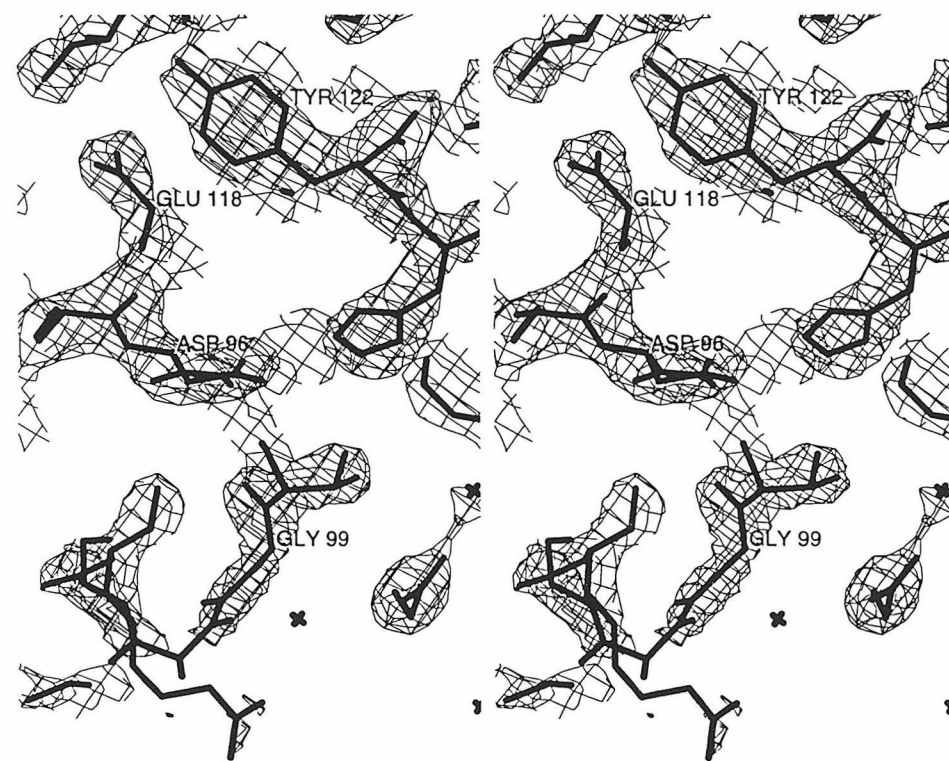


Figure 2.12 i-j *Model_13-phased map.*

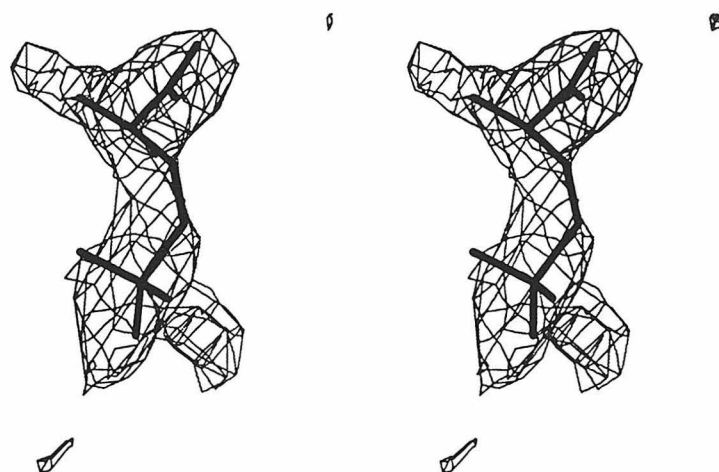


Figure 2.13 Stereoview of omit-refine difference density for 3-PG bound to the basic patch on the N-domain of yeast R65Q PGK. The contour level shown is 2.0σ . The extra density adjacent to the phosphate group may arise from a population of 3-PG molecules bound with the phosphate and carboxyl groups reversed.

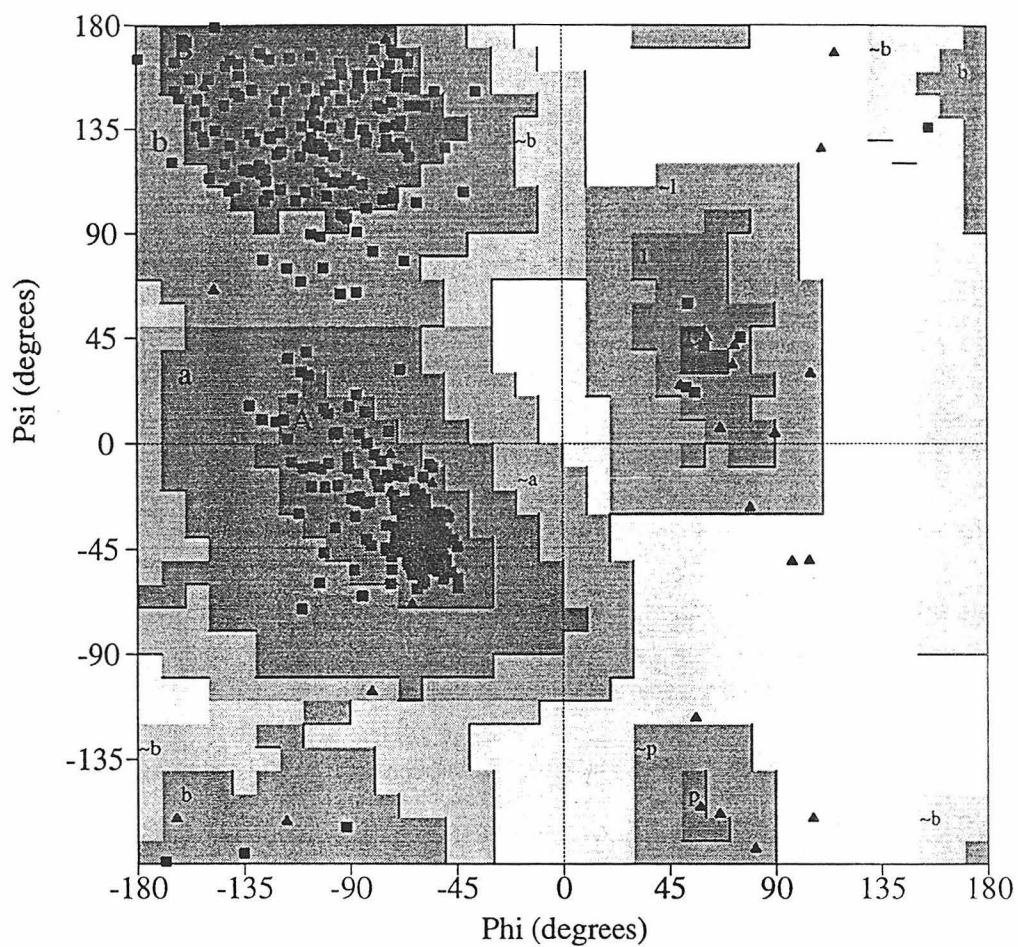


Figure 2.14 Ramachandran diagram for the final refined coordinates. Glycines are shown as triangles, non-glycine residues as squares. Drawn with PROCHECK (Laskowski et al., 1993).

Chapter 3

Structure of the R65Q Mutant of Yeast Phosphoglycerate Kinase Complexed with Adenylylimido-diphosphate and 3-Phosphoglycerate*

The crystal structure of the R65Q mutant of yeast PGK exhibits substantial variations from the previously reported wild-type yeast structure, PDB entry 3PGK (rms deviations between equivalent α -carbons in the individual domains of 2.9-3.2 Å). However, the R65Q structure is very similar to structures of PGK isolated from horse, pig, and *Bacillus stearothermophilus* (rms deviations between equivalent α -carbons in the individual domains < 1.0 Å), with the most significant differences occurring in the relative orientations of the N- and C-domains. It appears that the R65Q mutation has negligible effect on the tertiary structure of the individual domains and that the 1QPG coordinates are a more reliable model of yeast PGK than the 3PGK coordinates. In addition, the current structure represents the first view of a ternary complex of PGK with both substrates. The 3-PG is located in a 'basic patch' of residues on the N-terminal domain, while the Mg-AMP-PNP interacts with residues on the C-terminal domain. The overall structure is an 'open' configuration with the two substrates separated by ~11 Å across the interdomain cleft. The structure shows that the high

* Parts of this chapter have been published in McPhillips, T.M., Hsu, B.T., Sherman, M.A., Mas, M.T., Rees, D.C. (1996) Structure of the R65Q Mutant of Yeast 3-Phosphoglycerate Kinase Complexed with Mg-AMP-PNP and 3-Phospho-D-Glycerate. *Biochemistry* **35**:4118-4127.

activity of the R65Q mutant in the absence of activating anions cannot be due to the mutant enzyme assuming a partially closed conformation.

A. Conformation of the Polypeptide Chain

A.1. Yeast R65Q PGK

The single polypeptide chain of yeast R65Q PGK is organized into two domains, the N-domain and C-domain, consisting of residues 1-184 and 200-393, respectively. The remaining residues of the 415 amino acids in the sequence constitute the interdomain linkage. Each domain folds as an α/β , twisted open-sheet structure composed of a single sheet of six parallel β -strands sandwiched between four long α -helices, two on each side. In addition, each domain has a pair of antiparallel β -strands located above the C-termini of the parallel β -strands. Both pairs of antiparallel β -strands end in solvent-exposed loops. A ribbon diagram of the structure is given in Figure 3.1.

The secondary structural elements are listed in Table 3.1, labeled according to the equine PGK nomenclature (Banks et al., 1979). The elements are also labeled in Figure 3.1b. β -strands A-F form the core of the N-domain in the order CDBAEF (counting towards the C-domain). This sheet is shielded by helices 1 and 2 on one side, and helices 3 and 4 on the other. The core of the C-domain is made up of β -strands G-L in the order IHGJKL (counting towards the N-domain). Helices 8 and 10 shield one side of this sheet, while helices 11 and 12 border the opposite face. In both domains, the final β -strand is the one nearest the opposite domain. Each precedes a helix that crosses the interdomain space: helix 7 follows strand F, and helix 14 follows strand L.

There are five α -helices in addition to those previously described. The short helix 15 follows helix 14 and associates with the N-domain next to the very short helical element 6. Helix 5 consists of a single turn preceding helix 6 and forms an integral part of the interdomain junction. Elements 9 and 13 are helices of medium length in the C-domain perpendicular to the direction of the core strands. Helix 9 is on a skew line parallel to the β -sheet. The N-terminus of this helix forms part of the hydrophobic pocket in which the nucleotide binds. Helix 13 is normal to the β -sheet, with the N-terminus near the triphosphate group of the nucleotide.

A.2. Differences with the 3PGK Model

As alluded to in Chapter 2, comparison of the present model of yeast PGK, PDB (Bernstein et al., 1977) entry 1QPG, with PDB entry 3PGK (Watson et al., 1982) revealed significant differences. While the overall topology of the two yeast structures is nearly the same, least-squares superposition of the α -carbon positions in the two structures yields an rms deviation of 3.4 Å. Superposition of the N- and C-domains separately provides comparable values (2.9 Å for residues 1-185, and 3.2 Å for residues 199-393), indicating that the differences between the two structures reflect differences within the domains, not in the relative orientation of the domains. In particular, the β -strands in the core of the C-domain in our model are much less curved along their lengths than in 3PGK, and this behavior gives rise to most of the other differences in the C-domain. For instance, the single helix 9 in our model replaces helices VII and VIII of the 3PGK C-domain. (See Figures 2.8 and 2.9 in Chapter 2 for more details.)

Many of the discrepancies between the two yeast PGK structures are likely to be artifactual, however, and not a reflection of genuine biochemical differences or due to different crystallization conditions. The 3PGK model is one of the least reliable structures in the PDB (Morris et al., 1992). According to the header of the PDB file itself, the coordinates are unrefined, the result of 'normalizing' a model built into an experimental electron density map. Geometric analysis by PROCHECK (Laskowski et al., 1993) indicates numerous bad contacts (55 per 100 residues) and poor average side-chain parameters (rms deviation of 30° from ideal χ_1 values). Only 60% of the non-glycine, non-proline residues fall within the most favored region of the Ramachandran plot; 9% of the residues are in disallowed regions (see Figure 3.2). Finally, two segments of the 3PGK polypeptide have the wrong sequence. Residues 259 to 264 have the sequence VGPEIA in 3PGK; the actual sequence for yeast PGK is GAEIVP. Residues 329 to 332 in 3PGK have the sequence TVIL; the actual sequence is KTIV (Watson et al., 1982).

The apparent problems with the 3PGK coordinates potentially give the R65Q model, PDB entry 1QPG, special significance. All site-directed mutagenesis studies and most spectroscopic investigations of PGK have been carried out on the yeast enzyme (Joao & Williams, 1993). Up till now, the only model for the wild-type yeast structure was the 3PGK coordinates, and interpretation of these experiments were in light of this structure. Assuming the R65Q mutation does not significantly effect the tertiary structure of the enzyme, the 1QPG coordinates would be the only accurate representation of the structure of yeast PGK currently available. In this case, not only

would all future studies of the yeast enzyme need to be interpreted in terms of this structure, but past studies might benefit from reinterpretation in this light. The question then becomes, how can we tell if the 1QPG coordinates are representative of wild-type PGK?

A.3. Comparison with Equine, Porcine, and Bacterial PGK

Since the yeast R65Q structure was solved, the coordinates have become available for the structures of PGK from three other species: horse (Blake & Rice, 1981), pig (Harlos et al., 1992), and *B. stearothermophilus* (Davies et al., 1994). The R65Q yeast structure is very similar to these structures. Table 3.2 indicates the results of least-squares superpositions of the R65Q PGK model onto equine, porcine, and bacterial PGK. Calculations based on single domains resulted in each case in an rms deviation of less than 1 Å, with an average rmsd of 0.87 Å for the six single domain comparisons. The three rms deviations resulting from whole-molecule comparisons are also significantly lower (the largest is 2.0 Å) than the whole-molecule rmsd value for 3PGK (3.4 Å).

Chothia and Lesk (1986) have noted that the expected rmsd between two homologous structures can be estimated to within ~20% using the equation

$$D = 0.40 e^{1.87H}$$

where H is the fraction of the sequence different between the two homologous structures, and D is the predicted rms deviation in mainchain atom positions in Ångstroms (Chothia & Lesk, 1986). Estimates for the deviations between yeast PGK

and PGK from pig, horse, and bacteria are included in Table 3.2. The observed rmsd values for each single-domain superposition average only 4% higher than the predicted values and differ no more than $\pm 22\%$ from them. Thus, the structure of each domain of yeast R65Q PGK appears to be equivalent to that observed in each of the other species. Consequently, it may be assumed that the R65Q mutation does not disrupt the structure of the individual domains. Within these domains, at least, the 1QPG coordinates should be considered a reliable representation of wild-type yeast PGK. The question that remains is, how significant is the observed domain orientation?

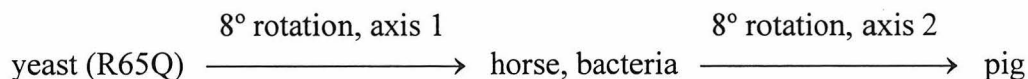
A.4. Orientation of the Domains in Different PGK Structures

In contrast to the individual domains, Table 3.2 shows that the observed rms deviations resulting from whole-molecule comparisons are larger than predicted from sequence differences, with bacterial PGK having the smallest deviation and porcine PGK having the largest deviation from yeast PGK. Since the individual domains superimpose well, differences in the relative orientation of the domains must be responsible for the larger sizes of the whole-molecule rms deviations. Larger deviations indicate correspondingly greater differences in relative orientation. Thus, the rmsd values suggest that the relative domain orientation of the yeast R65Q structure is most like that of *B. stearothermophilus* PGK and least like that of porcine PGK.

Further analysis of the relative orientation of the domains of PGK in the yeast R65Q, equine, porcine, and bacterial PGK models clarifies the relationships between the various structures. Table 3.3 shows the result of comparing the relative orientation of

the N-domain and C-domain in the yeast R65Q structure with the orientations observed in structures of PGK from horse (Blake & Rice, 1981), pig (Harlos et al., 1992), and *B. stearothermophilis* (Davies et al., 1994). The comparison was carried out in the following way. First, the four structures were superimposed (Kabsch, 1976) using only atoms from the N-domains in the superposition algorithm. This overlaid the N-domains on each other but left the C-domains askew. Next, each C-domain was transformed to superimpose on the C-domains of the other structures. Finally, each of these transformations was resolved into rotational and translational components. The transformations are listed in Table 3.3.

The rotational components of the transformations listed in Table 3.3 may be summarized as follows:



The porcine and yeast PGK structures represent the most divergent conformations of this set of structures, being related by a 15° rotation (see Table 3.3). The horse and bacterial PGK structures represent an intermediate conformation between the pig and yeast R65Q PGK structures. Since the relative positions of the two domains are similar in the horse and bacterial structures, transformations involving the horse and bacterial C-domains are also similar. The average rotation axis relating all three conformations of PGK is shown in Figure 3.3. The significance of these observations with respect to the function of PGK will be discussed in Chapter 4.

B. Binding Mode of AMP-PNP

B.1. Binding site and conformation of AMP-PNP

The residues following the C-termini of the five core β -strands in the C-domain form a hydrophobic pocket on the surface of the protein facing the opposite domain. The pocket is bordered by residues 210-220, 235-239, 334-341, 370-373, and 393-397, following β -strands G, H, J, K, and L, respectively. Residues 253, 254, 310, and 311 also contribute to the surface of the pocket. A molecule of AMP-PNP is bound in this pocket, stretched out along the top of the sheet. The nucleotide is oriented with the adenine ring distal to the N-domain and the triphosphate moiety adjacent to the interdomain cleft. Figure 3.4 shows the nucleotide and the residues that directly interact with it. Pertinent hydrogen bond distance information is given in Table 3.4.

Figure 3.5 shows the nucleotide alone with atoms labeled. The conformation of the nucleotide is currently modeled with a C2'-*endo* ribose group *anti* to the adenine ring. In the 1QPG model, the Mg^{2+} ion is coordinated by a single, non-bridging oxygen atom from each phosphate group. The phosphate oxygen ligands occupy *cis* coordination sites of a distorted octahedron. This α,β,γ -tridentate coordination of a divalent metal ion by a triphosphate group has been observed previously in the crystal structures of the magnesium and manganese complexes of ATP (Cini et al., 1984; Sabat et al., 1985). Unlike the small molecule structures of metal-ATP complexes, however, the coordination of the ion in the 1QPG model is unequal. The α -phosphate refines to a position 2.19 Å from the metal ion. This is consistent with the 2.0 to 2.2 Å coordination distances observed in the small molecule structures (Cini et al., 1984). In contrast, the

coordination distances corresponding to the β and γ phosphates are 2.46 Å and 2.41 Å, respectively, or significantly longer than the small-molecule distances. Figure 3.6 shows the final $2F_o - F_c$ density for the triphosphate group and Mg^{2+} ion.

B.2. Specific Interactions Between AMP-PNP and the Protein

In addition to Van der Waals interactions provided by the residues lining the nucleotide binding pocket, the protein makes several hydrogen bonding interactions with the AMP-PNP molecule. The adenine ring makes two such interactions with the protein: a weak hydrogen bond between atom N6 and the carbonyl oxygen of glycine-310 (3.31 Å) and a water-mediated interaction between N6 and the carbonyl oxygen of proline-337 through water-545. Figure 3.7 illustrates both of these sets of hydrogen bonds. The primary interaction between the protein and the sugar moiety involves the sidechain of glutamate-341 which accepts hydrogen bonds from atoms O2' and O3' of the sugar (see Figure 3.8).

The primary interaction between the triphosphate moiety of the nucleotide and the protein is with the N-terminus of helix 13. Figure 3.9 shows these interactions in detail. The amide groups of the first three residues of this helix donate hydrogen bonds to three oxygen atoms of the phosphate groups. The glycine-371 amide donates a hydrogen bond primarily to O2G from the γ -phosphate. The bond may be bifurcated, with the O1B atom of the β -phosphate accepting part of it. The O1B atom also accepts a full hydrogen bond from the aspartate-372 backbone amide. Finally, the O2B atom of the β -phosphate accepts two hydrogen bonds, one from the threonine-373 amide, and

another from the threonine-373 hydroxyl group. Both glycine-371 and aspartate-372 are conserved in over 20 known species, while position 373 is either a threonine or a serine in the yeast, horse, pig, and *B. stearothermophilus* sequences.

The triphosphate group and the protein also interact via the sidechain of asparagine-334, which is conserved in the four sequences just mentioned. Figure 3.10 shows that the N3B atom, the non-hydrolyzable link between the β - and γ -phosphate groups of the AMP-PNP molecule, donates a hydrogen bond to the OD1 atom of the asparagine, while the O3G atom of the γ -phosphate accepts a long hydrogen bond from the ND2 atom of the same amino acid.

In total, the β - and γ -phosphate groups are involved in seven hydrogen bonds with the protein, while the α -phosphate has no direct interactions with the protein whatsoever. This is consistent with the observation made above, that the Mg^{2+} is coordinated most strongly by the α -phosphate which accepts no protein hydrogen bonds, and least strongly by the β -phosphate which accepts as many as four. The combined steric, electrostatic, and electronic effects of the triphosphate-protein interactions appear to determine the degree of coordination to the metal ion in this complex.

It is apparent in Figures 3.4 and 3.9 that the sidechain of aspartate-372 is located near the magnesium ion. The distance between the carboxylate OD2 atom and the metal is 3.65 Å, indicating an electrostatic interaction but no actual coordination bond. It is clear, however, that a slight change in the protein or nucleotide conformations could easily lead to full coordination of the metal ion by the aspartate, *trans* to the γ -phosphate

oxygen. Aspartate-372 is completely conserved in all available sequences of PGK and mutation to asparagine results in a 90% decrease in enzymatic activity (Minard et al., 1990). Thus, coordination of the Mg^{2+} by this residue may play a role in phosphoryl transfer.

C. Binding Mode of 3-PG

C.1. Basic Patch Residues

Just as the residues following the five core β -strands of the C-domain form the nucleotide binding pocket, the residues following four of the core β -strands of the N-domain form the anion binding pocket, or basic patch. Residues 21-25, 62-67, and 121 are on turns following β -strands A, B, and D, respectively, and form one side of the pocket. Residues 162-171 follow strand E, include helix 5, and make up the opposite face of the pocket. Charged and polar side chains in the basic patch are contributed by arginines 21, 38, 121, and 168; histidines 62, 167, and 170; asparagines 25 and 67; glutamine-65; threonine-165; and aspartate-23. Electron density for a 3-PG molecule occupies the center of the basic patch. Figure 3.11 shows the refined position of the 3-PG and the residues that interact with it in the basic patch. Figure 3.12 shows the 3-PG alone with all atoms labeled.

C.2. Interactions between 3-PG and the Protein

Figure 3.13 indicates the distances between the 3-PG molecule and neighboring protein atoms. The 3-PG carboxyl group is stabilized by electrostatic interactions with

the sidechains of asparagine-25, arginine-38, and histidine-62. The carboxyl oxygens (O1 and O2) are not in position to hydrogen bond to either asparagine-25 or histidine-62. However, a hydrogen bond is possible between O2 and arginine-38 (3.38 Å). The 3-PG hydroxyl group (O3) interacts favorably with side-chains of asparagine-25 and arginine-38, both of which may participate in long hydrogen bonds to the group. Finally, the 3-PG phosphate is stabilized by histidine-62 and arginine-121. The latter appears to weakly donate at least one hydrogen bond to the O2P atom of the phosphate group.

Electrostatic attraction appears to play a significant role in the binding of 3-PG to the basic patch in this structure. There are only three hydrogen bond distances that are less than 3.5 Å. In contrast, there are four residues that certainly have a +1 charge (the arginines), three that may have a +1 charge (the three histidines), but only one residue that is likely to have -1 charge (the aspartic acid). Thus, with no anion bound, the basic patch has a net charge between +3 and +6. The charge of 3-PG at neutral pH is -3, giving the occupied basic patch a net charge between 0 and +3. Coulombic attraction between the 3-PG and the basic patch is therefore likely to be strong.

C.3. Comparison with the Porcine Structure

Comparison of the interactions between yeast R65Q PGK and 3-PG and those between wild-type porcine PGK (a 2.0 Å structure) and 3-PG reveals interesting differences. First, the conformation of 3-PG in the basic patch differs between the two structures. The conformations are related by a 180° torsion around the C2-C3 bond that

swaps the positions of the hydroxyl and carboxyl groups. More interesting are the differences in the types of interactions stabilizing the complexes. In the yeast R65Q case, the interactions are primarily electrostatic with perhaps six mostly weak (i.e., longer than 3.2 Å) hydrogen bonds between the 3-PG and the protein. However, in the porcine case, there are as many as nine strong (i.e., shorter than 3.2 Å) hydrogen bonds between the substrate and the protein (Harlos et al., 1992). The porcine analogs of arginines 65, 121, and 168 each donate two hydrogen bonds, and histidine-62 donates one hydrogen bond to the phosphate group of the 3-PG. In addition, the 3-PG hydroxyl group participates in hydrogen bonds with both asparagine-25 and aspartate-23. Thus, the hydrogen bonds between the 3-PG and the protein in the porcine PGK structure contribute much more to the stability of the complex.

There are several possible explanations for the difference in 3-PG binding modes in these two structures. The most obvious is that the R65Q mutation removes one of the arginines that binds strongly to the 3-PG phosphate group in the porcine structure. This could account for the partial occupancy of the site in the mutant structure and could result in a loss of orientational specificity leading to the ambiguity described in the next section. The other explanation is that the geometry of the basic patch is somewhat different in the two structures. In particular, helix 5 is shifted in the porcine structure, bringing arginine-168 closer to the bound 3-PG. This shift may be related to the difference in interdomain orientation between the two structures. These possibilities and their implications for the function and activation of PGK will be discussed fully in Chapter 4.

C.4. Alternative Orientations of 3-PG

Although the general size and shape of the density is appropriate for 3-phosphoglycerate, determination of the precise orientation of 3-PG in this site was problematic. In addition to the orientation selected for the final refinement of the 1QPG model, two other orientations were considered. Figure 3.14 compares these two alternative orientations (panels b and c) with the 1QPG position (panel a). The electron density shown is the same omit-refine density in each case. In none of the orientations examined does the 3-PG bind as strongly to the basic patch residues as observed in the porcine structure. The electron density for the substrate in the R65Q yeast structure is simply further away from the sidechains.

The first alternate orientation (Figure 3.14b) resulted from refining the 3-PG orientation observed in the porcine PGK structure against the yeast R65Q data. As described above, this 3-PG conformation is related to the 1QPG orientation by a 180° torsion around the C2-C3 bond. The figure shows that the hydroxyl group cannot account for all of the omit-map density at that position, while the carboxyl group suffers from too little density. Moreover, the refinement resulted in the inversion of the C2 chiral center to reflect the chirality of the observed density. Based on the experimental data, this test orientation for the 3-PG appears less plausible than the 1QPG orientation. However, a partial occupancy of the basic patch by 3-PG in this orientation cannot be ruled out.

The second alternate orientation (Figure 3.14c) resulted from refining a 3-PG molecule that had been rotated 180° around the C3 atom, swapping the carboxyl and phosphate groups. Although the carboxyl and phosphate groups fit the omit-map density no worse than the 1QPG orientation, the hydroxyl group protrudes from the density. Furthermore, no atoms occupy the density occupied by the hydroxyl group in the 1QPG model. Again, the 1QPG orientation appears superior, though a fraction of the 3-PG molecules in the crystal could be in the second alternate orientation.

D. Conclusion

Considering the kinetic properties of the R65Q enzyme, its structure has significant implications for the function of PGK. Of the horse, pig, and bacterial structures, the N- and C-domains of PGK are closest together in the porcine model (Harlos et al., 1992). Consequently, porcine PGK can be described as being partially closed. In contrast, the bacterial model is clearly open (Davies et al., 1993). Because the yeast R65Q model is least similar to the porcine structure and most similar to the bacterial structure, the yeast R65Q PGK should also be described as being open. Measuring the distance between substrates in this structure confirms this conjecture: the 3-PG and AMP-PNP molecules are separated by over 11 Å.

Since the R65Q PGK is relatively open, its high activity cannot be due to the enzyme being in a permanently closed conformation. Since the mutant PGK is no closer to a closed conformation than the wild-type structures of PGK, the hypothesis that the high activity of the mutant is due to the enzyme being closer in conformation to

a 'closed' form is inconsistent with the crystal structure. In fact, the R65Q mutation may cause the enzyme to adopt a more open conformation than the wild-type ternary complex. The structure of wild-type porcine PGK complexed with both 3-PG and Mn-AMP-PNP has been reported recently (May et al., 1996). Although the coordinates for this structure have not been released, the authors claim that the domain orientation observed is similar to that seen in the binary complex of porcine PGK with 3-PG, the structure compared with the R65Q model in Table 3.3. The implications of these observations for the hinge-bending model will be discussed fully in Chapter 4.

References

- Banks, R.D., Blake, C.C.F., Evans, P.R., Haser, R., Rice, D.W., Hardy, G.W., Merrett, M., Phillips, A.W. (1979) Sequence, structure and activity of phosphoglycerate kinase: a possible hinge-bending enzyme. *Nature* **279**:773-777.
- Bernstein, B.E., Michels, P.A.M., Hol, W.G.J. (1997) Synergistic effects of substrate-induced conformational changes in phosphoglycerate kinase activation. *Nature* **385**:275-278.
- Bernstein, F.C., Tasumi, M. 1977. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Appl. Cryst.* **26**: 282-291.
- Blake, C.C.F., Rice, D.W. (1981) Phosphoglycerate kinase. *Phil. Trans. R. Soc. Lond. A* **293**: 93-104.
- Chothia, C., Lesk, A.M. (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J.* **5**:823-826.
- Cini, R., Burla, M.C., Nunzi, A., Polidori, G.P., Zanazzi, P.F. (1984) *J. Chem. Soc. Dalton Trans.* 2467-2476.
- Davies, G.J., Gamblin, S.J., Littlechild, J.A., Dauter, Z., Wilson, K.S., Watson, H.C. (1994) Structure of the ADP complex of the 3-phosphoglycerate kinase from *Bacillus stearothermophilus* at 1.65Å. *Acta Cryst.* **D50**:202-209.
- Harlos, K., Vas, M., Blake, C.F. (1992) Crystal structure of the binary complex of pig muscle phosphoglycerate kinase and its substrate 3-phospho-D-glycerate. *Proteins Struct. Func. Genet.* **12**:133-144.
- Joao, H.C., Williams, R.J.P. (1993) The anatomy of a kinase and the control of phosphate transfer. *Eur. J. Biochem.* **216**:1-18.
- Kabsch, W. (1976) A solution for the best rotation to relate two sets of vectors. *Acta Cryst.* **A32**:922-923.
- Kraulis, P.J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**:946-950.
- Laskowski, R.A., McArthur, M.W., Moss, D.S., Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**:282-291.
- May, A., Vas, M., Harlos, K., Blake, C. (1996) 2.0 Å resolution structure of a ternary complex of pig muscle phosphoglycerate kinase containing 3-phospho-D-glycerate

and the nucleotide Mn adenylylimidodiphosphate. *Proteins Struct. Func. Genet.* **24**:292-303.

Minard, P., Bowen, D.J., Hall, L., Littlechild, J.A., Watson, H.C. (1990) Site-directed mutagenesis of aspartic-acid 372 at the ATP binding-site of yeast phosphoglycerate kinase—over-expression and characterization of the mutant enzyme. *Protein Eng.* **3**:515-521.

Morris, A.L., MacArthur, M.W., Hutchinson, E.G., Thornton, J.M. 1992. Stereochemical quality of protein structure coordinates. *Proteins* **12**:345-364.

Sabat, M., Cini, R., Haromy, T., Sundaralingam, M. (1985) Crystal structure of the α,β,γ -tridentate manganese complex of adenosine 5'-triphosphate cocrystallized with 2,2'-dipyridylamine. *Biochemistry* **24**:7827-7833.

Watson, H.C., Walker, N.P.C., Shaw, P.J., Bryant, T.N., Wendell, P.L., Fothergill, L.A., Perkins, R.E., Conroy, S.C., Dobson, M.J., Tuite, M.F., Kingsman, A.J., Kingsman, S.M. (1982) Sequence and structure of yeast phosphoglycerate kinase. *EMBO J.* **1**:1635-1640.

Table 3.1 *Constituent residues of secondary elements in yeast phosphoglycerate kinase^a.*

N-domain				C-domain			
β -strands		α -helices		β -strands		α -helices	
A	16-23	1a	36-41	G	205-210	8	217-227
		1b	42-52			9	236-246
B	56-63	2	76-88			10	257-273
C	89-94	3	100-108	I	275-280		
D	113-118			o	282-286		
m	129-132			p	294-299	11	315-327
n	135-139	4	141-154	J	330-334	12	346-362
E	157-162	5	163-167	K	365-369	13	371-380
		6	171-174			14	394-401
F	180-184	7	185-199			15	406-410

^aParallel β -strands in each domain are labeled sequentially by uppercase letters. Antiparallel β -strands are labeled with lowercase letters. Helices are numbered in sequential order. Helix 1 is broken by proline 44 into two sections labeled 1a and 1b.

Table 3.2 Results of superimposing highly refined PGK coordinates from other species on the yeast 3PGK and R65Q models. Note that the 3PGK coordinates differ significantly from the equine, porcine, and bacterial structures. The single-domain superpositions of these models on the 3PGK coordinates result in rms deviations averaging 2.96 Å and average three times the predicted values^a.

	N-domain	C-domain	Both
<i>B. stearothermophilus</i>			
α-carbons used	156	186	372
sequence identity	0.51	0.54	0.53
predicted rmsd	1.00	0.95	0.96
observed rmsd to 3PGK	2.68	3.08	3.10
observed rmsd to 1QPG	0.78	0.91	1.22
horse			
α-carbons used	173	185	390
sequence identity	0.62	0.67	0.65
predicted rmsd	0.81	0.74	0.77
observed rmsd to 3PGK	2.76	3.22	3.29
observed rmsd to 1QPG	0.93	0.89	1.46
pig			
α-carbons used	173	200	405
sequence identity	0.62	0.66	0.65
predicted rmsd	0.81	0.76	0.77
observed rmsd to 3PGK	2.82	3.17	3.66
observed rmsd to 1QPG	0.94	0.78	2.02

^aSuperposition was performed for residues in the N-domain, the C-domain, and the entire molecule. Alpha-carbons corresponding to yeast residues 1-4 (N-terminal segment), 186-198 (helix 7), and 400-415 (C-terminal segment) were included in the whole molecule superpositions only. Residues in antiparallel b-strands m and n, and residues near insertions or deletions were not included in any of the calculations. Residues in helix 13 were not used in calculations involving equine PGK. The rms deviations are in Ångstroms.

Table 3.3 *Rotational and translational components of coordinate transformations relating the C-domain orientations in PGK from four different crystal structures^a.*

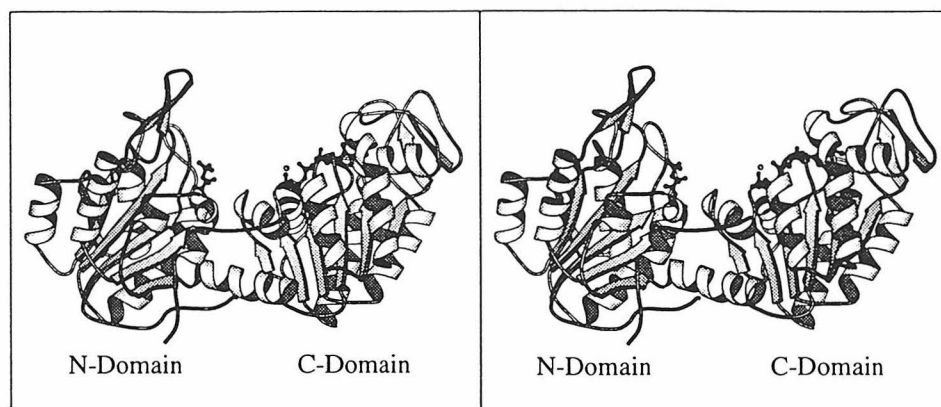
Starting orientations	Final orientations		
	<i>pig</i>	<i>horse</i>	<i>bacteria</i>
<i>yeast R65Q</i>	$\phi = -91^\circ$	$\phi = -92^\circ$	$\phi = -72^\circ$
	$\psi = 113^\circ$	$\psi = 89^\circ$	$\psi = 89^\circ$
	$\kappa = 15^\circ$	$\kappa = 9^\circ$	$\kappa = 7^\circ$
	$\Delta x = 4.0$	$\Delta x = 0.6$	$\Delta x = 0.9$
	$\Delta y = -2.8$	$\Delta y = 0.3$	$\Delta y = 0.2$
	$\Delta z = -0.1$	$\Delta z = 0.9$	$\Delta z = -0.2$
<i>bacteria</i>	$\phi = -109^\circ$	$\phi = -154^\circ$	
	$\psi = 128^\circ$	$\psi = 77^\circ$	
	$\kappa = 9^\circ$	$\kappa = 3^\circ$	
	$\Delta x = 3.2$	$\Delta x = -0.3$	
	$\Delta y = -2.9$	$\Delta y = 0.2$	
	$\Delta z = 0.0$	$\Delta z = 1.1$	
<i>horse</i>	$\phi = -84^\circ$		
	$\psi = 138^\circ$		
	$\kappa = 8^\circ$		
	$\Delta x = 3.5$		
	$\Delta y = -3.2$		
	$\Delta z = -1.1$		

^aEach transformation superimposes a C-domain in a starting orientation on a C-domain in a final orientation. Rotations are represented as spherical polar angles measured from the y-axis. Translations are represented as components along the coordinate axes of an orthogonal reference frame in Ångstroms.

Table 3.4 *Hydrogen bond distances between Mg-AMP-PNP and yeast R65Q PGK.*

Atom from nucleotide	Atom from protein/solvent		Distance (Å)	Atom from nucleotide	Atom from protein/solvent		Distance (Å)
N6	O	Gly 310	3.35	O2A	OD2	Asp 372	3.72
	OH	Wat 545	2.57				
O2*				O2B	OG1	Thr 373	2.59
	OE1	Glu 341	3.43		N	Gly 371	3.58
	OE2	Glu 341	2.85		N	Asp 372	3.30
	O	Gly 338	3.07		N	Thr 373	2.91
	N	Phe 340	3.44				
O3*				O2G	N	Gly 371	2.75
	OE1	Glu 341	2.87				
	OE2	Glu 341	3.08	O3G	ND2	Asn 334	3.35
O1B	N	Gly 371	3.16	N3B	OD1	Asn 334	2.76
	N	Asp 372	2.71				

(a)



(b)

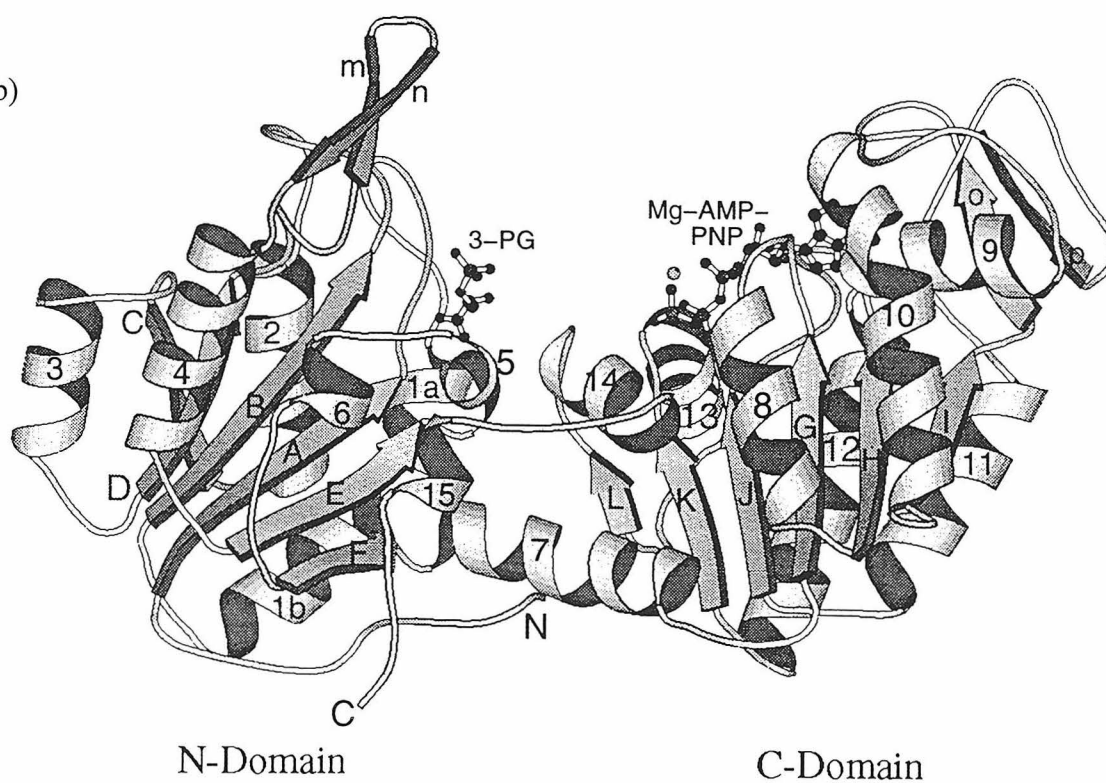


Figure 3.1 (a) Stereoview of the overall structure of yeast R65Q PGK. The ball-and-stick structures represent the nucleotide substrate analog Mg-AMP-PNP bound to the C-domain and 3-PG bound to the basic patch of the N-domain. (b) Magnified view with secondary structural elements labeled. Drawn with MOLSCRIPT (Kraulis, 1991).

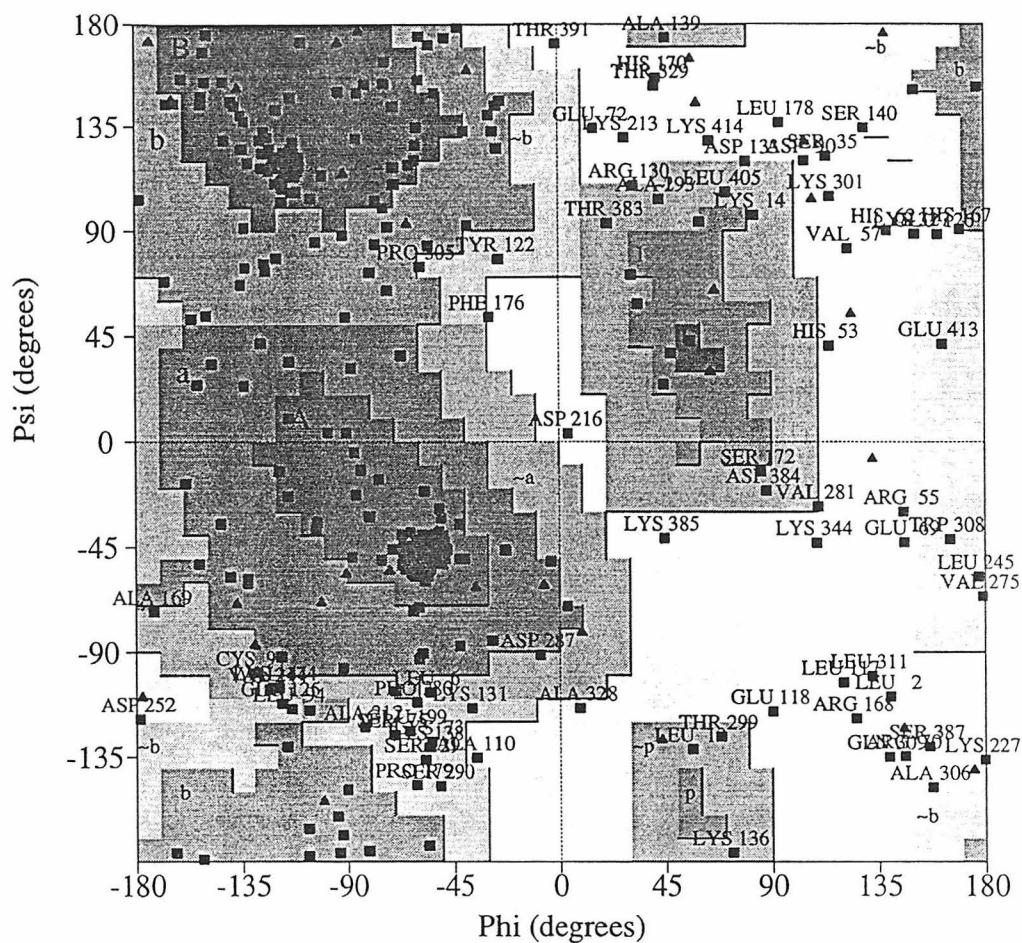


Figure 3.2 Ramachandran diagram for yeast coordinates 3PGK (Watson et al., 1982). Glycines are shown as triangles, non-glycine residues as squares. Labeled residues fall outside the most favored and additional allowed regions. Drawn with PROCHECK (Laskowski et al., 1993).

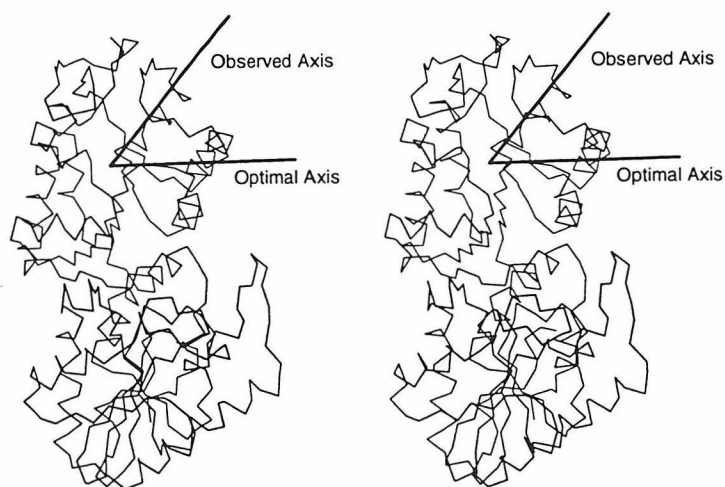


Figure 3.3 *Direction of the observed, average interdomain rotation axis in PGK, and comparison with the optimal rotation axis for domain closure by 'hinge-bending.'*

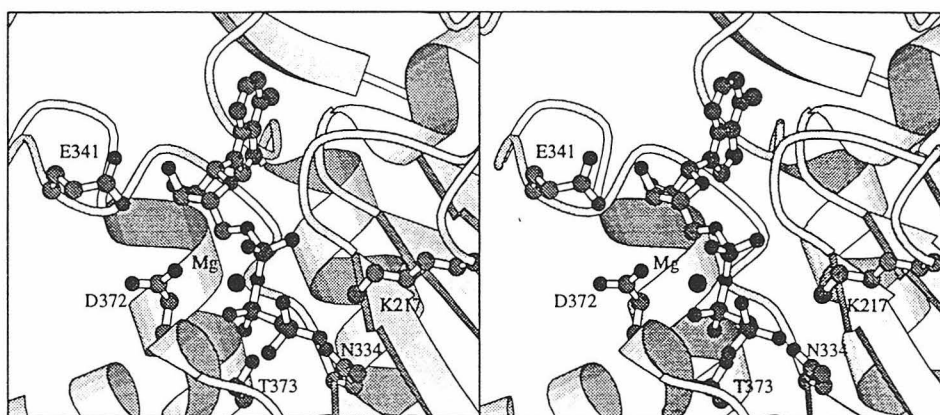


Figure 3.4 Stereoview of Mg-AMP-PNP bound to the nucleotide binding site on the C-domain of yeast R65Q PGK. Side chains that interact with the nucleotide are labeled. Drawn with MOLSCRIPT (Kraulis, 1991).

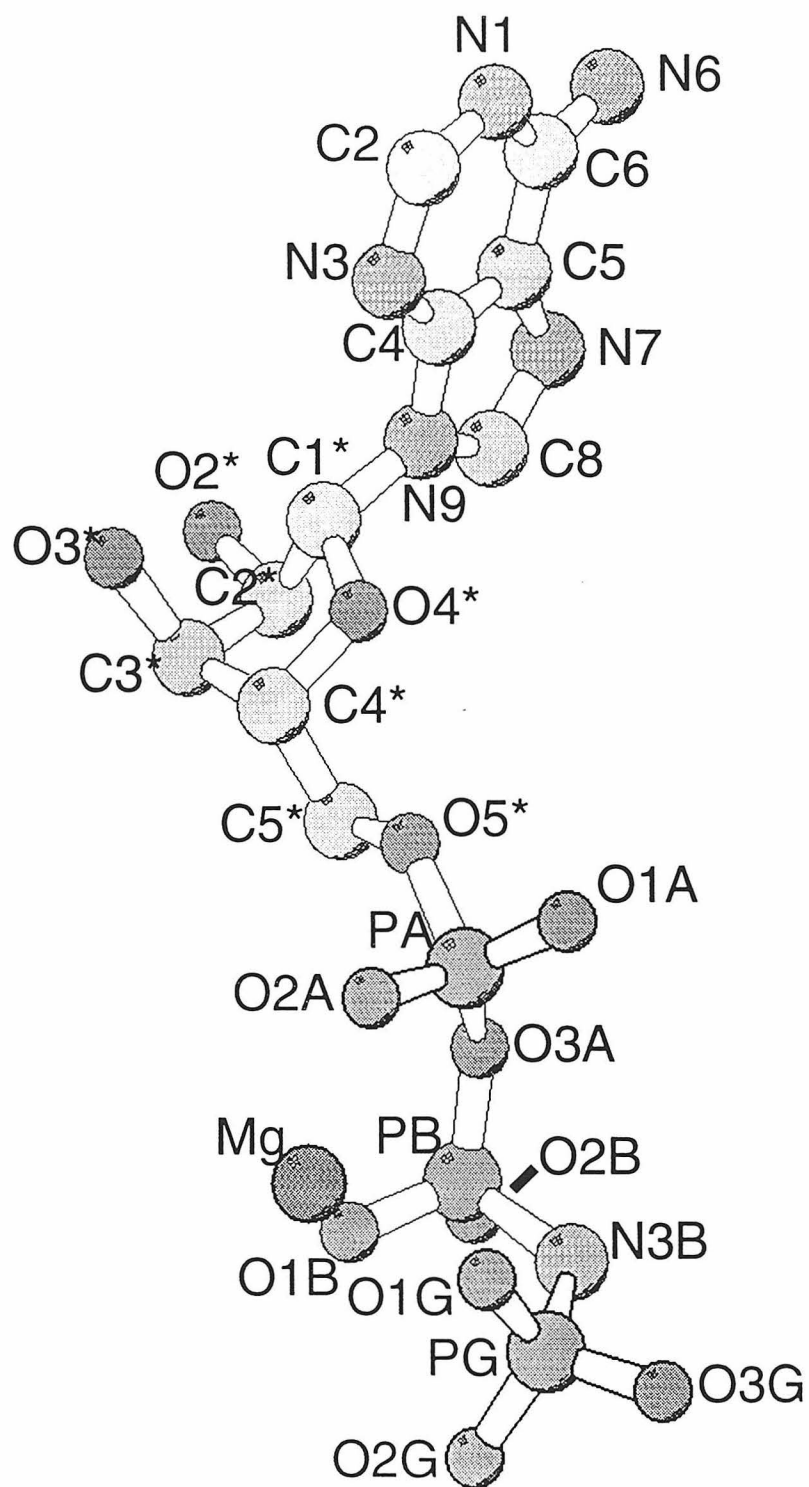


Figure 3.5 *Atomic naming scheme for Mg-AMP-PNP.*

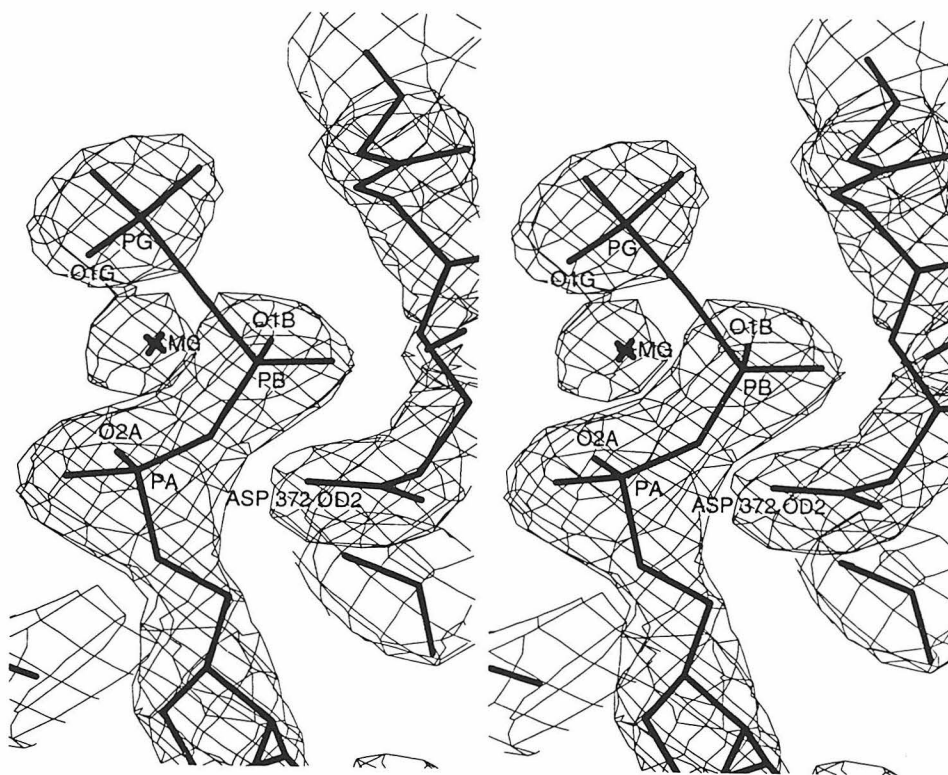


Figure 3.6 Final $2F_o - F_c$ density for the triphosphate group and Mg^{2+} ion of Mg-AMP-PNP bound to the protein. The contour level is 1.5σ .

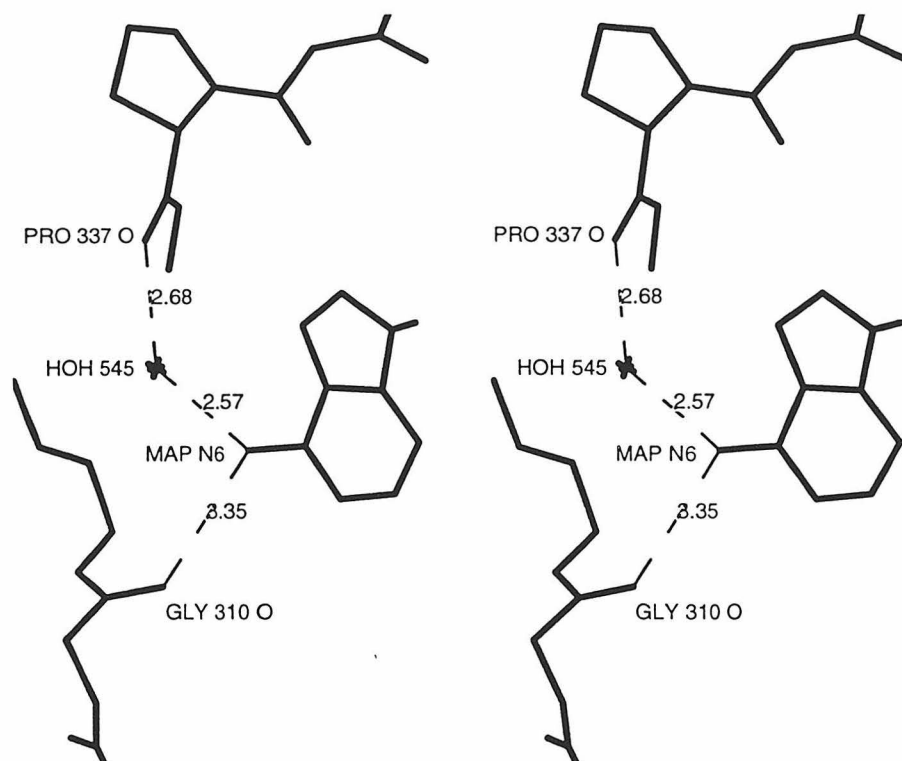


Figure 3.7 *Hydrogen bonding interactions between the adenine ring of the Mg-AMP-PNP (MAP) and the protein. There is a weak hydrogen bond between atom N6 of the nucleotide and the carbonyl oxygen of glycine-310 (3.31 Å) and a water-mediated interaction between the same nucleotide atom and the carbonyl oxygen of proline-337 through water-545. Distances shown are in Ångstroms.*

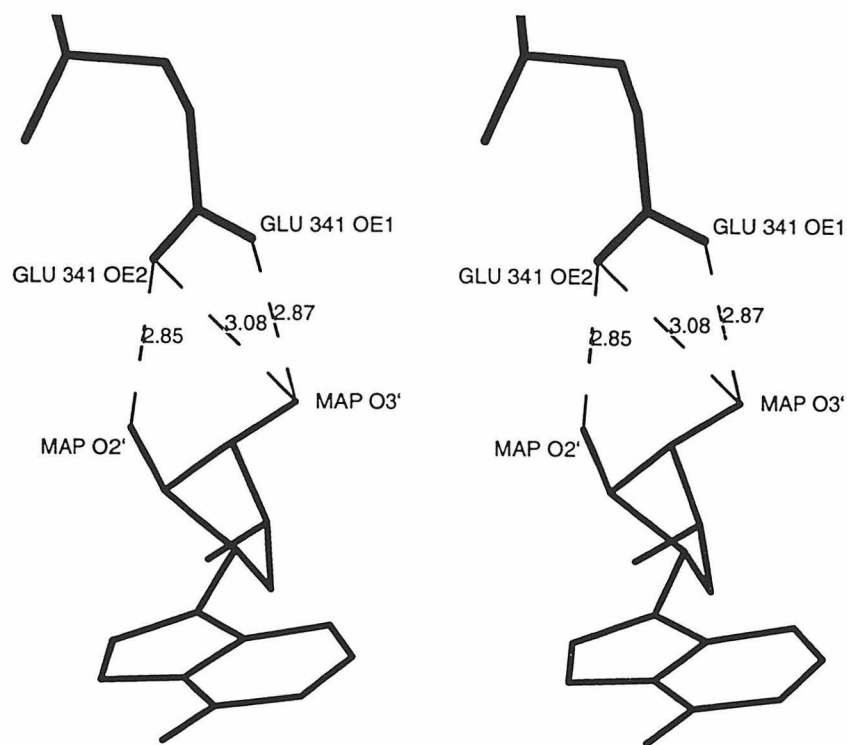


Figure 3.8 Primary interaction between the protein and the sugar moiety of the Mg-AMP-PNP (MAP). The sidechain of glutamate-341 accepts hydrogen bonds from atoms O2' and O3' of the sugar. Distances shown are in Ångstroms.

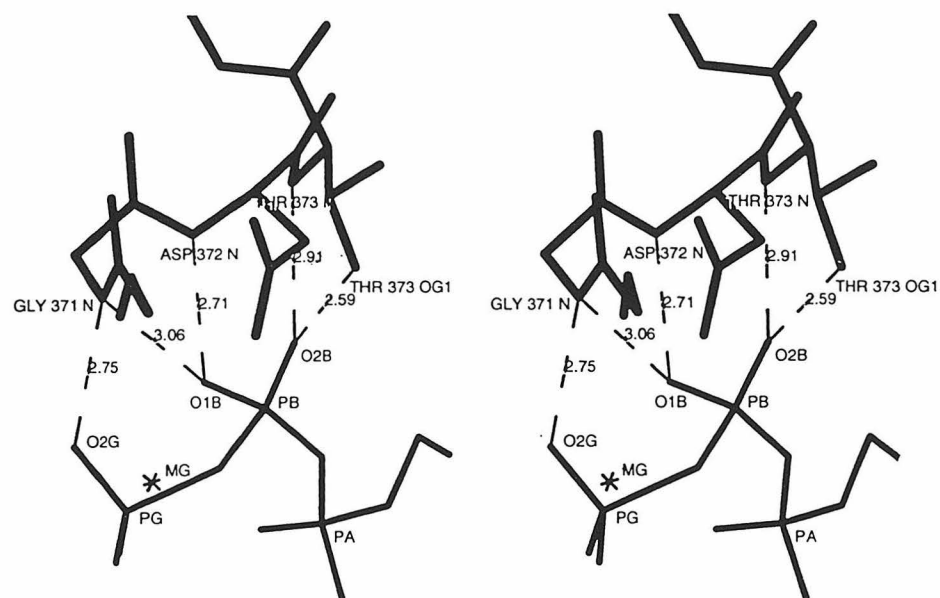


Figure 3.9 *Interaction between the triphosphate moiety of the Mg-AMP-PNP (MAP) and the N-terminus of helix 13. Distances shown are in Angstroms.*

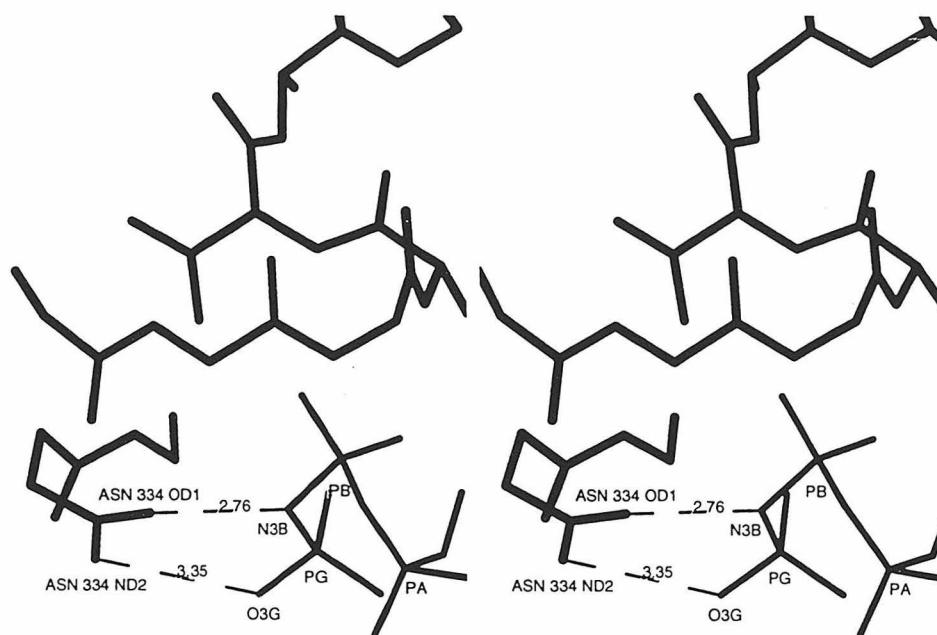


Figure 3.10 *Interaction between the γ -phosphate of the Mg-AMP-PNP (MAP) and the sidechain of asparagine-334. Distances shown are in Ångstroms.*

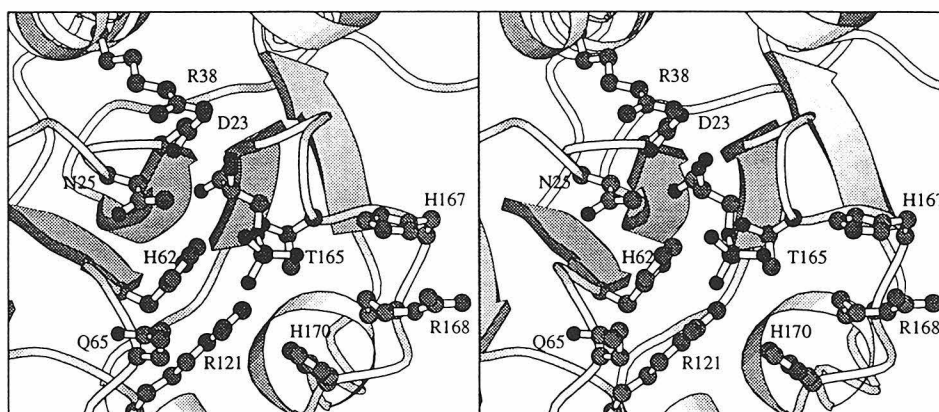


Figure 3.11 Stereoview of 3-PG bound to the basic patch on the N-domain of yeast R65Q PGK. Side chains that interact with the substrate are labeled. Drawn with MOLSCRIPT (Kraulis, 1991).

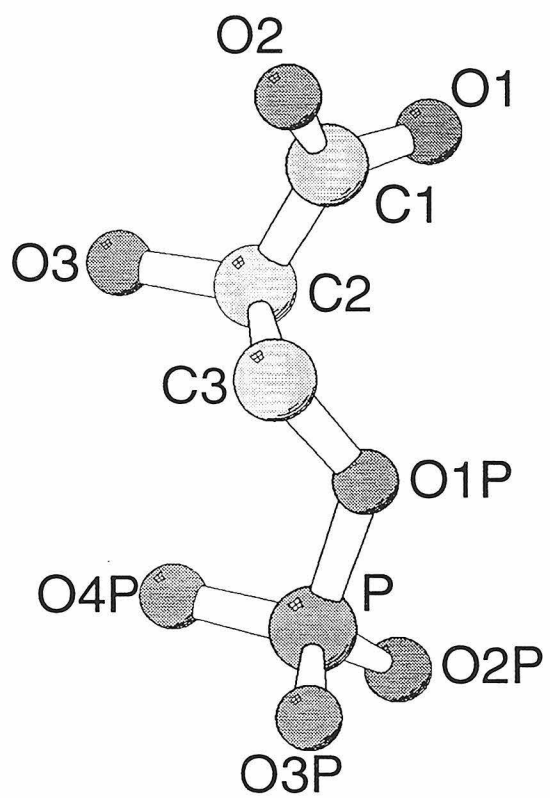


Figure 3.12 *Atomic naming scheme for 3-PG.*

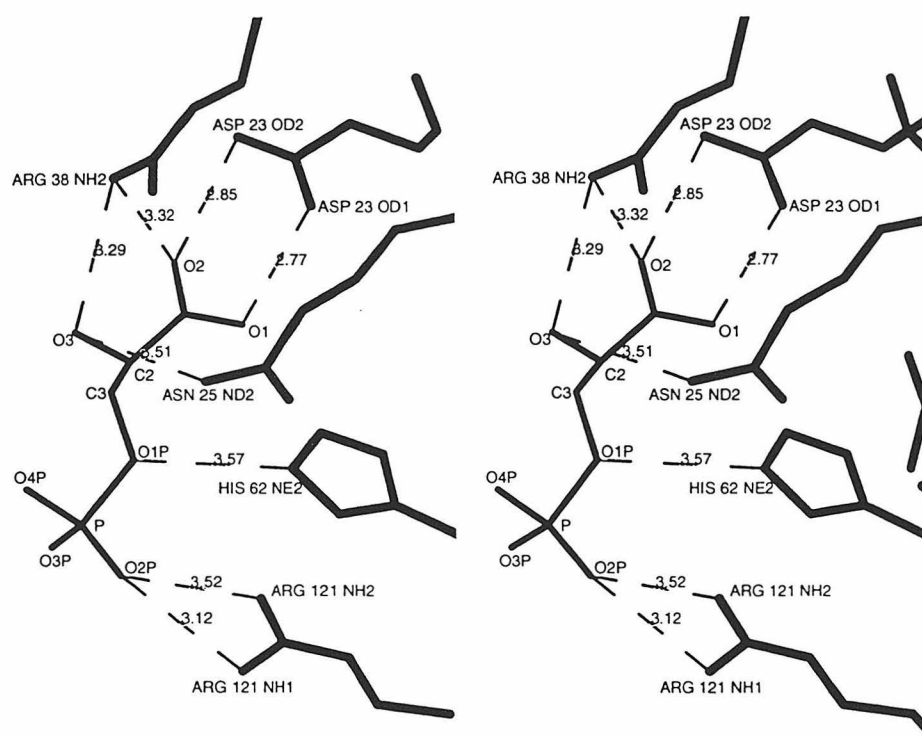


Figure 3.13 Distances between the 3-PG molecule and neighboring protein atoms in the basic patch. Distances shown are in Ångstroms.

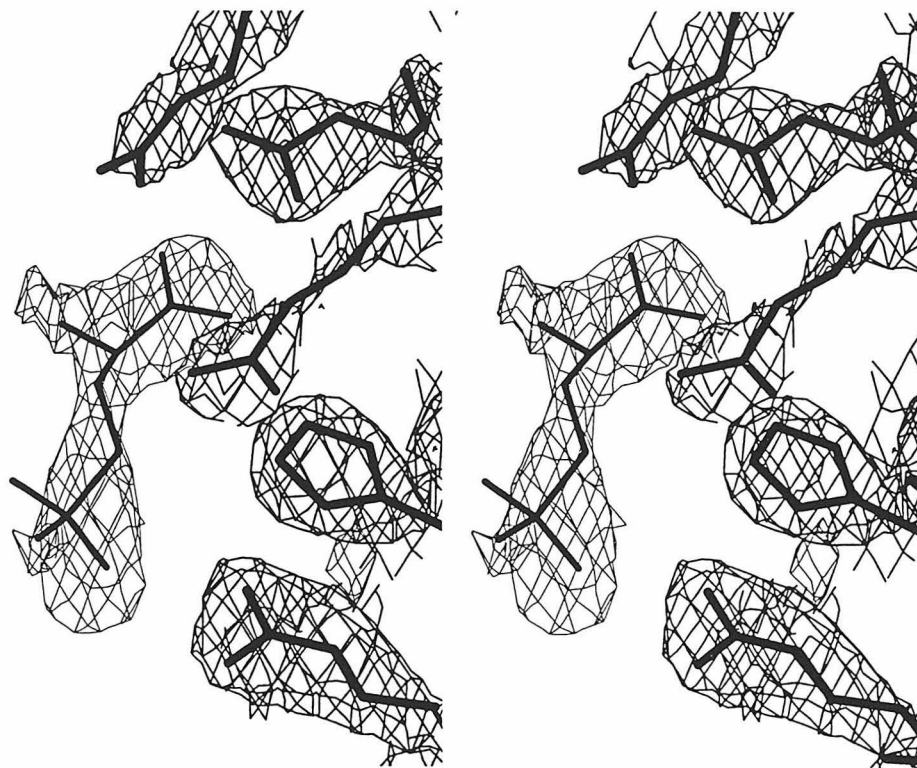


Figure 3.14 Comparison of two alternative orientations of 3-PG in the basic patch with the orientation chosen for the 1QPG model. In each case, the darker electron density contours represent 2.0σ $2Fo-Fc$ omit-map density, while the lighter density represents 2.0σ $Fo-Fc$ omit-map density, where 3-PG was omitted from the model phase calculations. (a) 1QPG model.

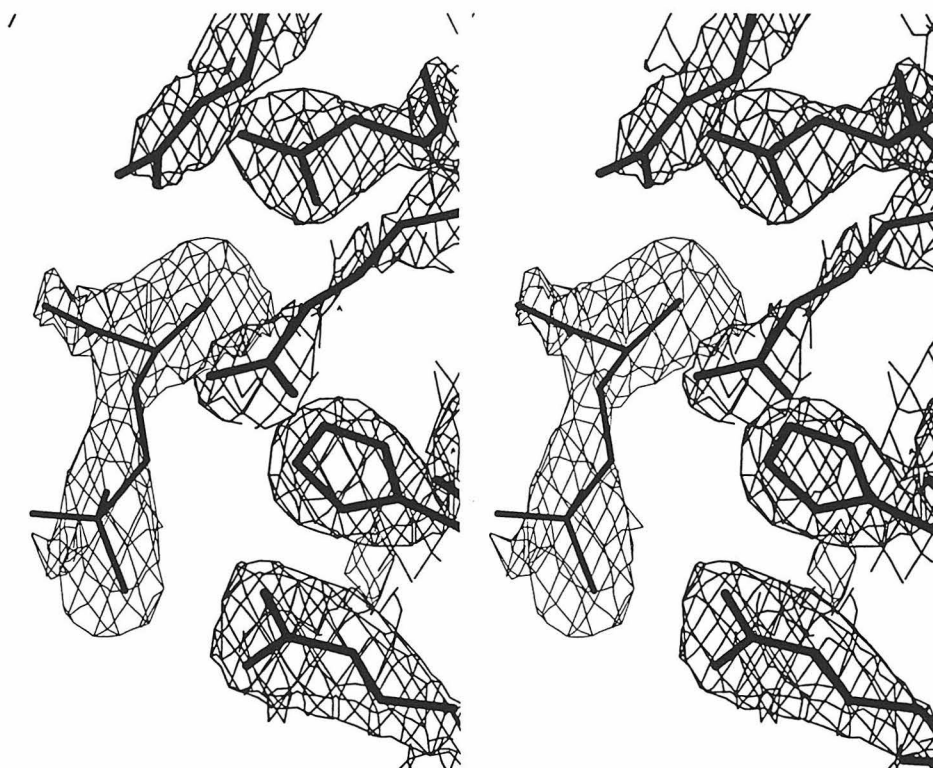


Figure 3.14 (b) *First alternative orientation.*

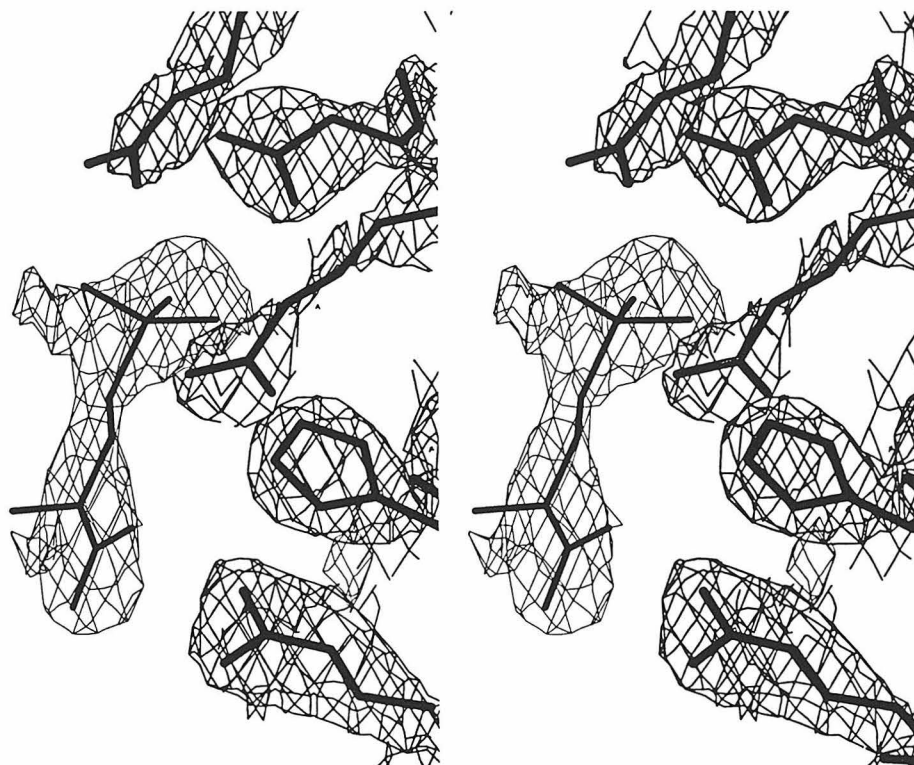


Figure 3.14 (c) *Second alternative orientation.*

Chapter 4

A New Model of PGK Function and Activation*

Chapter 3 demonstrated that the ternary complex of yeast R65Q PGK with Mg-AMP-PNP and 3-PG exhibits an open conformation with the substrates widely separated by the interdomain cleft. This observation is inconsistent with the hinge-bending model of PGK function which predicts a closed configuration for the ternary complex. This chapter shows that the relationships between the crystallographically observed conformations of PGK for which coordinates are available are also inconsistent with a 'hinge-bending' behavior that would close the interdomain cleft. However, these relationships are consistent with solution studies of the dynamics of PGK in the presence of various substrates.

Based on these observations and available site-directed mutagenesis data, an alternate model of PGK activity and activation is proposed. In this model, the basic patch represents the site of anion activation--not the catalytically active binding site for 3-PG. The active site of PGK is proposed to be the combination of the nucleotide binding site on the C-domain and a possible 3-PG binding site between the domains, near the N-terminus of helix 14 and the sidechain of arginine-38. In this model, anions bound to the basic patch activate PGK by favoring a shift of helices 5 and 14 with respect to the N-domain, thereby stabilizing conformations of the enzyme which bind 3-

PG in the active site. The kinetic properties and crystal structure of the R65Q mutant of yeast PGK are consistent with this proposal.

A. Implications of the Yeast R65Q Structure

According to the hinge-bending model of the PGK activity, substrate 3-PG binds to the basic patch on the N-domain while ATP binds to the nucleotide binding site on the C-domain. After both substrates bind, a large interdomain motion brings the occupied substrate binding sites together to form the active site; the domain closure is required for phosphoryl transfer. This model predicts that the ternary complex of PGK with its substrates should adopt a closed conformation. However, as described in Chapter 3, the ternary complex of yeast R65Q PGK with AMP-PNP and 3-PG is open. Therefore, this structure casts serious doubt on the hinge-bending model.

Actually, the structure is not surprising in light of the numerous site-directed mutagenesis studies of basic patch residues described in Chapter 1. The hinge-bending model assumes that the basic patch is the catalytic binding site for 3-PG. However, the results of mutagenesis studies of basic patch residues suggest that these residues, except perhaps for arginine-38, do not bind the substrate 3-PG during phosphoryl transfer. Only arginine-38 is essential for catalysis. If the basic patch residues do not bind substrate 3-PG during phosphoryl transfer, then there is no need for the basic patch to approach the nucleotide binding site so closely, and domain closure by hinge-bending motion may not be required for activity.

* Parts of this chapter have been published in McPhillips, T.M., Hsu, B.T., Sherman, M.A., Mas, M.T., Rees, D.C. (1996) Structure of the R65Q Mutant of Yeast 3-Phosphoglycerate Kinase Complexed

On the other hand, if the basic patch is not the substrate binding site for 3-PG and if PGK activity does not require alternation between completely open and completely closed conformations, then three key mechanistic issues need to be addressed:

- (1) Domain motions are observed in PGK. What is the significance of the different conformational states of the enzyme?
- (2) If the catalytic binding site for 3-PG is not in the basic patch, where is it?
- (3) The basic patch binds 3-PG. If this site is not the catalytic site, then what is its role?

The proposed answers to these questions form a new model of PGK function and activation. The final section of this chapter discusses this model in light of the recently reported structure of a closed conformation of PGK (Bernstein et al., 1997).

B. Significance of the Conformational States of PGK

B.1. Crystallographically Observed Domain Motions

What is the significance of the observed domain motions in PGK? Various spectroscopic experiments indicate that the conformation of PGK depends on which substrates are present. However, the most detailed information about the orientations of the domains in PGK under different conditions come from the crystal structures of the enzyme. By examining the domain orientations in these structures, one can answer the following question: Are the different domain orientations observed in the various

crystal structures of PGK consistent with a simple hinge-bending motion that brings the basic patch and nucleotide binding site together?

As shown in Chapter 3, comparing the yeast, equine, porcine, and bacterial forms of PGK demonstrates that, within the domains, all four species have essentially the same structure. The small structural variations observed within the domains are consistent with predicted variations among homologous proteins with different sequences. In contrast, substantial variations are observed in the relative orientations of the N- and C-domains of the available PGK structures. While sequence differences complicate the analysis, the observed variations in domain positions probably reflect different natural conformations of PGK selected by ligand and solution conditions. In other words, the crystal structures of yeast, equine, porcine, and bacterial PGK may represent samples of PGK conformations in solution. Assuming that the crystallographically observed domain orientations represent a few possible conformations out of a continuous range of motion, a series of PGK structures that encompass the observed orientations were generated by applying the transformations of Table 3.3 to the yeast R65Q structure. A similar analysis has been described recently by Vonnrhein et al. (1995) to characterize domain motions in adenylate kinase.

The PGK structures generated were examined to determine whether these transformations can close the domains or decrease the distance between the basic patch and the nucleotide binding site. The distance between the centroids of the two domains and the distance between 3-PG and AMP-PNP in the generated structures are plotted in Figure 4.1. The distance between the domain centroids decreases only 1.7 Å over 24° of rotation, and the closest the carboxylate oxygen of 3-PG and the phosphate of the

nucleotide approach each other over the same range of motion is 8.3 Å. Continuing the rotations and translations beyond those observed crystallographically does not result in either the domains or the substrates moving significantly closer.

Furthermore, this analysis indicates that the rotations and translations relating the observed conformations of PGK differ significantly from those optimal for domain closure or 'hinge-bending.' The optimal translation vector for domain closure would be directed along the vector between the centroids of the two domains. Similarly, the optimal translation vector for joining the basic patch and nucleotide binding site would be along the vector between the 3-PG and AMP-PNP molecules. The optimal rotation axis for both operations would be one normal to both the inter-centroid and inter-substrate vectors. Table 4.1 compares these optimal rotation and translation axes with those relating the crystallographically observed orientations of the domains, and Figure 3.3 shows the direction of the optimal rotation axis along with that of the observed average rotation axis. In the first 8° of rotation/translation, the observed translation axis is fairly close to the optimal translations for domain closure. However, the actual translation along this axis is only 1.0 Å, and the observed rotation axis deviates more than 70° from the optimal axis. In the final 8° of rotation/translation, the rotation axis is less than 30° from the optimal axis, but the translation vector is nearly 80° away from the inter-centroid vector and nearly 70° away from the inter-substrate vector. Thus, the 4.6 Å translation along this axis results in only a 1 Å decrease in the distance between the centroids and less than a 2 Å decrease in the distance between the substrates. Consequently, the motion interconverting the observed structures is better described as a 'twisting' around the interdomain vector than as a 'hinge-bending.'

B.2. PGK Dynamics in Solution

It was assumed in the analysis above that the crystal structures represent different natural conformations of PGK selected by ligand and solution conditions. The conclusions made above can be compared to time-resolved fluorescence energy transfer experiments that have explored the properties of different equilibrium conformations of yeast PGK in the presence of various ligands in solution (Haran et al., 1992). The results of the solution study, summarized in Table 4.2, are consistent with the proposal that the interdomain dynamics of PGK consists chiefly of twisting.

The fluorescent probes were attached to residue 135, on the loop connecting antiparallel strands m and n, and to residue 290, on the loop connecting antiparallel strands o and p (see Figure 4.2). The probes were located on opposite sides of the interdomain cleft so that the distance between them would indicate the relative orientation of the domains. A simple closure of the interdomain cleft would decrease the distance between the probes. The inter-probe distance distributions were measured in the presence of the following combinations of substrates: (1) no substrates, (2) 3-PG only, (3) ATP only, and (4) both 3-PG and AMP-PNP. Table 4.2 lists both the mean distance between the probes and the width of the distance distribution under each condition.

In light of the hinge-bending model, it was expected that the mean distance between the probes would be largest in the substrate-free case, and smallest in the ternary complex. In fact, the opposite result is observed. On average, the probes are separated

by 31.5 Å when no substrates are present but by 38.2 Å when both substrates are present. The presence of ATP without 3-PG yields results similar to the ternary case. An intermediate value of 35.2 Å is observed when only 3-PG is present. Thus, the data indicate that substrate binding dramatically *increases* the average distance between the probes over that observed for substrate-free enzyme. Since only one intramolecular distance was measured in this experiment, one cannot tell what conformational changes are actually occurring in solution. The effects could result from a hinge-bending motion that causes the domains to *open* on substrate binding. Another explanation is that the domains undergo the twisting motion consistent with the crystal structures.

The latter possibility seems more likely. The fourth column of Table 4.2 shows the crystallographically determined distances between the residues in the porcine and R65Q yeast enzymes corresponding to Q135 and S290 in the native, yeast enzyme. The distances in the porcine structure (a binary complex with 3-PG) and the yeast R65Q structure (a ternary complex with 3-PG and AMP-PNP) are amazingly close to the solution values. Thus, not only does the solution study contradict the hypothesis that substrate binding favors domain closure, it is consistent with the proposal that the primary mode of domain motion is a twisting motion.

In addition to showing that the average orientation of the domains depends on the combination of substrates bound to the enzyme, the solution study also shows that PGK is an extremely flexible molecule. The widths of the interprobe distance distributions (third column of Table 4.2) indicate that the domains of PGK undergo motions 10 Å or more in amplitude around the average structures, i.e., 20 Å or more between extreme positions. Furthermore, the distance distributions indicate that substrate binding has a

dramatic effect on the flexibility of the protein, reducing the amplitude of interdomain motion by as much as 35%. Substrate binding thus appears both to select the average conformation of the enzyme and to narrow the distribution of conformations around the average. Further analysis of the fluorescence data indicates that the rate of interconversion among conformational states is several orders of magnitude faster than catalysis. Thus, the conformation of the protein is in constant flux regardless of whether substrates are being turned over or not.

This flexibility of PGK has another important implication. Namely, the various crystal structures may not represent equilibrium, solution conformations of the enzyme. True, the fluorescence data discussed above represent equilibrium structures and agree well with two crystal structures. But the range of domain motion observed in solution implies that the energy wells about the equilibrium conformations are both shallow and broad. Thus, crystal packing forces may select the crystallographically observed conformations from a large range of interdomain motion. The protein sequences may play a greater role in selecting the different crystallographically observed conformations than anything else. If so, then the crystal structures should be viewed as snap-shots of PGK conformations relatively uncorrelated with the nature of any ligands bound to the enzyme.

B.3. Role of Conformational States in PGK Function

The results of the solution study provide additional evidence against models of PGK function in which substrate binding triggers a conformational change from an open,

inactive form to a closed, catalytically competent form. The contradiction is twofold. First, the high rate of interconversion among conformational states in solution contradicts the view that the enzyme closes around newly bound substrates and opens again only after phosphoryl transfer has occurred. Rather, the enzyme appears to continuously sample a wide range of conformations, catalytically active and not. One may hypothesize that turnover occurs when the enzyme happens to assume the catalytically active conformation while both substrates are bound. Substrate binding narrows the distribution of conformations sampled by the enzyme but does not cause the protein to assume a single conformation.

Second, it appears unlikely that the hypothesized, catalytically active conformation is a 'closed' one. The narrowing of the distribution of domain orientations upon substrate binding presumably increases the chances that the active conformation will occur. Yet, the average distance between the two probes appears to increase in response to substrate binding. Consequently, the probability of the enzyme assuming a completely closed conformation is *reduced* on substrate binding. Thus, the catalytic state is expected to be similar to the average structures observed for the substrate-bound forms of PGK in solution and in the crystal structures.

C. The Catalytic Binding Site for 3-PG

C.1. Conformations of the Interdomain Region

The twisting motion consistent with the available crystal structures cannot bring 3-PG bound to the basic patch any closer than 7 Å from the nucleotide substrate.

Furthermore, the distances observed in the fluorescence energy transfer experiments described above are consistent with the distances this twisting motion predicts. In other words, neither the crystal structures nor the fluorescence data support the primary assumption made by the hinge-bending model, that the basic patch forms part of the active site of PGK. More important, the results of numerous site-directed mutagenesis studies of basic patch residues have shown that, with the exception of arginine-38, none of the basic patch residues studied are essential for catalysis. Thus, the basic patch is unlikely to be the catalytic binding site for 3-PG.

But if the basic patch is not the catalytic binding site for 3-PG, where is it? It was hypothesized in the previous section that substrate binding favors average domain orientations close to the active conformation. It is possible that the catalytic binding site for 3-PG forms transiently when such conformations occur. This binding site would have to be located in a part of the protein that varies significantly as a function of domain orientation. Since the internal structure of the two domains of the protein are well conserved regardless of domain orientation, the obvious place to look for such a site is *between* the two domains.

Examination of the available crystal structures confirms that the conformation of the polypeptide chain in the interdomain junction changes significantly to accommodate domain motion. Helix 7 either rotates about an axis perpendicular to its length or bends significantly. Similarly, helix 14 translates in a piston-like fashion along its axis to keep the domains connected. Contrasting the yeast and porcine structures illustrates the motions of the interdomain helices (see Figure 4.3). Helix 7 is nearly straight in the yeast structure but clearly bent in the porcine structure. And, helix 14 in the porcine

structure is shifted 1.5 Å toward its N-terminus with respect to the helix in the yeast structure. Since the triphosphate group of enzyme-bound ATP is located near these variable regions of the interdomain region, the catalytic binding site for 3-PG may form transiently in this area.

C.2. The N-Terminus of Helix 14

A likely candidate for the catalytic binding site for 3-PG has been identified in the interdomain region (see Figure 4.4). As a result of the changes in the interdomain region described above, a potential binding site for the phosphate group of 3-PG appears at the N-terminal end of helix 14. Three sequential glycines, residues 392-394, are positioned at this location, and they form a structure similar to the P-loop, a phosphate binding structural element found in many nucleotide binding proteins (Schulz, 1992). Glycine-392 is conserved in the bacterial, horse, pig, and yeast sequences, and glycines 393 and 394 are conserved in all known sequences of the enzyme. These glycines could be expected to donate hydrogen bonds to the phosphate group of 3-PG bound at this site. The helix dipole would further stabilize the negatively charged substrate.

Locating of the catalytic binding site for 3-PG at this position would explain why the R38A mutation is catalytically inactive. The sidechain of arginine-38 is near the porcine counterpart of yeast glycine-394 at the N-terminus of helix 14 (minimum distance = 4.6Å). If the phosphate group of a 3-PG molecule were bound to the N-terminus of helix-14, the guanidium moiety of the arginine-38 sidechain could help hold it there with its positive charge and hydrogen-bond donating properties. Thus, arginine-38 may be essential because it binds 3-PG during phosphoryl transfer. Since none of

the other basic patch residues are this close to the N-terminus of helix 14, this model of 3-PG binding explains why arginine-38 is the only basic patch residue essential for catalysis (Sherman et al., 1992).

This role for arginine-38 is consistent with the hypothesis that only certain conformations of PGK are catalytically active. Since arginine-38 is part of the N-domain, the distance between arginine-38 and the N-terminus of helix 14 varies among the different crystal structures. While the shortest distance between arginine-38 and glycine-394 is only 4.6 Å in the porcine structure, it is 8.1 Å in the bacterial structure and 9.2 Å in the yeast R65Q structure. Thus, domain motion and the accompanying changes in the interdomain region appear to control the catalytic ability of the enzyme by modulating the affinity of this site for 3-PG.

Modeling 3-PG into this proposed site in the porcine structure emphasizes the most significant feature of the site. When the phosphate of 3-PG is positioned between arginine-38 and the N-terminus of helix 14, the carboxyl end of the substrate can be placed very near the nucleotide binding site. If the Mg-AMP-PNP molecule from the R65Q yeast model is modeled in the porcine structure as shown in Figure 4.4, the Mg^{2+} ion falls within coordination range of the 3-PG carboxyl group. Coordination of the metal ion by 3-PG in this way could be the first step in the phosphoryl transfer mechanism of PGK as discussed in Chapter 1. Thus, the combination of the nucleotide binding site with this potential 3-PG binding site may represent the true active site of PGK.

The problem with this model of 3-PG binding is that 3-PG does not occupy this site in any crystal structure, including that of porcine PGK. Given the variability of the geometry of this site, however, it is possible that none of the crystallographically observed conformations of PGK are close enough to the active conformation to bind 3-PG at this location.

D. The Basic Patch

D.1. An Activatory Role for the Basic Patch

If the catalytic binding site for 3-PG is located in the interdomain region, then what is the role of the basic patch? And why does 3-PG bind there? PGK is activated by low concentrations of anions, including 3-PG (Scopes, 1978b). NMR studies have shown that the basic patch is the primary site for both 3-PG and non-substrate anion binding and activation (Fairbrother et al., 1990). In addition, mutation of basic patch residues often results in complete loss of anion activatory behavior (Sherman et al., 1992). Thus, the basic patch appears to be the site of anion activation.

How does anion binding at the basic patch activate PGK? The fluorescence data discussed above shows that 3-PG binding to PGK in solution reduces the flexibility of the enzyme and changes the equilibrium conformation of the enzyme. It is possible that anion binding at the basic patch modulates the flexibility of the protein similarly. In this case, binding of either 3-PG or non-substrate anions at the basic patch could activate PGK the same way: by reducing the range of domain motion and shifting the

average conformation towards the catalytically active form. The question then becomes, by what mechanism does anion binding affect the dynamics of PGK?

D.2. Effect of Anion Binding on the Basic Patch Structure

Of the equine, porcine, bacterial and yeast PGK structures available, only the porcine and yeast R65Q structures have a 3-PG molecule occupying the basic patch. Comparing the basic patch region in these four PGK crystal structures provides a clue to how anion binding to the basic patch affects the dynamics of the protein. Visual inspection of the four structures, using superpositioned N-domains, reveals that the side of the anion-binding pocket formed by the turns following β -strands A, B, and D is virtually identical in each species. The opposite side formed by helix 5 is also the same in the yeast R65Q, equine, and *B. stearothermophilus* PGK structures. However, this side is significantly different in the porcine structure, as illustrated in Figure 4.5. (Section D.3 will discuss why the basic patch of the yeast R65Q structure does not look like that of the porcine model, despite the presence of 3-PG.)

The difference between the porcine structure and the other species can be summarized as follows (see Figure 4.5). Helix 5 of porcine PGK is translated about 1.5 Å with respect to the rest of the N-domain (Harlos et al., 1992). The direction of this translation is towards the N-terminus of the helix. This translation causes arginine-168 at the C-terminus of helix 5 to shift 1.8 Å (C_α to C_α) from the *B. stearothermophilus* position, allowing the sidechain to bind the phosphate group of the 3-PG molecule, held tightly in place by the sidechains of arginines 65 and 121. The translation of helix 5

also brings threonine-167 into hydrogen-bonding range of aspartate-23. No other large effects of this shift are observed in the N-domain of the pig structure. However, the translation of helix 5 in the N-domain has a large effect on the orientation of the C-domain. In all of the structures interdomain helix 14 is packed against helix 5. Consequently, helix 14 in the pig structure moves with helix 5 to preserve favorable contacts. The entire C-domain moves with helix 14, resulting in the rotation and translation of the domain described in Chapter 3. Thus, the shift of helix 5 in the basic patch of the porcine structure appears to be related to the domain orientation observed in this structure.

A model of anion activation can be deduced from these observations. First, note that binding 3-PG to the basic patch appears to favor the shift of helix 5 in the porcine structure. The shift of helix 5 and the accompanying domain reorientation allows a charged hydrogen bond to form between arginine-168 and the 3-PG phosphate group. Since this favorable interaction is not possible in the absence of the anion, 3-PG bound to the basic patch appears to stabilize the conformation of porcine PGK. Activatory sulfate ion has been shown by both crystallography (Blake & Rice, 1981) and NMR (Fairbrother et al., 1990) to bind to arginines 65, 121, and 168, as observed for the 3-PG phosphate in the porcine structure. Thus, both 3-PG and non-substrate anions such as sulfate may activate PGK by stabilizing domain orientations similar to the conformation observed in the porcine structure.

The problem with this interpretation is obvious. If 3-PG binding to the basic patch favors this shift of helix 5 and the resulting orientation of the C-domain, why does the R65Q structure not look like the porcine model? There are two possible explanations.

First, as discussed above in the context of the fluorescence experiments, PGK appears to be very flexible. It is not known whether the crystal structures of PGK represent equilibrium conformations. In other words, crystal packing forces may select the domain orientations observed in the porcine and yeast R65Q structures, regardless of what substrates are present in them. The other explanation is that the basic patch mutation decouples the interdomain orientation from the presence or absence of anions in the basic patch. As discussed below, this possibility would explain the kinetic properties of the mutant enzyme as well as its structure.

D.3. The R65Q Basic Patch Mutation

The properties of the R65Q mutant of yeast PGK are consistent with the differences observed between the basic patch regions of the porcine and yeast R65Q PGK structures. Recall from Chapter 1 that the R65Q mutant is not activated by anions and has a $V_{\max}$ twice that of wild-type, but a catalytic efficiency lower than wild-type. All of these properties can be explained in terms of the model of anion activation proposed above.

First consider why the loss of one arginine sidechain should result in a significant difference in basic patch structure. Comparison of the four available crystal structures shows that the position of arginine-121 is constant in the three wild-type PGK structures, as well as the R65Q yeast structure (see Figure 4.6). The sidechain is oriented to bind 3-PG or sulfate ion regardless of domain orientation. In contrast, the positions and sidechain conformations of residues arginine-65 and arginine-168 vary

between the four structures. They approach the site of 3-PG binding most closely in the 3-PG bound porcine structure as shown in Figure 4.6. Residue 65 is a glutamine in the R65Q mutant, so it does not bind the 3-PG in this structure. The question is, why doesn't arginine-168 bind the 3-PG in the R65Q structure?

The loss of arginine-65 means that one of the anion-binding interactions is eliminated. This may result in 3-PG binding with lower affinity or less orientational specificity. The affinity of the site for non-substrate anions may be reduced as well. If anions are more poorly anchored in the mutated basic patch, they may not be able to stabilize the alternative position of arginine-168 and the conformation of the enzyme seen in the porcine structure. This would explain why arginine-168 is observed in the 'anion-free' orientation in the R65Q structure, despite the presence of 3-PG in the basic patch.

The relatively high activity of the R65Q mutant can be explained. First, one must assume that the domain orientation observed in porcine PGK is similar to the active conformation. The porcine model places three arginine sidechains in close proximity. Without an anion bound between these positively charged groups, the arginine sidechains are electrostatically repelled by each other, and the porcine conformation of PGK is disfavored. Binding a divalent anion to this site, either a sulfate ion or the phosphate group of a 3-PG molecule partially neutralizes the charges of the sidechains and lowers the energy of the porcine configuration. Elimination of one of the positive charges repelling arginine-168, that of arginine-65, lowers the energy of the porcine conformation and obviates the need for a stabilizing anion.

The lower efficiency of the mutant can be explained if one considers that not only does the mutant not *need* an anion in the basic patch, the binding of anions *cannot* improve the activity of the mutant. Anions bind too loosely to the mutant basic patch to significantly stabilize the active conformation. And since the positive charge of arginine-121 remains when anions are absent from the site, the active conformation of the mutant is not as stable as the anion-bound, wild-type case. Consequently, the mutant enzyme spends less time in the active conformation, the K_m values for the substrates are higher, and the catalytic efficiency, correspondingly lower.

E. A Closed Conformation of PGK

E.1. Making Sense of the Trypanosomal PGK Structure

The model of PGK activity and activation proposed in this chapter is consistent with the kinetic properties of the enzyme, the results of site-directed mutagenesis experiments, and the crystal structures for which coordinates are available. But how does the model hold up in light of the recently reported crystal structure of a closed conformation of trypanosomal PGK (Bernstein et al., 1997)?

At first glance, this structure appears to lend credence to the hinge-bending model. The structure of PGK from *Trypanosoma brucei* is a ternary complex of the enzyme, ADP, and 3-PG. The two ligands bind to the basic patch and nucleotide binding site, as previously observed. Significantly, the interdomain cleft is closed, and the binding sites are substantially closer than previously observed. The authors argue that the structure represents the first view of PGK in a catalytic conformation. The implication is that the

hinge-bending model of PGK now has the experimental support it needs. Before accepting this conclusion, however, one should consider this structure in light of the topics of this chapter, namely, the significance of the dynamics of PGK, the location of the catalytic binding site for 3-PG, and the role of the basic patch in anion activation.

E.2. PGK Dynamics and the Trypanosomal PGK

First, consider the implications of the new structure for the dynamics of PGK. How did this new conformation arise if the enzyme chiefly undergoes twisting motions? The authors state that a newly discovered point of flexibility is responsible for the domain closure. The protein bends at the C-terminus of interdomain helix 7, causing the C-domain to rotate 26° with respect to the N-domain. This domain motion cannot be analyzed fully until coordinates are made available. However, it seems likely that the direction of the rotation axis differs significantly from those described earlier in this chapter. If so, then there are two primary modes of domain motion in PGK. One is the twisting of the domains that relates the previously observed crystal structures. The second is a direct domain closure that can occur at only one point along the twisting trajectory. The two motions would be analogous to the operations necessary to join two BNC connectors—a twist until the connections match perfectly, and then closure.

But, what is the significance of the new conformation itself? Bernstein et al. describe the physiologically relevant conformations of PGK in terms of the horse (ligand free), bacterial (binary complex with ADP), and porcine (binary complex with 3-PG) models analyzed in this chapter, in addition to the new structure (ternary complex

with 3-PG and ADP). They assume that the crystallographically observed conformations of the enzyme in these structures reflect the equilibrium conformations of PGK in solution in the presence of the respective ligands. The problem with this interpretation is that there are two other ternary complexes, the yeast R65Q structure (3-PG and Mg-AMP-PNP) and a recently reported porcine structure (3-PG and Mn-AMP-PNP). Neither of these structures assume the closed conformation observed by Bernstein et al. Since the combinations of ligands included in these complexes are arguably more catalytically relevant than in the trypanosomal PGK structure (not much phosphoryl transfer can occur between 3-PG and ADP), the assumption that the ligands determine the domain orientations in the crystal structure actually implies that the active conformation is an open conformation and that the unproductive ternary complex is closed.

Fortunately, this assumption that the ligands determine the crystallographically observed domain orientations is difficult to support. The only species for which crystal structures of two different complexes have been determined is pig. The binary complex of porcine PGK with 3-PG, and the ternary complex with 3-PG and Mn-AMP-PNP are reported to be very similar. On the other hand, a similar ternary complex, that of yeast R65Q PGK with 3-PG and Mg-AMP-PNP, is quite different from the porcine structures. Thus, there is more evidence that the protein sequence selects the conformations of PGK that crystallize than that the ligands determine the gross domain orientations in the structures. The real significance of the new structure is that a new conformation of PGK has been revealed.

E.3. The Active Site of PGK and the Trypanosomal Structure

Next, consider the implications of the new structure for the active site of PGK. Bernstein et al. report that the space between the bound ADP and 3-PG molecules is bracketed by the N-terminus of helix 14 and the sidechain of arginine-39 (which corresponds to arginine-38 in the yeast sequence). They propose that the transition state pentacoordinate phosphoryl group is stabilized by these motifs during phosphoryl transfer. Thus, both the model proposed in this chapter and the Bernstein model predict that arginine-38 and the N-terminus of helix 14 bind a phosphate group during phosphoryl transfer. This difference is which phosphate group is held at the terminus of helix 14 there.

Bernstein et al. predict that the γ -phosphate of ATP is transferred to the carboxyl group of a basic patch-bound 3-PG via the stabilization provided by helix 14 and arginine-38. This phosphate becomes the 1-phosphate group of the product 1,3-BPG. However, this leaves the catalyzed 3-PG/1,3-BPG bound to the basic patch during phosphoryl transfer. It is difficult to explain why mutating the basic patch residues binding the 3-phosphate group of the catalyzed species would not prevent catalysis. After all, the 3-PG substrate must be aligned appropriately to accept a phosphate group.

In contrast, the model presented in this chapter predicts that helix-14 and arginine-38 bind the 3-phosphate group of 3-PG, not the γ -phosphate of ATP. This model predicts that the only basic patch residue binding the substrate 3-PG is arginine-38. Thus, this model agrees well with the mutagenesis data which indicate that arginine-38 is the only basic patch residue critical for enzymatic function.

E.4. Anion Activation and Trypanosomal PGK

Finally, there is the matter of anion activation and non-linear kinetics. The non-linear kinetics of PGK are difficult to explain in terms of a single binding site for 3-PG. Bernstein et al. imply that substrate binding activates the enzyme. But how can substrate binding to a catalytic site activate an enzyme? If there is only one binding site for 3-PG, then the substrate is either bound or it is not. Why then should the kinetic plots be non-linear?

The activatory effects of non-substrate anions such as sulfate must also be explained. Mutagenesis data indicate that anions, both substrate and non-substrate, require the basic patch residues to activate wild-type PGK. So, how can both an activatory sulfate ion and a substrate 3-PG occupy the basic patch at the same time? Also, it is difficult to understand how mutagenesis of residues in this region could dramatically effect anion activation without dramatically reducing catalytic activity. Unfortunately, Bernstein et al. address neither the kinetic evidence for multiple binding sites for 3-PG nor the issue of non-substrate anion activation.

Thus, the closed conformation of PGK alone cannot rescue the hinge-bending hypothesis. However, the new conformation must play a part in any new models of PGK function that are proposed. The structure may indeed represent the catalytic conformation of PGK. But, if this is the case, many questions remain to be answered. And, as always, the interpretation of crystal structures must be performed in light of the available biochemical data.

References

- Bernstein, B.E., Michels, P.A.M., Hol, W.G.J. (1997) Synergistic effects of substrate-induced conformational changes in phosphoglycerate kinase activation. *Nature* **385**:275-278.
- Blake, C.C.F., Rice, D.W. (1981) Phosphoglycerate kinase. *Phil. Trans. R. Soc. Lond. A* **293**: 93-104.
- Fairbrother, W.J., Graham, H.C., Williams, R.J.P. (1990) An NMR study of anion binding to yeast phosphoglycerate kinase. *Eur. J. Biochem.* **190**:161-169.
- Haran, G., Haas, E., Szpikowska, B.K., Mas, M.T. (1992) Domain motions in phosphoglycerate kinase: determination of interdomain distance distributions by site-specific labeling and time-resolved fluorescence energy transfer. *Proc. Natl. Acad. Sci. USA* **89**:11764-11768.
- Harlos, K., Vas, M., Blake, C.F. (1992) Crystal structure of the binary complex of pig muscle phosphoglycerate kinase and its substrate 3-phospho-D-glycerate. *Proteins Struct. Func. Genet.* **12**:133-144.
- Kraulis, P.J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**:946-950.
- Schulz, G.E. (1992) Binding of nucleotides by proteins. *Current Op. Struct. Biol.* **2**:61-67.
- Scopes, R.K. (1978) The steady-state kinetics of yeast phosphoglycerate kinase. Anomalous kinetic plots and the effects of salts on activity. *Eur. J. Biochem.* **85**:503-516.
- Sherman, M.A., Fairbrother, W.J., Mas, M.T. (1992) Characterization of the structure and properties of the His 62→Ala and Arg 38→Ala mutants of yeast phosphoglycerate kinase: An investigation of the catalytic and activatory sites by site-directed mutagenesis and NMR. *Protein Sci.* **1**:1-9.
- Vonrhein, C., Schlauderer, G.J., Schulz, G.Z. (1995) Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinase. *Structure* **3**, 483-490.

Table 4.1 Comparison of the rotation and translation axes optimal for domain closure with the observed rotations and translations.^a

	Optimal Direction	Observed Direction	Angular Deviation
<i>First 8° of rotation:</i>			
Rotation axis	$\phi = -109^\circ, \psi = 160^\circ$	$\phi = -82^\circ, \psi = 89^\circ$	73°
Translation vectors			
(inter-centroid)	$\phi = 129^\circ, \psi = 101^\circ$	$\phi = 154^\circ, \psi = 105^\circ$	26°
(inter-substrate)	$\phi = 138^\circ, \psi = 98^\circ$		22°
<i>Last 8° of rotation:</i>			
Rotation axis	$\phi = -131^\circ, \psi = 132^\circ$	$\phi = -97^\circ, \psi = 133^\circ$	27°
Translation axes			
(inter-centroid)	$\phi = 127^\circ, \psi = 102^\circ$	$\phi = 170^\circ, \psi = 48^\circ$	77°
(inter-substrate)	$\phi = 135^\circ, \psi = 94^\circ$		69°
<i>Full Rotation:</i>			
Rotation axis	$\phi = -109^\circ, \psi = 160^\circ$	$\phi = -91^\circ, \psi = 113^\circ$	48°
Translation axes			
(inter-centroid)	$\phi = 129^\circ, \psi = 101^\circ$	$\phi = 179^\circ, \psi = 55^\circ$	67°
(inter-substrate)	$\phi = 138^\circ, \psi = 98^\circ$		59°

^aThe optimal translation axis for domain closure is defined as a vector directed from the centroid of the C-domain towards the centroid of the N-domain. The optimal translation axis for joining the basic patch with the nucleotide binding pocket is defined as the vector from the γ -phosphate of nucleotide on the C-domain to the carboxyl group of 3-PG on the N-domain. The optimal rotation axis is defined as a vector normal to the two optimal translation vectors.^a

Table 4.2 *Interprobe distance distribution parameters for yeast PGK mutant (C97S, Q135C, S290C) modified with 5-(2-((iodoacetyl)-aminoethyl)amino)naphthalene-1-sulfonic acid (IAEDANS) and 5-iodoacetamidofluorescein (IAF) (Haran et al., 1992). Standard deviations are given in parentheses. The last column gives the distance between residues 135 and 290 in crystal structures of PGK complexed to the combination of substrates listed in column 1. There are no crystal structures to compare with the first and third mean distances.*

Substrates, analogs, and effectors present	Distribution Full Width at Half Maximum (Å)	Mean Distance in Solution (Å)	Mean Distance in Crystal Structure(Å)
None	31 (2)	31.5 (0.5)	
3-PG	24 (2)	35.2 (0.5)	35.5 (porcine PGK + 3-PG)
ATP	21 (2)	38.3 (0.4)	
AMP-PNP, 3-PG	20 (2)	38.2 (0.5)	38.7 (yeast R65Q PGK + Mg-AMP-PNP + 3-PG)

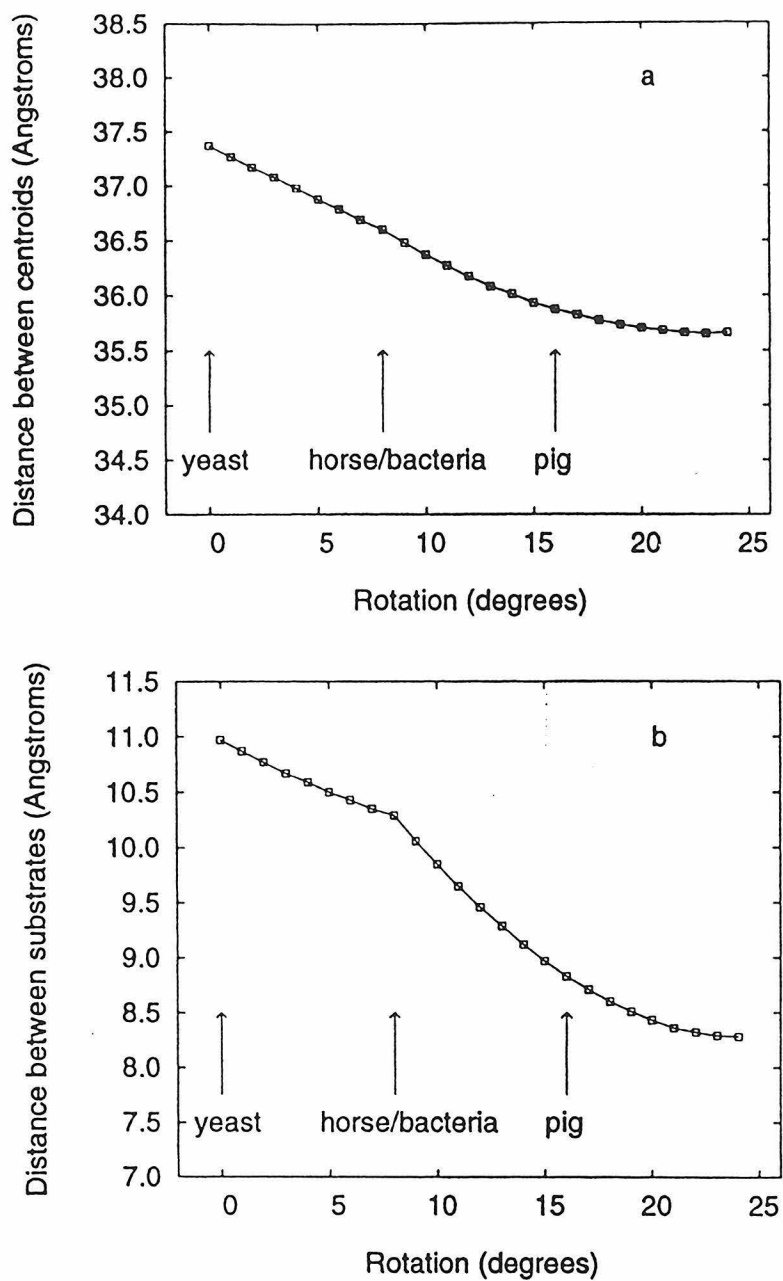


Figure 4.1 (a) Distance between the centroids of the N and C-domains of the transformed coordinates of yeast R65Q PGK as a function of rotation angle. (b) Distance between substrates in the transformed coordinates of yeast R65Q PGK as a function of rotation angle.

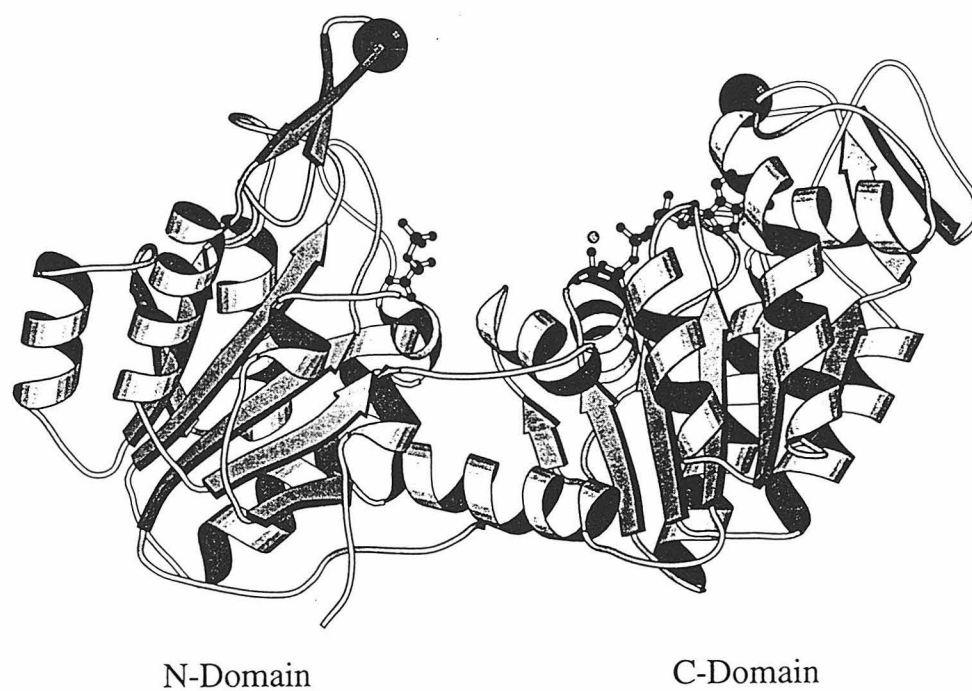


Figure 4.2 *Location of probes for fluorescent energy transfer experiments (Haran et al., 1992). The two large spheres represent residues 135 and 290 on the N- and C-domains, respectively. See Table 4.2 for details.*

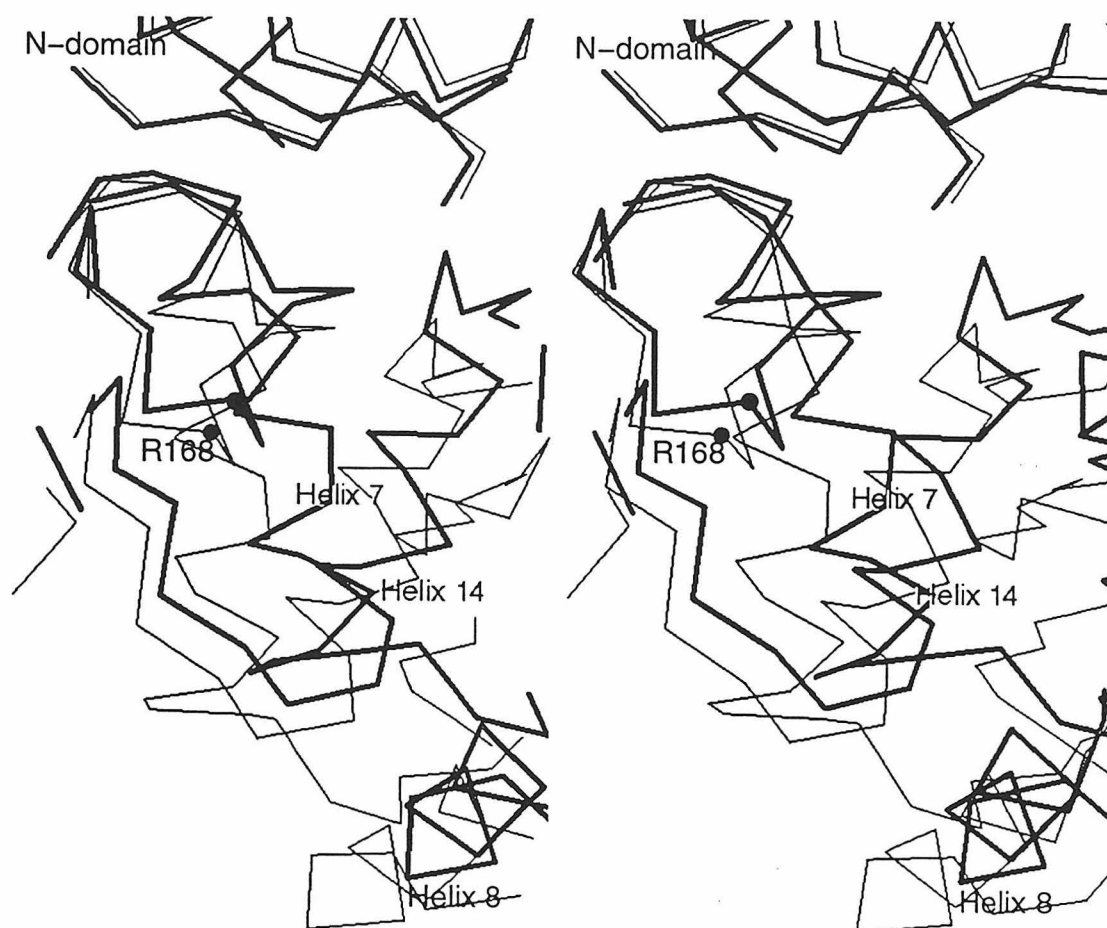


Figure 4.3 Superposition of the $C\alpha$ traces of the yeast R65Q (thin line) and porcine (thick lines) PGK structures showing the interdomain region. Only the N-domain residues were included in the least-squares superposition.

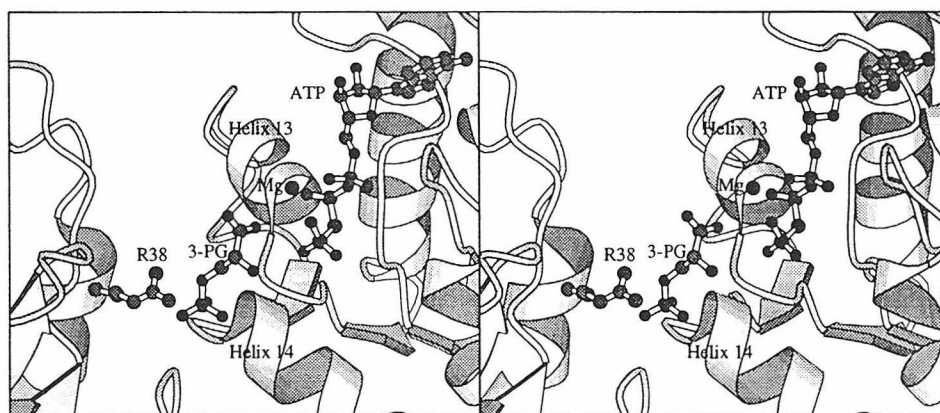


Figure 4.4 *The result of modeling 3-PG into the porcine PGK coordinates adjacent to the N-terminus of helix 14. The protein structure shown, including the conformation of R38, is based on the refined coordinates of porcine PGK (Harlos et al., 1992). The nucleotide coordinates result from superpositioning the R65Q yeast PGK coordinates on the C-domain of the porcine coordinates and transferring the Mg-AMP-PNP coordinates to the porcine model. The 3-PG molecule was positioned by hand. Drawn with MOLSCRIPT (Kraulis, 1991).*

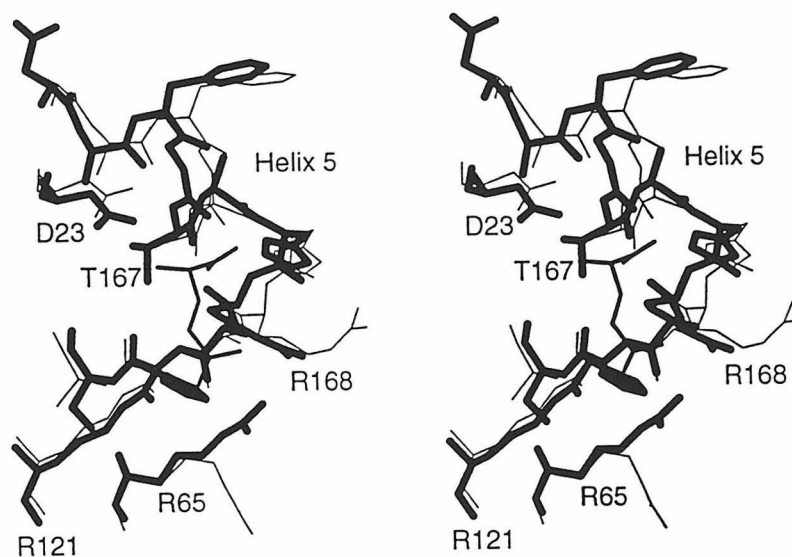


Figure 4.5 Shift of helix 5 between the bacterial (thin lines) and porcine PGK structures (thick lines). Pertinent residues are labeled according to the yeast sequence. The 3-PG molecule is shown as observed in the porcine structure. Drawn with MOLSCRIPT (Kraulis, 1991).

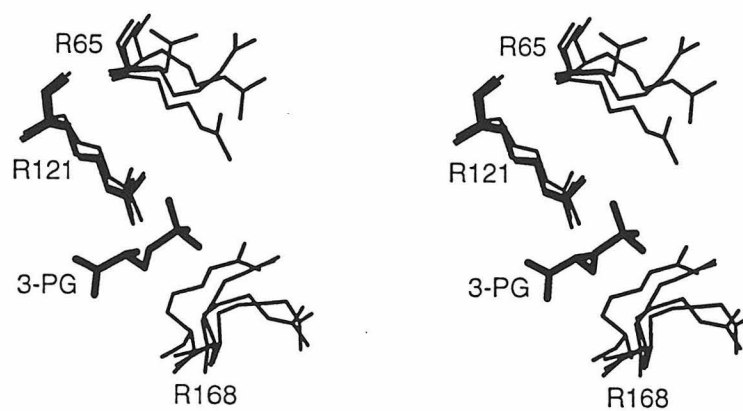


Figure 4.6 Positions of arginines 65, 121, and 168 in the yeast R65Q, horse, pig, and bacterial PGK structures. The 3-PG molecule is shown as observed in the yeast R65Q structure. For residues 65 and 168, the porcine conformations are those that most closely approach the 3-PG. Drawn with MOLSCRIPT (Kraulis, 1991).

Appendix 1

Command and Geometry Definition Files for Refinement of PGK Using TNT

Refinement Command File (XYZREF_5A.COM):

```

$!
$!      XYZREF_5A.COM
$!
$!      FOR TNT VERSION 5A
$!
$ SET NOVERIFY
$ !
$ !   Initialize the process.
$ !
$ @HOME:TNTLOGICALS
$ ON ERROR THEN GOTO BADNEWS
$ !
$ @TNT_COMMAND:SETUP
$ !
$ SET DEF USER3:[TIM.PGK1.TNT]
$ !
$ ! ***** First the R factor gradient *****
$ !
$ !   CALCULATE STRUCTURE FACTORS USING INITIAL COORDINATES
$ !
$ !
$ !   Now run Rfactor.
$ !
$ ass/user sys$input for001
$ RUN REFROOT:[tnt_5a.RFACTOR]RFACTOR
CELL 96.0 70.1 82.3 90.0 122.2 90.0
RESOLUTION 20.0 2.7
FO PGK1:PGK1.HKL
INCLUDE PGK1:WEIGHTS.DAT
INCLUDE PGK1:C2.DAT
INCLUDE PGK1:INIT.COR
INCLUDE TNT_DATA:FORMFACTOR.DAT
set b 0
REPORT RFACTOR
curvature RFACTOR.DAT
$ !
$ !
$ ! ***** Now the Bcorrelation Part *****
$ !
$ RUN REFROOT:[TNT_5A.BCORREL]BCORREL
INCLUDE PGK1:INIT.COR
INCLUDE PGK1:PGK1_MODEL_7.SEQ
INCLUDE REFROOT:[TNT_5A.BCORREL]BCORREL.GEO
INCLUDE PGK1:WEIGHTS.DAT
curvature BCORREL.DAT
$ !
$ purge
$ ! ***** Now the geometry gradient *****
$ !
$ ass/user sys$input for001

```

```

$ RUN REFROOT:[TNT 5A.GEOMETRY]GEOMETRY
CELL 96.0 70.1 82.3 90.0 122.2 90.0
INCLUDE PGK1:INIT.COR
INCLUDE PGK1:PGKGEO.DAT
INCLUDE PGK1:WEIGHTS.DAT
curvature GEOMETRY.DAT
$ !
$ WRITE      SYS$OUTPUT  "Weights used for refinement:"
$ TYPE PGK1:WEIGHTS.DAT
$ !
$ !
$ ! ***** End of modules and into control program *****
$ !
$ ! Now move the atoms as directed.  Makes a new SHIFTED.cor
$ !
$ ass/user sys$input for001
$ RUN REFROOT:[tnt_5a.SHIFT]SHIFT
INCLUDE      RFACTOR.DAT
INCLUDE      GEOMETRY.DAT
INCLUDE      BCORREL.DAT
INCLUDE      PGK1:INIT.COR
CONSTANT OCC
$ !
$ ! More clean up.
$ !
$ DELETE      OLDDIR.DAT; ,GEOMETRY.DAT; ,RFACTOR.DAT;
$ purge
$ ! *****
$ !
$ ! Loop around this short refinement loop.
$ !
$ LOOPCOUNT == 0
$ !
$ START:
$ LOOPCOUNT == LOOPCOUNT + 1
$ !
$ !
$ ass/user sys$input for001
$ RUN REFROOT:[tnt_5a.RFACTOR]RFACTOR
CELL 96.0 70.1 82.3 90.0 122.2 90.0
RESOLUTION 20.0 2.7
FO PGK1:PGK1.HKL
INCLUDE RCTET:WEIGHTS.DAT
INCLUDE SCALES.DAT
INCLUDE PGK1:C2.DAT
INCLUDE SHIFTED.COR
INCLUDE TNT_DATA:FORMFACTOR.DAT
REPORT RFACTOR
FUNCTION RFACTOR.DAT
$ !
$ ! Clean up.
$ !
$ purge
$ !
$ ! ***** Now the Bcorrelation Part *****
$ !
$ RUN REFROOT:[TNT_5A.BCORREL]BCORREL
INCLUDE SHIFTED.COR
INCLUDE PGK1:PGK1_MODEL_7.SEQ
INCLUDE REFROOT:[TNT_5A.BCORREL]BCORREL.GEO
INCLUDE PGK1:WEIGHTS.DAT

```

```

FUNCTION BCORREL.DAT
$ !
$ !***** Now run geometry *****
$ !
$ ass/user sys$input for001
$ RUN REFROOT:[TNT_5A.GEOMETRY]GEOMETRY
CELL 96.0 70.1 82.3 90.0 122.2 90.0
INCLUDE SHIFTED.COR
INCLUDE PGK1:PGKGEO.DAT
INCLUDE PGK1:WEIGHTS.DAT
FUNCTION GEOMETRY.DAT
$ !
$ !***** Get rid of the old SHIFTED.cor.
$ !
$ !DELETE    SHIFTED.cor;*
$ !
$ !***** Now move the atoms as directed.
$ !
$ ass/user sys$input for001
$ ASSIGN/USER      SHIFT.OUT    SYS$OUTPUT
$ RUN REFROOT:[tnt_5a.SHIFT]SHIFT
INCLUDE      STPFIL.DAT
INCLUDE      RFACTOR.DAT
INCLUDE      GEOMETRY.DAT
INCLUDE      BCORREL.DAT
INCLUDE      PGK1:INIT.COR
INCLUDE      NEWDIR.DAT
$ !
$ TYPE        SHIFT.OUT
$ !
$ ! More clean up
$ !
$ delete RFACTOR.DAT;* , geometry.dat;*
$ PURGE
$ !
$ ! End of loop.  Are we done?
$ !
$ @TNT_COMMAND:ALLDONE          ! Returns true if the cycle is complete
$ !
$ IF .NOT. RETURN THEN GOTO START
$ !
$ ! If we are here then we are all done, notify the log file.
$ !
$ MESSAGE      := "All done.  "'LOOPCOUNT'" short cycles to settle down"
$ WRITE        SYS$OUTPUT MESSAGE
$ !
$ EXIT
$ !
$ !*****
$ BADNEWS:
$      WRITE SYS$OUTPUT "Error encounter in runstream, run aborted"
$!$      STOP  'F$GETJPI(0,"PRCNAM")'
$ EXIT
$ !

```

Geometry Definition File for AMP-PNP (PNPGEO.DAT):

GEOMETRY	PNP	BOND	1.338	0.02	N1, C2
GEOMETRY	PNP	BOND	1.332	0.02	C2, N3
GEOMETRY	PNP	BOND	1.342	0.02	N3, C4
GEOMETRY	PNP	BOND	1.382	0.02	C4, C5
GEOMETRY	PNP	BOND	1.409	0.02	C5, C6
GEOMETRY	PNP	BOND	1.349	0.02	C6, N1
GEOMETRY	PNP	BOND	1.385	0.02	C5, N7
GEOMETRY	PNP	BOND	1.312	0.02	N7, C8
GEOMETRY	PNP	BOND	1.367	0.02	C8, N9
GEOMETRY	PNP	BOND	1.376	0.02	N9, C4
GEOMETRY	PNP	BOND	1.337	0.02	C6, N6
GEOMETRY	PNP	BOND	1.46	0.02	N9, C1*
GEOMETRY	PNP	BOND	1.60	0.02	PA, O5*
GEOMETRY	PNP	BOND	1.48	0.02	PA, O1A
GEOMETRY	PNP	BOND	1.48	0.02	PA, O2A
GEOMETRY	PNP	BOND	1.60	0.02	PA, O3A
GEOMETRY	PNP	BOND	1.60	0.02	O3A, PB
GEOMETRY	PNP	BOND	1.48	0.02	PB, O1B
GEOMETRY	PNP	BOND	1.48	0.02	PB, O2B
GEOMETRY	PNP	BOND	1.60	0.06	PB, N3B
GEOMETRY	PNP	BOND	1.60	0.06	N3B, PG
GEOMETRY	PNP	BOND	1.52	0.02	PG, O1G
GEOMETRY	PNP	BOND	1.52	0.02	PG, O2G
GEOMETRY	PNP	BOND	1.52	0.02	PG, O3G
GEOMETRY	PNP	BOND	1.525	0.02	C1*, C2*
GEOMETRY	PNP	BOND	1.525	0.02	C2*, C3*
GEOMETRY	PNP	BOND	1.523	0.02	C3*, C4*
GEOMETRY	PNP	BOND	1.422	0.03	C3*, O3*
GEOMETRY	PNP	BOND	1.518	0.02	C4*, C5*
GEOMETRY	PNP	BOND	1.450	0.02	C4*, O4*
GEOMETRY	PNP	BOND	1.414	0.02	O4*, C1*
GEOMETRY	PNP	BOND	1.423	0.03	C5*, O5*
GEOMETRY	PNP	BOND	1.414	0.02	C2*, O2*
GEOMETRY	PNP	ANGLE	118.8	3	C6, N1, C2
GEOMETRY	PNP	ANGLE	129.0	3	N1, C2, N3
GEOMETRY	PNP	ANGLE	110.8	3	C2, N3, C4
GEOMETRY	PNP	ANGLE	126.9	3	N3, C4, C5
GEOMETRY	PNP	ANGLE	116.9	3	C4, C5, C6
GEOMETRY	PNP	ANGLE	117.6	3	C5, C6, N1
GEOMETRY	PNP	ANGLE	110.7	3	C4, C5, N7
GEOMETRY	PNP	ANGLE	103.9	3	C5, N7, C8
GEOMETRY	PNP	ANGLE	113.8	3	N7, C8, N9
GEOMETRY	PNP	ANGLE	105.9	3	C8, N9, C4
GEOMETRY	PNP	ANGLE	105.7	3	N9, C4, C5
GEOMETRY	PNP	ANGLE	127.4	3	N3, C4, N9
GEOMETRY	PNP	ANGLE	132.3	3	C6, C5, N7
GEOMETRY	PNP	ANGLE	119.0	3	N1, C6, N6
GEOMETRY	PNP	ANGLE	123.4	3	C5, C6, N6
GEOMETRY	PNP	ANGLE	108.4	3	N9, C1*, O4*
GEOMETRY	PNP	ANGLE	113.5	3	N9, C1*, C2*
GEOMETRY	PNP	ANGLE	106.4	3	O4*, C1*, C2*
GEOMETRY	PNP	ANGLE	101.6	3	C1*, C2*, C3*
GEOMETRY	PNP	ANGLE	102.4	3	C2*, C3*, C4*
GEOMETRY	PNP	ANGLE	111.5	3	C2*, C3*, O3*

GEOMETRY	PNP	ANGLE	110.0	3	O3*, C3*, C4*
GEOMETRY	PNP	ANGLE	114.9	3	C3*, C4*, C5*
GEOMETRY	PNP	ANGLE	105.4	3	C3*, C4*, O4*
GEOMETRY	PNP	ANGLE	109.4	3	O4*, C4*, C5*
GEOMETRY	PNP	ANGLE	109.7	3	C4*, O4*, C1*
GEOMETRY	PNP	ANGLE	110.8	3	C4*, C5*, O5*
GEOMETRY	PNP	ANGLE	120	4	C5*, O5*, PA
GEOMETRY	PNP	ANGLE	108	3	O5*, PA, O1A
GEOMETRY	PNP	ANGLE	108	3	O5*, PA, O2A
GEOMETRY	PNP	ANGLE	120	3	O5*, PA, O3A
GEOMETRY	PNP	ANGLE	110	3	O1A, PA, O2A
GEOMETRY	PNP	ANGLE	106	3	O1A, PA, O3A
GEOMETRY	PNP	ANGLE	106	3	O2A, PA, O3A
GEOMETRY	PNP	ANGLE	108	3	O3A, PB, O1B
GEOMETRY	PNP	ANGLE	108	3	O3A, PB, O2B
GEOMETRY	PNP	ANGLE	120	3	O3A, PB, N3B
GEOMETRY	PNP	ANGLE	107	3	O1B, PB, O2B
GEOMETRY	PNP	ANGLE	106	3	O1B, PB, N3B
GEOMETRY	PNP	ANGLE	106	3	O2B, PB, N3B
GEOMETRY	PNP	ANGLE	112	3	N3B, PG, O1G
GEOMETRY	PNP	ANGLE	112	3	N3B, PG, O2G
GEOMETRY	PNP	ANGLE	112	3	N3B, PG, O3G
GEOMETRY	PNP	ANGLE	107	3	O1G, PG, O2G
GEOMETRY	PNP	ANGLE	107	3	O1G, PG, O3G
GEOMETRY	PNP	ANGLE	107	3	O2G, PG, O3G
GEOMETRY	PNP	ANGLE	120	3	PA, O3A, PB
GEOMETRY	PNP	ANGLE	120	6	PB, N3B, PG
GEOMETRY	PNP	ANGLE	110.2	3	C1*, C2*, O2*
GEOMETRY	PNP	ANGLE	112.4	3	O2*, C2*, C3*
GEOMETRY	PNP	TORS	3055	20	O5*, C5*, C4*, C3* (GAMMA)
GEOMETRY	PNP	TORS	3120	30	C5*, C4*, C3*, O3* (DELTA)
GEOMETRY	PNP	TORS	3000	40	C4*, O4*, C1*, C2* (NU 0)
GEOMETRY	PNP	TORS	3000	40	O4*, C1*, C2*, C3* (NU 1)
GEOMETRY	PNP	TORS	3000	40	C1*, C2*, C3*, C4* (NU 2)
GEOMETRY	PNP	TORS	3000	40	C2*, C3*, C4*, O4* (NU 3)
GEOMETRY	PNP	TORS	3000	40	C3*, C4*, O4*, C1* (NU 4)
GEOMETRY	PNP	TORS	-2150	40	O4*, C1*, N9, C4 (CHI) PURINE
GEOMETRY	PNP	PLANE	11	0.02	N1,C2,N3,C4,C5,C6,N6,N7,C8,N9,C1*
GEOMETRY	PNP	CHIRAL	1	1	C3*, C4*, O3*, C2*
GEOMETRY	PNP	CHIRAL	1	1	C4*, C5*, O4*, C3*
GEOMETRY	PNP	CHIRAL	1	1	C2*, C3*, O2*, C1*
GEOMETRY	PNP	CHIRAL	1	1	C1*, O4*, N9, C2*

Appendix 2

Command and Geometry Definition Files for Refinement of PGK Using X-PLOR

Positional Refinement Script (xyzref):

```

#! /bin/sh
# xyzref -- shell script to refine xyz coordinates of
# protein model with xplor
# usage: xyzref inputpdb outputpdb resolution cycles

xplor31 << END

parameter    @parhcsdx_jp.pro
              @substrates.par
              @water.par

end

structure    @generate.psf end
coordinates  @$1

vector do ( charge=0.0 ) ( resname LYS and
    ( name ce or name nz or name hz* ) )
vector do ( charge=0.0 ) ( resname GLU and
    ( name cg or name cd or name oe* ) )
vector do ( charge=0.0 ) ( resname ASP and
    ( name cb or name cg or name od* ) )
vector do ( charge=0.0 ) ( resname ARG and
    ( name cd or name *E or name cz or name NH* or name HH* ) )

flags
    include pele pvdw xref    { In addition to the empirical    }
    ?                         { potential energy terms which    }
end                           { are turned on initially.  This }
                              { statement turns on the      }
                              { crystallographic residual term }
                              { and packing term.              }

xrefine

a=96.00 b=70.10 c=82.30 alpha=90.0 beta=122.2 gamma=90.0

symmetry=(x,y,z)
symmetry=(-x,y,-z)
symmetry=(x+1/2,y+1/2,z)
symmetry=(-x+1/2,y+1/2,-z)

SCATter ( chemical C* )
    2.31000 20.8439 1.02000 10.2075
    1.58860 .568700 .865000 51.6512 .215600
SCATter ( chemical N* )
    12.2126 .005700 3.13220 9.89330

```



```

2.01250 28.9975 1.16630 .582600 -11.529
SCATter ( chemical O* )
3.04850 13.2771 2.28680 5.70110
1.54630 .323900 .867000 32.9089 .250800
SCATter ( chemical S* )
6.90530 1.46790 5.20340 22.2151
1.43790 .253600 1.58630 56.1720 .866900
SCATter ( chemical P* )
6.43450 1.90670 4.17910 27.1570
1.78000 0.52600 1.49080 68.1645 1.11490
SCATter ( chemical MG* )
3.49880 2.16760 3.83780 4.75420
1.32840 0.18500 0.84970 10.1411 .48530

nreflections=17000
reflection @pgkl.fob end

resolution 8.0 $3

reduce

fwind=0.001=100000

method=FFT          { Use the FFT method }

fft
memory=1000000      { This tells the FFT routine how much memory }
end                  { is available; refers to DOUBLE COMPLEX      }
                    { words, the memory is allocated from HEAP    }

tolerance=0.0 lookup=false

wa=360000

end                  { Terminates the diffraction data parser      }

minimize powell

nstep=$4
drop=10.0

end

write coordinates output=$2 end

xrefin
mbins=10
resolution 8.0 $3
update
print R
end

constraints interaction (all) (all) end
print threshold=0.1 bonds
print threshold=15.0 angles
print threshold=90.0 dihedrals
print threshold=20.0 impropers

stop

END

```

Temperature Factor Refinement Script (bref):

```

#!/bin/sh
# bref -- shell script to refine b-factors of protein model with xplor
# usage: bref inputpdb outputpdb resolution cycles

xplor31 << END

parameter    @parhcsdx_jp.pro
              @substrates.par
              @water.par
end

structure    @generate.psf end
coordinates @$1

xrefine

  selection=(known)

  a=96.00 b=70.10 c=82.30 alpha=90.0 beta=122.2 gamma=90.0

  symmetry=(x,y,z)
  symmetry=(-x,y,-z)
  symmetry=(x+1/2,y+1/2,z)
  symmetry=(-x+1/2,y+1/2,-z)

  SCATter ( chemical C* )
    2.31000 20.8439 1.02000 10.2075
    1.58860 .568700 .865000 51.6512 .215600
  SCATter ( chemical N* )
    12.2126 .005700 3.13220 9.89330
    2.01250 28.9975 1.16630 .582600 -11.529
  SCATter ( chemical O* )
    3.04850 13.2771 2.28680 5.70110
    1.54630 .323900 .867000 32.9089 .250800
  SCATter ( chemical S* )
    6.90530 1.46790 5.20340 22.2151
    1.43790 .253600 1.58630 56.1720 .866900
  SCATter ( chemical P* )
    6.43450 1.90670 4.17910 27.1570
    1.78000 0.52600 1.49080 68.1645 1.11490
  SCATter ( chemical MG* )
    3.49880 2.16760 3.83780 4.75420
    1.32840 0.18500 0.84970 10.1411 .48530

  nreflections=17000
  reflection @pgkl.fob end

  resolution 8.0 $3

  reduce
  !do amplitude ( fobs = fobs * heavy(fobs - 2.0*sigma))
  fwind=0.001=100000
  method=FFT          { Use the FFT method          }

  fft

```

```

        memory=1000000    { This tells the FFT routine how much memory }
end                      { is available; refers to DOUBLE COMPLEX    }
                        { words; memory is allocated from the HEAP    }
tolerance=0.0  lookup=false

wa=800000

end                      { Terminates the diffraction data parser    }

xrefin
  update
  optimize overall
    drop=0.01
    nstep=15
  end
end

vector do (b=b+gauss(1.5)) (name ca or name n or name c )
vector do (b=b+gauss(2.0)) (not hydrogen and not
                           (name ca or name n or name c))

xrefin
  optimize bfactors
    nstep=$4
    drop=10.0
    bsigma=(name ca or name n or name c or name o)=1.5
    bsigma=(not (name ca or name n or name c or name o))=2.0
    asigma=(name ca or name n or name c or name o)=2.0
    asigma=(not (name ca or name n or name c or name o))=2.0
  end
end

write coordinates output=$2 end

xrefin
  mbins=10
  resolution 8.0 $3
  update
  print R
end

vector show ave ( b ) ( all )
vector show ave ( b ) ( resid 1:184 )
vector show ave ( b ) ( resid 200:393 )
vector show ave ( b ) ( resid 1:184 and
                      ( name ca or name n or name c or name o or name hydrogen ) )
vector show ave ( b ) ( resid 200:393 and
                      ( name ca or name n or name c or name o or name hydrogen ) )

stop

END

```

Topology File for Mg-AMP-PNP and 3-PG (substrates.top):

```
autogenerate angles=true end
```

```
set echo=false end
```

```
mass  CN    12.01100!  corresp. to C
mass  CH    13.01900!  corresp. to CH1E
mass  C2    14.02700!  corresp. to CH2E
mass  CA    12.01100!  carbon in ring bound to two or three nitrogens
mass  CB    12.01100!  carbon in ring ("bridge" e.g. in GUA)
mass  CE    13.01900!  extended atom carbon in (5) ring with one H
mass  CF    13.01900!  extended atom carbon in (6) ring with one H
mass  C3    15.03500!  corresp. to CH3E
mass  CS    12.01100!  carbon in ring >C-C in THY
mass  N2    14.00670!  nitrogen in -NH2
mass  NA    14.00670!  corresp. to NH1
mass  NB    14.00670!  corresp. to NR in (5) ring
mass  NC    14.00670!  corresp. to NR
mass  NS    14.00670!  nitrogen in ring >N-
mass  ON    15.99940!  corresp. to O
mass  O2    15.99940!  O in phosphate
mass  OS    15.99940!  ester -P-O-C-
mass  OH    15.99940!  corresp. to OH1
mass  P     30.97400!  phosphorus
mass  MG    24.31200!  magnesium
```

```
mass  O113PG 15.99940
mass  C13PG  12.01100
mass  C23PG  13.01900
mass  C33PG  14.02700
mass  OP43PG 15.99940
mass  OP13PG 15.99940
mass  O213PG 15.99940
mass  P3PG   30.97400
```

```
residue 3PG
```

```
group
  atom O1    type=O113PG    charge=0.0    end
  atom O2    type=O113PG    charge=0.0    end
  atom C1    type=C13PG     charge=0.0    end
```

```
group
  atom C2    type=C23PG     charge=0.0    end
  atom C3    type=C33PG     charge=0.0    end
  atom O3    type=O213PG    charge=0.0    end
```

```
group
  atom P      type=P3PG      charge=0.0    end
  atom O1P    type=OP43PG    charge=0.0    end
  atom O2P    type=OP13PG    charge=0.0    end
  atom O3P    type=OP13PG    charge=0.0    end
  atom O4P    type=OP13PG    charge=0.0    end
```

```
bond C1 O1
bond C1 O2
bond C1 C2
bond C2 O3
bond C2 C3
bond C3 O1P
bond O1P P
bond P O2P
```

```

bond P    O3P
bond P    O4P

improper C1  O1 O2 C2

end {3PG}

residue MAP

group
atom MG      type=MG      charge=2.00      end
group
atom PA      type=P       charge=0.00      end
atom O1A     type=O2      charge=-.50      end
atom O2A     type=O2      charge=-.50      end
atom O3A     type=O2      charge=0.00      end
group
atom PB      type=P       charge=0.00      end
atom O1B     type=O2      charge=-.50      end
atom O2B     type=O2      charge=-.50      end
atom N3B     type=N2      charge=0.00      end
group
atom PG      type=P       charge=0.00      end
atom O1G     type=O2      charge=-.33      end
atom O2G     type=O2      charge=-.33      end
atom O3G     type=O2      charge=-.33      end
group
atom O5*     type=OS      charge=-0.36     END
group
atom C5*     type=C2      charge=0.00      END
group
atom C4*     type=CH      charge=0.10      END
atom O4*     type=OS      charge=-0.30     END
atom C1*     type=CH      charge=0.20      END
group
atom N9      type=NS      charge=-0.19     END
atom C4      type=CB      charge=0.19      EXCLusion=( N1 )  END
group
atom N3      type=NC      charge=-0.26     EXCLusion=( C6 )  END
atom C2      type=CE      charge=0.26      EXCLusion=( C5 )  END
group
atom N1      type=NC      charge=-0.28     END    !H
atom C6      type=CA      charge=0.28      END    !H
group
atom N6      type=N2      charge=-0.42     END    !H
group
atom C5      type=CB      charge=0.02      END
atom N7      type=NB      charge=-0.25     END
atom C8      type=CE      charge=0.23      END
group
atom C2*     type=CH      charge=0.15      END
atom O2*     type=OH      charge=-0.40     END
group
atom C3*     type=CH      charge=0.00      END
group
atom O3*     type=OS      charge=-0.36     end

bond N1 C2      bond C2 N3      bond N3 C4      bond C4 C5
bond C5 C6      bond C6 N1      bond C5 N7      bond N7 C8

```

```

bond C8 N9      bond N9 C4      bond C6 N6      bond N9 C1*
bond PA O5*     bond PA O1A     bond PA O2A     bond PA O3A
bond O3A PB     bond PB O1B     bond PB O2B     bond PB N3B
bond N3B PG     bond PG O1G     bond PG O2G     bond PG O3G
bond C1* C2*    bond C2* C3*    bond C3* C4*    bond C3* O3*
bond C4* C5*    bond C4* O4*    bond O4* C1*    bond C5* O5*
bond C2* O2*    bond MG O2A     bond MG O1B     bond MG O1G

```

```

dihedral PA O5* C5* C4*
dihedral O5* C5* C4* C3*
dihedral C4* O4* C1* C2*
dihedral C1* C2* C3* C4*
dihedral C5* C4* C3* C2*
dihedral O4* C1* N9 C4
dihedral C8 C4 C5 N1
dihedral N3 C4 C5 N7
improper C5* O4* C3* C4*
improper N9 C2* O4* C1*
improper N9 C4 C5 N7
improper C5 N7 C8 N9
improper C8 N9 C4 C5
improper C4 N3 C2 N1
improper N3 C2 N1 C6
improper N1 C6 C5 C4
improper C5 C4 N3 C2

```

```
end {MAP}
```

```
set echo=true end
```

```

dihedral O5* C5* C4* O4*
dihedral C3* C4* O4* C1*
dihedral O4* C1* C2* C3*
dihedral O4* C4* C3* O3*
dihedral O2* C2* C3* O3*

dihedral C8 C5 C4 C2
dihedral C6 C5 C4 N9
improper O3* C2* C4* C3*
improper C1* C4 C8 N9
improper C4 C5 N7 C8
improper N7 C8 N9 C4
improper N6 N1 C5 C6

improper C2 N1 C6 C5
improper C6 C5 C4 N3
improper C2* C3* C1* O2*

```

Parameter File for Mg-AMP-PNP and 3-PG (substrates.par):

```
set echo=false end
```

bond	P	N2	200.0000	1.8000
bond	P	O2	528.0000	1.4800
bond	P	OS	237.0000	1.6100
bond	P	OH	237.0000	1.6100
bond	OS	C2	292.0000	1.4300
bond	OS	CH	292.0000	1.4300
bond	ON	CN	612.0000	1.2290
bond	OH	C2	292.0000	1.4300
bond	OH	CH	292.0000	1.4300
bond	C2	CH	201.0000	1.5250
bond	C2	C2	201.0000	1.5250
bond	CH	CH	201.0000	1.5250
bond	C3	CH	201.0000	1.5250
bond	C3	C2	201.0000	1.5250
bond	CB	CB	550.0000	1.3700
bond	CA	CB	386.0000	1.4040
bond	CN	CB	386.0000	1.4190
bond	CF	CN	386.0000	1.4440
bond	CF	CA	386.0000	1.4330
bond	CF	CF	684.0000	1.3500
bond	CN	CS	386.0000	1.4440
bond	CS	C3	201.0000	1.5250
bond	CS	CF	386.0000	1.3430
bond	NS	CH	300.0000	1.4750
bond	NS	CB	337.0000	1.3740
bond	NS	CE	300.0000	1.3710
bond	NS	CN	337.0000	1.3830
bond	NS	CF	337.0000	1.3650
bond	NC	CB	400.0000	1.3540
bond	NC	CE	400.0000	1.3240
bond	NC	CA	400.0000	1.3390
bond	NC	CN	400.0000	1.3580
bond	N2	CA	400.0000	1.3330
bond	NB	CB	400.0000	1.3910
bond	NB	CE	400.0000	1.3040
bond	NA	CA	400.0000	1.3810
bond	NA	CN	337.0000	1.3880
bond	NH2E	CH	422.0000	1.49
bond	BR	CS	300.0000	1.880
bond	O2	MG	50.0000	2.300
bond	O2	C2	300.0000	1.240
bond	PA	O5*	200.0000	1.6000
bond	PA	O1A	200.0000	1.4800
bond	PA	O2A	200.0000	1.4800
bond	PA	O3A	100.0000	1.7500
bond	O3A	PB	100.0000	1.7500
bond	PB	O1B	200.0000	1.4800
bond	PB	O2B	200.0000	1.4800
bond	PB	N3B	100.0000	1.8000
bond	N3B	PG	100.0000	1.8000
bond	PG	O1G	200.0000	1.5200
bond	PG	O2G	200.0000	1.5200
bond	PG	O3G	200.0000	1.5200
bond	O113PG	C13PG	200.00	1.23

bond	C13PG	C23PG	200.00	1.52
bond	C23PG	O213PG	200.00	1.46
bond	C23PG	C33PG	100.00	1.50
bond	C33PG	OP43PG	100.00	1.43
bond	OP43PG	P3PG	100.00	1.66
bond	OP13PG	P3PG	200.00	1.51

angle	P	O2	P	47.6000	120.0000
angle	O2	P	N2	47.6000	130.0000
angle	P	N2	P	47.6000	120.0000
angle	OS	CH	OS	47.6000	111.0000
angle	CH	OS	C2	47.6000	111.5000
angle	O2	C2	O2	47.6000	120.0000
angle	CH	C2	O2	47.6000	120.0000
angle	C2	CH	C2	47.6000	109.0000

angle	O2	MG	O2	20.0000	90.0000
angle	P	O2	MG	20.0000	109.0000
angle	P	OS	C2	47.6000	120.5000
angle	P	OS	C3	47.6000	120.5000
angle	CH	OS	CH	46.5000	111.5000
angle	P	OS	CH	46.5000	120.5000
angle	OS	P	OS	48.1000	102.6000
angle	OH	P	OS	48.1000	102.6000
angle	OS	P	O2	98.9000	108.2300
angle	OH	P	O2	98.9000	108.2300
angle	O2	P	O2	140.5000	119.9000
angle	CN	NA	CN	116.0000	126.4000
angle	CA	NA	CN	116.0000	125.2000
angle	CB	NB	CE	70.0000	103.8000
angle	CB	NC	CE	70.0000	111.0000
angle	CA	NC	CE	70.0000	118.6000
angle	CN	NC	CA	70.0000	120.5000
angle	CA	NC	CB	70.0000	112.2000
angle	CH	NS	CB	70.0000	125.8000
angle	CH	NS	CE	70.0000	128.8000
angle	CB	NS	CE	116.0000	105.4000
angle	CN	NS	CH	70.0000	117.6000
angle	CF	NS	CH	70.0000	121.2000
angle	CN	NS	CF	116.0000	121.6000
angle	NS	CN	ON	70.0000	120.9000
angle	NA	CN	ON	70.0000	120.6000
angle	CB	CN	ON	86.0000	128.8000
angle	NC	CN	ON	70.0000	122.5000
angle	CF	CN	ON	86.0000	125.3000
angle	CS	CN	ON	86.0000	125.3000
angle	CF	CN	NA	62.0000	114.1000
angle	CS	CN	NA	62.0000	114.1000
angle	NS	CN	NA	62.0000	115.4000
angle	CB	CN	NA	62.0000	111.3000
angle	NS	CN	NC	70.0000	118.6000
angle	N2	CA	NB	70.0000	118.3000
angle	N2	CA	NC	70.0000	119.8000
angle	CB	CA	NC	70.0000	117.3000
angle	CB	CA	N2	70.0000	123.5000
angle	CF	CA	NC	70.0000	121.5000
angle	N2	CA	CF	70.0000	120.1000
angle	NC	CA	NA	70.0000	123.3000
angle	N2	CA	NA	70.0000	116.0000
angle	CB	CB	NS	70.0000	106.2000
angle	CB	CB	NA	70.0000	106.2000

angle	NC	CB	NS	70.0000	126.0000
angle	NC	CB	CB	116.0000	127.7000
angle	CA	CB	NB	116.0000	132.4000
angle	CA	CB	CB	116.0000	117.3000
angle	CB	CB	NB	116.0000	110.4000
angle	CN	CB	NB	116.0000	130.0000
angle	CN	CB	CB	116.0000	119.2000
angle	NC	CE	NC	70.0000	129.1000
angle	NB	CE	NS	70.0000	113.9000
angle	NS	CF	CF	116.0000	121.2000
angle	NA	CF	CF	116.0000	121.2000
angle	CA	CF	CF	150.0000	117.0000
angle	CS	CF	NS	150.0000	121.2000
angle	CN	CF	CF	150.0000	120.7000
angle	C2	CH	OS	46.5000	111.0000
angle	C2	CH	OH	46.5000	111.0000
angle	CH	CH	OH	46.5000	111.0000
angle	C2	CH	CH	46.5000	111.0000
angle	CH	CH	CH	46.5000	111.0000
angle	CH	CH	OS	46.5000	111.0000
angle	C3	CH	OS	46.5000	111.0000
angle	OS	CH	NS	46.5000	111.0000
angle	C2	CH	NS	46.5000	111.0000
angle	CH	CH	NS	46.5000	111.0000
angle	C2	CH	NH2E	46.5000	111.0000
angle	CH	CH	NH2E	46.5000	111.0000
angle	OS	CH	NH2E	46.5000	111.0000
angle	C3	CH	CH	46.5000	111.0000
angle	OS	C2	C3	150.5000	111.0000
angle	OH	C2	CH	70.0000	112.0000
angle	CH	C2	CH	70.0000	112.0000
angle	CN	CS	C3	70.0000	119.7000
angle	CN	CS	CF	70.0000	120.7000
angle	C3	CS	CF	70.0000	119.7000
angle	BR	CS	CF	70.0000	119.7000
angle	BR	CS	CA	70.0000	119.7000

angle	O5*	PA	O1A	47.6000	108.0000
angle	O5*	PA	O2A	47.6000	108.0000
angle	O5*	PA	O3A	47.6000	120.0000
angle	O1A	PA	O2A	47.6000	110.0000
angle	O1A	PA	O3A	47.6000	106.0000
angle	O2A	PA	O3A	47.6000	106.0000
angle	O3A	PB	O1B	47.6000	108.0000
angle	O3A	PB	O2B	47.6000	108.0000
angle	O3A	PB	N3B	47.6000	120.0000
angle	O1B	PB	O2B	47.6000	107.0000
angle	O1B	PB	N3B	47.6000	106.0000
angle	O2B	PB	N3B	47.6000	106.0000
angle	N3B	PG	O1G	47.6000	112.0000
angle	N3B	PG	O2G	47.6000	112.0000
angle	N3B	PG	O3G	47.6000	112.0000
angle	O1G	PG	O2G	47.6000	107.0000
angle	O1G	PG	O3G	47.6000	107.0000
angle	O2G	PG	O3G	47.6000	107.0000
angle	PA	O3A	PB	23.6000	135.0000
angle	PB	N3B	PG	23.6000	135.0000

angle	O113PG	C13PG	O113PG	50.000	118.70
-------	--------	-------	--------	--------	--------

angle O113PG	C13PG	C23PG	50.000	121.15
angle C13PG	C23PG	O213PG	50.000	113.33
angle C33PG	C23PG	O213PG	50.000	114.42
angle C13PG	C23PG	C33PG	50.000	106.54
angle C23PG	C33PG	OP43PG	25.000	112.75
angle C33PG	OP43PG	P3PG	25.000	121.84
angle OP43PG	P3PG	OP13PG	25.000	103.00
angle OP13PG	P3PG	OP13PG	25.000	115.00

dihedral	X	C2	CH	X	10.5000	3	0.0000
dihedral	X	CH	CH	X	10.5000	3	0.0000
dihedral	X	CH	OS	X	5.5000	3	0.0000
dihedral	X	CH	OH	X	5.5000	3	0.0000
dihedral	X	C2	OH	X	5.5000	3	0.0000
dihedral	X	C2	OS	X	5.5000	3	0.0000
dihedral	X	OS	P	X	8.7500	3	0.0000
dihedral	X	OH	P	X	8.7500	3	0.0000
dihedral	X	CH	NS	X	3.3000	2	0.0000
dihedral	X	CA	N2	X	60.0000	2	180.0000

dihedral	OS	P	OS	C2	8.7500	3	0.0000
dihedral	OS	P	OS	C2	8.7500	2	0.0000
dihedral	OS	P	OS	C3	8.7500	3	0.0000
dihedral	OS	P	OS	C3	8.7500	2	0.0000
dihedral	OS	P	OS	CH	8.7500	3	0.0000
dihedral	OS	P	OS	CH	8.7500	2	0.0000
dihedral	CH	OS	P	OH	8.7500	3	0.0000
dihedral	CH	OS	P	OH	8.7500	2	0.0000
dihedral	CE	CB	CB	NC	50.0	2	180.00
dihedral	CE	CB	CB	CE	50.0	2	180.00
dihedral	CE	CB	CB	CA	50.0	2	180.00
dihedral	CE	CB	CB	NA	50.0	2	180.00
dihedral	NC	CB	CB	NB	20.5	2	180.00
dihedral	CA	CB	CB	NS	20.5	2	180.00
dihedral	CN	CB	CB	NS	20.5	2	180.00

dihedral	CE	CB	CB	NS	50.0	2	180.00
dihedral	CE	CB	CB	CB	50.0	2	180.00
dihedral	NS	CB	CB	CH	20.5	2	180.00
dihedral	NA	CB	CB	NB	20.5	2	180.00
dihedral	NA	CB	NA	CA	20.5	2	180.00
dihedral	CN	NA	CB	NB	20.5	2	180.00
dihedral	CA	CB	NA	CB	20.5	2	180.00
dihedral	CA	NA	CB	CB	20.5	2	180.00

improper	C2	X	X	CH	310.5000	0	35.2600
improper	C3	X	X	CH	310.5000	0	35.2600
improper	OH	X	X	CH	310.5000	0	35.2600
improper	OS	X	X	CH	310.5000	0	35.2600
improper	CH	OS	C2	NS	310.5000	0	35.2600
improper	CH	OS	CH	NS	310.5000	0	35.2600
improper	CH	OS	C2	NH2E	310.5000	0	35.2600
improper	CH	OS	CH	NH2E	310.5000	0	35.2600
improper	HN	X	X	NA	430.0000	0	0.0000
improper	HN	X	X	NS	430.0000	0	0.0000
improper	H2	X	X	N2	430.0000	0	0.0000
improper	CN	X	X	CB	40.5000	0	0.0000
improper	CN	X	X	CS	40.5000	0	0.0000
improper	CN	X	X	CF	40.5000	0	0.0000
improper	CN	X	X	ON	1130.3000	0	0.0000
improper	CN	X	X	NC	40.5000	0	0.0000

improper	CN	X	X	NS	40.5000	0	0.0000
improper	CA	X	X	CB	40.5000	0	0.0000
improper	CA	X	X	NC	40.5000	0	0.0000
improper	CA	X	X	N2	430.0000	0	0.0000
improper	CA	X	X	NS	40.5000	0	0.0000
improper	CB	X	X	CE	40.5000	0	0.0000
improper	CB	X	X	NA	40.5000	0	0.0000
improper	CB	X	X	NB	40.5000	0	0.0000
improper	CB	X	X	NC	40.5000	0	0.0000
improper	CB	X	X	NS	40.5000	0	0.0000
improper	CF	X	X	NC	40.5000	0	0.0000
improper	CF	X	X	NA	40.5000	0	0.0000
improper	CH	X	X	NS	40.5000	0	0.0000
improper	C3	X	X	CS	430.0000	0	0.0000
improper	NB	X	X	NS	40.5000	0	0.0000
improper	BR	X	X	CS	40.5000	0	0.0000

improper	C13PG	O113PG	O113PG	C23PG	300.0000	0	0.0000
----------	-------	--------	--------	-------	----------	---	--------

nonbonded	CN	0.1200	3.7418	0.1	3.3854
nonbonded	CA	0.1200	3.7418	0.1	3.3854
nonbonded	CB	0.1200	3.7418	0.1	3.3854
nonbonded	CS	0.1200	3.7418	0.1	3.3854
nonbonded	CH	0.0486	4.2140	0.1	3.3854
nonbonded	C2	0.1142	3.9823	0.1	3.3854
nonbonded	C3	0.1811	3.8576	0.1	3.3854
nonbonded	CE	0.1200	3.7418	0.1	3.3854
nonbonded	N2	0.2384	2.8509	0.2384	2.8509
nonbonded	NC	0.2384	2.8509	0.2384	2.8509
nonbonded	NS	0.2384	2.8509	0.2384	2.8509
nonbonded	NA	0.2384	2.8509	0.2384	2.8509
nonbonded	NB	0.2384	2.8509	0.2384	2.8509
nonbonded	ON	0.1591	2.8509	0.1591	2.8509
nonbonded	O2	0.1204	2.8509	0.1204	2.8509
nonbonded	OH	0.1591	2.8509	0.1591	2.8509
nonbonded	P	0.5849	3.3854	0.5849	3.3854
nonbonded	MG	0.2000	2.8509	0.2000	2.8509
nonbonded	O113PG	0.1204	2.8509	0.1204	2.8509
nonbonded	O213PG	0.1591	2.8509	0.1591	2.8509
nonbonded	OP13PG	0.1204	2.8509	0.1204	2.8509
nonbonded	OP43PG	0.1204	2.8509	0.1204	2.8509
nonbonded	C13PG	0.1142	3.9823	0.1	3.3854
nonbonded	C23PG	0.0486	4.2140	0.1	3.3854
nonbonded	C33PG	0.1142	3.9823	0.1	3.3854
nonbonded	P3PG	0.5849	3.3854	0.5849	3.3854

```
set echo=true end
```