

STRUCTURAL STUDIES OF GLYCERALDEHYDE-3-PHOSPHATE  
DEHYDROGENASE COMPLEXES AND THE *E. COLI* PutA DNA BINDING  
DOMAIN

---

A Dissertation  
Presented to  
The Faculty of the Graduate School  
University of Missouri-Columbia

---

In Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy

---

by  
JERMAINE L. JENKINS

Dr. John J. Tanner, Dissertation Supervisor

August 2006

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled

STRUCTURAL STUDIES OF GLYCERALDEHYDE-3-PHOSPHATE  
DEHYDROGENASE COMPLEXES AND THE *E. COLI* PutA DNA BINDING  
DOMAIN

presented by Jermaine Jenkins

a candidate for the degree of Doctor of Philosophy

and hereby certify that in their opinion it is worthy of acceptance.

John Tanner\_\_\_\_\_

Michael T. Henzl\_\_\_\_\_

Thomas Quinn\_\_\_\_\_

Lisa Beamer\_\_\_\_\_

Kent Gates\_\_\_\_\_

To Susan Santiago

## ACKNOWLEDGEMENTS

I would like to acknowledge my wife Susan for inspiring me to go back to school and for being a constant source of support. I want to say thank you to my family for their patience and support during my college years. The women in my family have taught me how to be a good man and have since childhood told me I could accomplish anything, for that I will be eternally grateful.

I would like to thank my committee members (listed above) for their support and help throughout graduate school. I want to thank my undergraduate research advisor Mark Holland for all his support and for recommending Mizzou to me.

To Jon Schuermann, Shorena Nadaraia, Chris Bottoms, Season Prewitt, Min Zhang, Tommi White, Dale Karr, John Larson, Zhonghui Ou, Levi Felts, Li Ma, and all the other members of the Tanner lab since 1999 I do and always will consider you family! I would also like to thank the whole Biochemistry department because there are too many people to list individually.

Finally, I want to say a heartfelt thanks to Jack Tanner my adviser for putting up with me for six plus years and for all he has taught me during that time. Thanks for all your support.



## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	ii
ABSTRACT.....	iii
LIST OF TABLES .....	v
LIST OF ILLUSTRATIONS.....	vi

## CHAPTER

### **1. High Resolution Structure of Human Placental D-Glyceraldehyde-3-**

#### **Phosphate Dehydrogenase**

Abstract.....	2
Introduction.....	3
Experimental Procedures.....	6
Cloning, Expression, and Purification.....	6
Dynamic Light Scattering and Gel Filtration.....	7
Crystallization and X-ray Diffraction Data Collection.....	8
Data Collection and Processing.....	9
Structure Determination and Refinement.....	12
CGP 3466 Docking to HsGAPDH .....	15
GAPDH-Siah1 Docking Calculations.....	16
Results.....	17
Quality of Model and Overall Fold.....	17
Active Site.....	25
Selectivity Cleft.....	30
Gel Filtration.....	34

Co-crystallization and Soaking Studies.....	34
Docking Studies of CGP-3466.....	42
Structural Insights into GAPDH-Siah1 Interaction.....	56
Discussion.....	62
Selectivity Cleft.....	62
Docking Studies of CGP-3466.....	63
NO-S-Nitrosylation-GAPDH-Siah1 Cascade.....	64
Acknowledgements.....	66
References.....	67
 2. <b>High Resolution Structures of <i>Thermus aquaticus</i> Glyceraldehyde-3- Phosphate Dehydrogenase: Conformational Changes Induced by Phosphate Binding</b>	
Introduction.....	72
Experimental Procedures.....	81
Purification and Crystallization.....	81
Data Collection, Model Building, and Refinement.....	82
Results and Discussion.....	86
Quality of Models.....	86
Structural comparison of the 1.65 Å structure with 1CER.....	98
Substrate Phosphate binding site and movement of Asp181, Arg194, and Arg230.....	102
Movement of the inorganic phosphate binding loop.....	110
References.....	133

<b>3. Crystal Structures of the DNA-Binding Domain of the Multifunctional PutA Flavoprotein</b>	
Abstract.....	136
Introduction.....	137
Experimental Procedures.....	140
Materials.....	140
Subcloning and Site-directed Mutagenesis of PutA and PutA52.....	140
Preparation of PutA Proteins for DNA-binding Studies.....	141
DNA Binding Assays.....	142
Preparation of Selenomethionyl PutA52K9M.....	142
Preparation of PutA52 for Crystallization Trials.....	142
Crystallization of PutA52 and Se-Met PutA52K9M.....	143
Structure Determination of PutA52K9M.....	145
Structure Determination of PutA52WT.....	146
Modeling of PutA52/DNA Complex.....	147
Results.....	151
Overall fold.....	151
Conserved RHH Sequence-Structure Relationships Present in PutA.....	155
Sequence-Structure Relationships Unique to the PutA RHH Subfamily.....	160
Residues Involved In Binding DNA.....	163
Discussion.....	171
The PutA RHH Subfamily.....	171
Structural Basis of PutA Dimerization.....	172

Structural Basis of DNA Recognition .....	173
Acknowledgment.....	175
References.....	176
4. VITA.....	178

## LIST OF TABLES

Figure	Page
1.1 Data Collection and Refinement Statistics.....	10
1.2 Molecular Replacement Results.....	14
1.3 Summary of CGP-3466 Docking Results.....	53
2.1 Data Collection and Refinement Statistics.....	84
2.2 C <sub>α</sub> Superposition statistics.....	100
2.3 Hydrogen bonding interactions found in the phosphate-free P <sub>s</sub> -site residues.....	104
2.4 Interactions between the substrate phosphate and active site residues.....	109
3.1 Data Collection and Refinement Statistics.....	149

## LIST OF ILLUSTRATIONS

Figure	Page
1.1 Sequence alignment of solved mammalian and <i>P. falciparum</i> GAPDHs of known structure.....	19
1.2 Ribbon drawing of the HsGAPDH tetramer viewed down the P-axis.....	22
1.3 NAD <sup>+</sup> conformation and interactions.....	24
1.4 Effect of NAD <sup>+</sup> binding on conformation of catalytic Cys152.....	28
1.5 Model of inhibitor NMDBA in the HsGAPDH active site (stereoview).....	33
1.6 Chemical structures of CGP-3466, DEP, and DES.....	37
1.7 Gel filtration of HsGAPDH in precrystallization buffer.....	39
1.8 Gel filtration of HsGAPDH in the presence of DEP and DES.....	41
1.9 HsGAPDH tetramer with space fill molecules representing the seven preferred clustering sites of CGP 3466.....	44
1.10 CASTP analysis of HsGAPDH pockets.....	46
1.11 Surface view of the top seven CASTP predicted pockets for HsGAPDH.....	49
1.12 Predicted binding sites for CGP-3466 in 270s binding pocket.....	51
1.13 Predicted binding sites for CGP-3466.....	55
1.14 Structures of GAPDH and Siah1 highlighting regions that are important for GAPDH-Siah1 association.....	59
1.15 Model of GAPDH-Siah1 complex.....	61
2.1 The postulated catalytic mechanism for GAPDH.....	75
2.2 Cartoon representation of the flip-flop model.....	78
2.3 Cartoon representation of the two tetramers found in the asymmetric unit of the 1.65 Å resolution TAGAPDH structure.....	88
2.4 Ribbon drawing of a TaGAPDH tetramer viewed down the Q-axis.....	90

2.5	A view of the interactions between PEG and a Na <sup>+</sup> ion with the surrounding solvent.....	92
2.6	Ligplot of the two isopropanol molecules (IPA) found in the 1.65 Å structure....	95
2.7	View of one of the three glycerol molecules (GOL) of the 1.65 Å structure.....	97
2.8	Ribbon drawing of the 3 TaGAPDH OPQR tetramers superimposed and viewed down the P-axis.....	100
2.9	The two conformations modeled in the phosphate-free P <sub>s</sub> -site.....	104
2.10	A view of the 50 mM Q subunit P <sub>s</sub> -site interactions.....	109
2.11	Sequence alignment of GAPDH structures that show movement of the P <sub>i</sub> -binding loop.....	114
2.12	Cartoon of the P <sub>i</sub> -binding loop (red) in TaGAPDH.....	116
2.13	The 50 mM subunit P <sub>i</sub> -binding loop in TaGAPDH (stereoview).....	118
2.14	Conformations of the P <sub>i</sub> -binding loops (stereoview).....	121
2.15	Superimposed conformations of the P <sub>i</sub> -binding loops from different GAPDHs (stereoview).....	124
2.16	Stereoview of superimposed conformations of the P <sub>i</sub> -binding loops from different GAPDHs.....	126
2.17	A view of the active site of 50 mM structure superposed on to the GAPDH hemiacetal structure (stereoview).....	130
3.1	Structure of the PutA RHH domain.....	153
3.2	Amino acid sequence alignments of RHH and PutA proteins.....	157
3.3	Comparison of the αA-αB loop regions of PutA52 and Arc.....	159
3.4	Gel mobility shift assay of wild type PutA and PutA mutant K9M.....	166
3.5	Model of the PutA RHH domain interacting with DNA.....	168

## CHAPTER 1

# HIGH RESOLUTION STRUCTURE OF HUMAN D- GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE

**Jermaine L. Jenkins and John J. Tanner**

Departments of Biochemistry and Chemistry

University of Missouri, Columbia, MO 65211

\*Portions of this work are reproduced from *Acta Crystallographica Section D* © 2006,  
Volume 62, pp 290-301 by copyright permission of The International Union of  
Crystallography.



## ABSTRACT

GAPDH (D-Glyceraldehyde-3-phosphate dehydrogenase) is a multifunctional protein that is a target for the design of antitrypanosomatid and anti-apoptosis drugs. Here, the first high-resolution (1.75 Å) structure of a human GAPDH is reported. The structure shows that the intersubunit selectivity cleft that has been leveraged in the design of antitrypanosomatid compounds is closed in human GAPDH. Modeling of an antitrypanosomatid GAPDH inhibitor in the human GAPDH active site provides insights into the basis for the observed selectivity of this class of inhibitor. Moreover, the high-resolution data revealed a new feature of the cleft: water-mediated intersubunit hydrogen bonds that assist closure of the cleft in the human enzyme. The structure was used in a computational ligand docking study of the drug CGP-3466, which inhibits apoptosis by preventing nuclear accumulation of GAPDH. Plausible binding sites were discovered in the adenosine pocket of the NAD<sup>+</sup>-binding site and in a hydrophobic channel located in the center of the tetramer near the intersection of the three molecular twofold axes. The structure is also used to build a qualitative model of the complex between GAPDH and the E3 ubiquitin ligase Siah1. The model suggests that the convex surface near GAPDH Lys227 interacts with a large shallow groove of the Siah1 dimer. These results are discussed in the context of the recently discovered NO–S-nitrosylation–GAPDH–Siah1 apoptosis cascade.

## INTRODUCTION

GAPDH is an  $\text{NAD}^+$ -dependent glycolytic enzyme that catalyzes the formation of 1,3-bisphosphoglycerate from glyceraldehyde-3-phosphate and inorganic phosphate. The enzyme is ubiquitous in nature and cellularly abundant, which has led to widespread study of its function, structure, and roles in human health and disease. For example, GAPDH is an attractive target for the design of drugs to combat protozoan parasites whose bloodstream forms depend solely on glycolysis for energy production (1-13). Enzymes that have been targeted for drug design include GAPDHs from trypanosomatid species *Leishmania mexicana*, *Trypanosoma cruzi* and *Trypanosoma brucei*, which are responsible for the debilitating illnesses leishmaniasis, Chagas disease and African sleeping sickness, respectively (2, 4, 5, 9, 11, 13, 14). Also, inhibitors of GAPDH from *Plasmodium falciparum* have been proposed in the fight against malaria (15). As part of this effort, crystal structures of GAPDHs from *L. mexicana* (7, 12, 13), *T. cruzi* (3, 16), *T. brucei* (17) and *P. falciparum* (PDB 1ZYA) have been determined to aid structure-based drug design.

Although once thought to play a role solely in glycolysis, GAPDH is now considered to be a classic example of a moonlighting protein (18). Mammalian GAPDH has been implicated in many cellular activities, including apoptosis, nuclear RNA transport, DNA replication, DNA repair, RNase activity, microtubule bundling and membrane fusion (19-21). Consistent with its many functions, GAPDH is thought to play roles in many diseases, including Parkinson's Disease (PD), Alzheimer's Disease, Huntington's Disease, dentatorubropallidoluysian atrophy and prostate cancer (19-26).

The involvement of GAPDH in apoptosis is particularly intriguing because apoptosis is one mechanism of neuronal cell death that is thought to contribute to neurodegenerative diseases such as PD (22-25, 27-32). The recent discovery of the NO-S-nitrosylation-GAPDH-Siah1 cascade answers two major outstanding questions about the role of GAPDH in apoptosis (27). First, nuclear accumulation of GAPDH had long been observed in cells stimulated to undergo apoptosis, however; the mechanism of nuclear translocation was unknown since GAPDH lacks a nuclear localization signal. Hara et al. showed that GAPDH enters the nucleus following S-nitrosylation of the active site Cys by nitric oxide (NO) and binding to the E3 ubiquitin ligase Siah1, which does contain a nuclear localization signal (27). Thus, GAPDH gains entry into the nucleus by virtue of Siah1's nuclear localization signal. The second major question about GAPDH and apoptosis concerns the role of nuclear GAPDH in apoptosis. The work of Hara and coworkers (27) suggests that GAPDH stabilizes the otherwise short-lived Siah1, which leads to increased degradation of nuclear targets and promotion of apoptosis.

Small molecule compounds that interfere with the proapoptotic activity of GAPDH by disrupting the NO-S-nitrosylation-GAPDH-Siah1 cascade are potential drug candidates for the treatment of PD and other neurodegenerative disease. For example, the small molecule compound CGP-3466 (dibenzo-[*b,f*]oxepin-10-ylmethyl-methyl-prop-2-ynyl-amine) reduces apoptosis by preventing nuclear localization of GAPDH (22, 25, 33); however, the molecular-level mechanism of CGP-3466 is poorly understood.

Knowledge of the three-dimensional structure of human GAPDH at high resolution would facilitate the development of specific inhibitors of GAPDHs from parasites as well as new antiapoptosis drugs for the treatment of neurodegenerative

diseases. Crystals of human liver GAPDH have been reported (34) and a moderate resolution (2.5 Å) structure of the human liver enzyme has been deposited in the Protein Data Bank (PDB code 1ZNQ (35)).

Here, we report the 1.75 Å structure of recombinant human placental GAPDH (HsGAPDH). The structure provides an updated view of the NAD<sup>+</sup>-binding site, which is the target of GAPDH inhibitors designed to combat parasitic diseases. The structure also provides a foundation for investigating the interactions between CGP3466 and human GAPDH. Computational docking studies performed with our structure suggest plausible binding sites for CGP-3466 in the adenosine pocket of the NAD<sup>+</sup>-binding site and a hydrophobic channel located at the intersection of the molecular twofold axes. We propose that binding of CGP-3466 in these sites may cause tertiary and quaternary structural changes in GAPDH that decrease affinity for the E3 ubiquitin ligase Siah1, which would result in inhibition of the NO-GAPDH-Siah1 apoptosis cascade (27). Finally, a qualitative model for the complex between GAPDH and Siah1 is proposed.

## EXPERIMENTAL PROCEDURES

### *Cloning, Expression, and Purification*

The pChug 20.2 construct (36) was used to engineer a His-tagged HsGAPDH. PCR was used to amplify the HsGAPDH coding sequence of pChug 20.2 and to introduce Nde I and BamH I restriction sites. The resulting fragment was subcloned into a pET14b vector (Novagen) carrying an N-terminal hexahistidine tag (His<sub>6</sub>). This His<sub>6</sub>-HsGAPDH construct was introduced into the *E. coli* expression strain BL21(DE3) pLysS. Protein was expressed by induction with isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) when the culture, which had been grown in LB broth containing carbenicillin and chloramphenicol, reached an OD of approximately 0.8. Induction with IPTG lasted for 4 h at 310 K.

Pelleted *E. coli* cells were frozen overnight and resuspended the next day in lysis buffer (50 mM sodium phosphate pH 8.0 containing 10 mM imidazole, 1 mM phenylmethanesulfonyl fluoride, and 0.3 M NaCl). The resuspended cells were disrupted by sonication in an ice-chilled container, and Benzonase Nuclease (Novagen) was added to cleave any DNA/RNA released as a result of sonication. The resulting cellular debris was removed by centrifugation at 10 000 rev min<sup>-1</sup> for 25 min at 277 K followed by vacuum filtration through a 0.45  $\mu$ m filter. Approximately 28 ml of Ni-NTA resin (Qiagen) was added to the cleared supernatant, and the mixture was stirred at 277 K for 4 hours. The suspension was allowed to settle, the supernatant was decanted, and the remaining slurry was poured into a gravity filtration column. Four column volumes of wash buffer (50 mM sodium phosphate pH 8.0, 20 mM imidazole, 0.3 M NaCl) were

applied to the packed column, and the protein was eluted with 50 mM sodium phosphate pH 8.0, 250 mM imidazole, 0.3 M NaCl.

All fractions containing protein were analyzed by denaturing SDS-PAGE gels, and those fractions containing His<sub>6</sub>-HsGAPDH were combined and dialyzed overnight into 20 mM Tris pH 8.0. The protein was loaded onto a HiTrap Blue column (Amersham Pharmacia) that had been equilibrated with 20 mM Tris pH 8.0. The column was washed with the start buffer, and HsGAPDH was eluted with the addition of 0.5 M NaCl in 20 mM Tris pH 8.0. The eluted fractions were pooled and dialyzed into the precrystallization buffer (2.5 mM Tris pH 7.7, 0.1 M NaCl, 1 mM  $\beta$ -mercaptoethanol, and 1 mM disodium EDTA) supplemented with 10 mM NAD<sup>+</sup>. To remove the His-tag from the purified His<sub>6</sub>-HsGAPDH, the protein was incubated with Biotinylated Thrombin (Novagen) for 12 h at 310 K. Streptavidin agarose and spin filters (Novagen) were used to remove the thrombin. The protein was batch purified as before with Ni-NTA and the fractions from the packing and wash steps containing the desired HsGAPDH were pooled. HsGAPDH was dialyzed into the precrystallization buffer and concentrated to 19.5 mg ml<sup>-1</sup> using centrifugal filters and tubes (Millipore). The protein concentration was measured using the Bradford assay (Pierce).

### *Gel filtration*

Determinations of the apparent oligomeric state of HsGAPDH (9 mg ml<sup>-1</sup>) in the aforementioned precrystallization buffer were carried out using a Superdex 200 and Sephacryl S-200 gel filtration columns (GE Biosciences) connected to an AKTA FPLC. To investigate the effects of the deprenyl-like compounds, R-(–)-deprenyl (DEP) and (–)-

desmethyldeprenyl (DES), on the oligomeric state of HsGAPDH they were incubated with HsGAPDH, the enzyme was incubated with DES or DEP at various concentrations for 30 or 60 minutes prior to gel filtration analysis.

#### *Crystallization and X-ray diffraction data collection*

All crystallization experiments were performed at 295 K using the sitting-drop method of vapor diffusion. Initial crystallization conditions were obtained with Wizard, Hampton, and Index crystal screen kits. Several crystal forms were obtained for His<sub>6</sub>-HsGAPDH. The most promising condition was 50% PEG 4000, 0.1 M Hepes buffer, and 10% isopropanol. Large, blemish-free crystals diffracted to only 3.4 Å resolution, therefore; the His<sub>6</sub>-tag was cleaved and the tag-free protein was input to crystal screens.

Screening of the tag-free protein was very successful. For example, crystals appeared in approximately 36 of the 96 conditions of the Index Screen. Many of the screen conditions that produced crystals with His<sub>6</sub>-HsGAPDH also gave crystals with tag-free HsGAPDH. The crystals used for structure determination grew directly from Index Screen condition 89 (0.1 M succinic acid pH 7.0 and 15 % PEG 3350) in 4 days. These crystals were cryoprotected with the addition of reservoir solution supplemented with 15 % glycerol. The cryoprotected crystals were frozen in liquid nitrogen.

Diffraction to 2.1 Å was observed using a Cu rotating anode system, and diffraction to 1.75 Å was obtained at beamline 19-ID of the Structural Biology Center at the Advanced Photon Source (APS). The space group is P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> with unit cell dimensions  $a = 85 \text{ Å}$ ,  $b = 126 \text{ Å}$ ,  $c = 132 \text{ Å}$ . There is one tetramer in the asymmetric unit and the solvent content is 44.8 (37).

We note that attempts to crystallize His<sub>6</sub>-HsGAPDH using the succinic acid/PEG 3350 recipe failed. Thus, removal of the affinity tag was critically important for obtaining high quality crystals.

### *Data Collection and Processing*

Molecular replacement and initial model building studies were based on a 2.15 Å resolution data set collected using an in-house Cu rotating anode system. The data collection consisted of 235 frames with a crystal-to-detector distance of 175 mm, an oscillation angle of 0.5° and an exposure time of 10 minutes per frame. The final data set consists of 309 453 observations of 77 527 unique reflections and is 99 % complete to 2.15 Å resolution. The overall  $R_{\text{merge}}$  on  $I$  is 6.3 %, with an average  $I/\sigma$  of 20.

Subsequently, a 1.75 Å resolution data set was collected at APS beamline 19-ID. The data were processed with HKL and HKL2000 (38). The data collection consisted of 930 frames obtained with a crystal-to-detector distance of 150 mm, an oscillation angle of 0.2°, and an exposure time of 5 s per frame. The data set had excellent redundancy (6.6), completeness (99.6), and signal-to-noise ratio ( $I/\sigma = 4$  in the high resolution bin). Processing statistics for the APS data set are listed in Table 1.1.



**Table 1.1.** Data collection and refinement statistics

PDB entry	1U8F
Wavelength (Å)	0.97857
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Unit cell dimensions (Å)	a = 84.95, b = 125.65, c = 132.33
Diffraction resolution (Å)	100–1.75 (1.81-1.75) <sup>a</sup>
No. of observations	934 983
No. of unique reflections	141 888
Redundancy	6.6 (6.1)
Completeness (%)	99.6 (99.6)
Average I/σ	23.0 (4.1)
R <sub>merge</sub> (I)	0.071 (0.436)
No. of protein atoms	10 133
No. of ligands	3
No. of water molecules	911
R <sub>cryst</sub>	19.1
R <sub>free</sub> <sup>b</sup>	21.7
RMSD <sup>c</sup>	
Bond lengths (Å)	0.0047
Bond angles (deg.)	1.388
Dihedral angles (deg.)	24.32
Improper dihedrals (deg.)	0.719
Ramachandran plot <sup>d</sup>	
Favored (%)	89.7
Allowed (%)	9.7

Generous (%)	0.3
Disallowed (%)	0.3
Average B-factors ( $\text{\AA}^2$ )	
Protein	15.42
Solvent	23.15

<sup>a</sup>Values for the outer resolution shell of data are given in parenthesis.

<sup>b</sup>10%  $R_{\text{free}}$  test set.

<sup>c</sup>Compared to the Engh and Huber force field. (39)

<sup>d</sup>The Ramachandran plot was generated with PROCHECK. (40)

### *Structure Determination and Refinement*

Molecular replacement calculations (in-house data set, 10-4 Å) were performed with the program AMoRe (41) using the *Thermus aquaticus* GAPDH tetramer (pdb code 1cer, (42)) as the search model. The *T. aquaticus* and HsGAPDH proteins share approximately 44% sequence identity. Molecular replacement calculations were carried out for all possible primitive orthorhombic space groups (Table 1.2). Space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> gave the best solution, as indicated by a correlation coefficient of 42.6 and an  $R_{factor}$  of 49.9. Rigid-body refinement in CNS (43)) lowered the  $R_{factor}$  to 48.3.

The phases from rigid body refinement were input to ARP/wARP (44) for automated model building. Side-chains were docked to the ARP/wARP model using the guiSIDE module of CCP4i (45). One complete subunit of the tetramer was built using O (46), and non-crystallographic symmetry was used to generate the full tetramer. The NCS can be described by 3 transformations that relate the O subunit to the P, Q and R subunits. The O subunit is related to the P subunit by a rotation in spherical polar coordinates of ( $\psi = 80.3$ ,  $\phi = 72.0$ ,  $\kappa = 178.9$ ) followed by a translation of ( $x = 71.7$ ,  $y = 76.8$ ,  $z = 37.1$ ). The O subunit is related to the Q subunit by a rotation in spherical polar coordinates of ( $\psi = 170.1$ ,  $\phi = 67.0$ ,  $\kappa = 179.8$ ) followed by a translation of ( $x = 69.6$ ,  $y = -3.1$ ,  $z = 48.8$ ). The O subunit is related to the R subunit by a rotation in spherical polar coordinates of ( $\psi = 90.3$ ,  $\phi = 162.0$ ,  $\kappa = 179.9$ ) followed by a translation of ( $x = -12.1$ ,  $y = 71.5$ ,  $z = 38.8$ ). The model was completed with several rounds of model building using O followed by refinement against the 1.75 Å resolution data using CNS (43). Non-crystallographic symmetry was not used during refinement. See Table 1.1 for refinement statistics.

The programs CNS, O, CCP4i, and Pymol (47) were used for structural analysis of HsGAPDH. Superimpositions of structures and RMSD calculations were done using CNS and the websites for CE (48), Mass (49), and COMPARE (50). For the purposes of RMSD calculations, the NAD binding domain was defined as residues 1-151 and 315-335 and the catalytic domain was defined as residues 152-314.

**Table 1.2.** *Molecular replacement results*

Space Group	$\alpha$	$\beta$	$\gamma$	$x$	$y$	$z$	Corr F	Rfac	Corr I
P2(1)2(1)2(1)	93.16	80.03	159.38	0.3783	0.2857	0.2299	42.6	49.9	46.9
	85.59	80.47	195.93	0.3783	0.2144	0.2704	42.5	50	46.8
P222	93.31	80.07	159.24	0.4264	0.1071	0.4318	27.8	55.2	28.6
	85.58	80.42	196.04	0.4274	0.3924	0.0693	28	55.2	28.8
P222(1) c	93.17	79.96	159.43	0.1288	0.2854	0.0004	31	54	31.6
	85.50	80.43	195.83	0.1276	0.2146	0.4418	31.5	54	33.3
P2(1)2(1)2 c	93.11	80.25	159.34	0.0204	0.2854	0.2296	32.6	53.4	34.2
	85.69	80.76	195.75	0.0219	0.2145	0.2706	32.6	53.5	34.1
P2(1)2(1)2 b	162.12	10.32	357.60	0.0203	0.0487	0.0357	33.1	53.4	34.8
	25.52	10.39	350.11	0.0206	0.4511	0.4642	33	53.5	34.9
P222(1) b	163.09	10.49	356.75	0.0202	0.1219	0.4227	32	53.5	33.2
	24.01	10.58	351.78	0.0200	0.3789	0.4137	32	53.6	33.1
P2(1)2(1)2 a	9.70	85.54	105.16	0.4640	0.1845	0.1289	31.7	53.7	33.1
	169.93	86.66	249.94	0.4641	0.3165	0.3711	31.5	53.7	33.1
P222(1) a	9.53	85.61	105.19	0.4638	0.2711	0.3462	32	53.8	34.2
	170.01	86.79	249.86	0.4646	0.2294	0.0768	32.3	53.3	33.9

### **1.1. CGP-3466 docking calculations**

AutoDock 3.0.5 (51) was used to identify plausible binding sites for CGP-3466 (dibenzo-[b,f]oxepin-10-ylmethyl-methyl-prop-2-ynyl-amine). In preparation for these studies, solvent and NAD<sup>+</sup> were removed from the final refined HsGAPDH model and polar hydrogen atoms were added with AutoDockTools (ADT). A model for CGP-3466 (hydrogen atoms included) was built using the PRODRG server (52). Partial charges for each atom of the protein and ligand were calculated with ADT. A global search for potential binding sites was performed using a 230 x 230 x 230 point grid (0.375 Å grid spacing) that covered the entire HsGAPDH tetramer. Promising sites identified by the global docking search were explored further using smaller grids (36 x 36 x 36) focused on the sites of interest. Default parameters for the genetic algorithm-local search method of AutoDock were used for both the global and focused calculations with the following exceptions: translation step = 0.2 Å, quaternion step = 4.0°, torsion step = 4.0°, maximum number of energy evaluations = 1 x 10<sup>7</sup>, number of generations = 2.7 x 10<sup>4</sup>, number of survivors = 5.0, and number of runs = 100. The results were analyzed using the default cluster analysis in ADT. The LPC server was used to analyze interactions formed by the docked ligands and to calculate surface area buried in ligand-protein interfaces (53). The top CGP-3466 poses from focused docking runs have been deposited in the PDB (PDB code 2feh).

### *GAPDH-Siah1 docking calculations*

The web server for PatchDock was used to build qualitative models for the complex between HsGAPDH and the E3 ubiquitin ligase Siah1 in order to predict which protein surfaces might interact. The receptor for docking calculations was the HsGAPDH QR dimer, corresponding to two subunits related by the P molecular twofold axis. The ligand was the homodimer of murine Siah1a obtained from PDB entry 1K2F (54). We note that 1K2F contains Siah1a residues 93–282, which includes the region that is essential for GAPDH–Siah1 complex formation (27). Site-directed mutagenesis results from Hara and coworkers show that GAPDH residue Lys227 is essential for GAPDH–Siah1 association (27); therefore, the docking calculations were constrained to return only complexes having Lys227 in the binding site. The top solution from PatchDock had a score of 10 388, which was 16% higher than the score for solution 2 (8992). For reference, solutions 3–5 had scores of 8854, 8206 and 8156.

## RESULTS

### *Quality of the Model and Overall Fold*

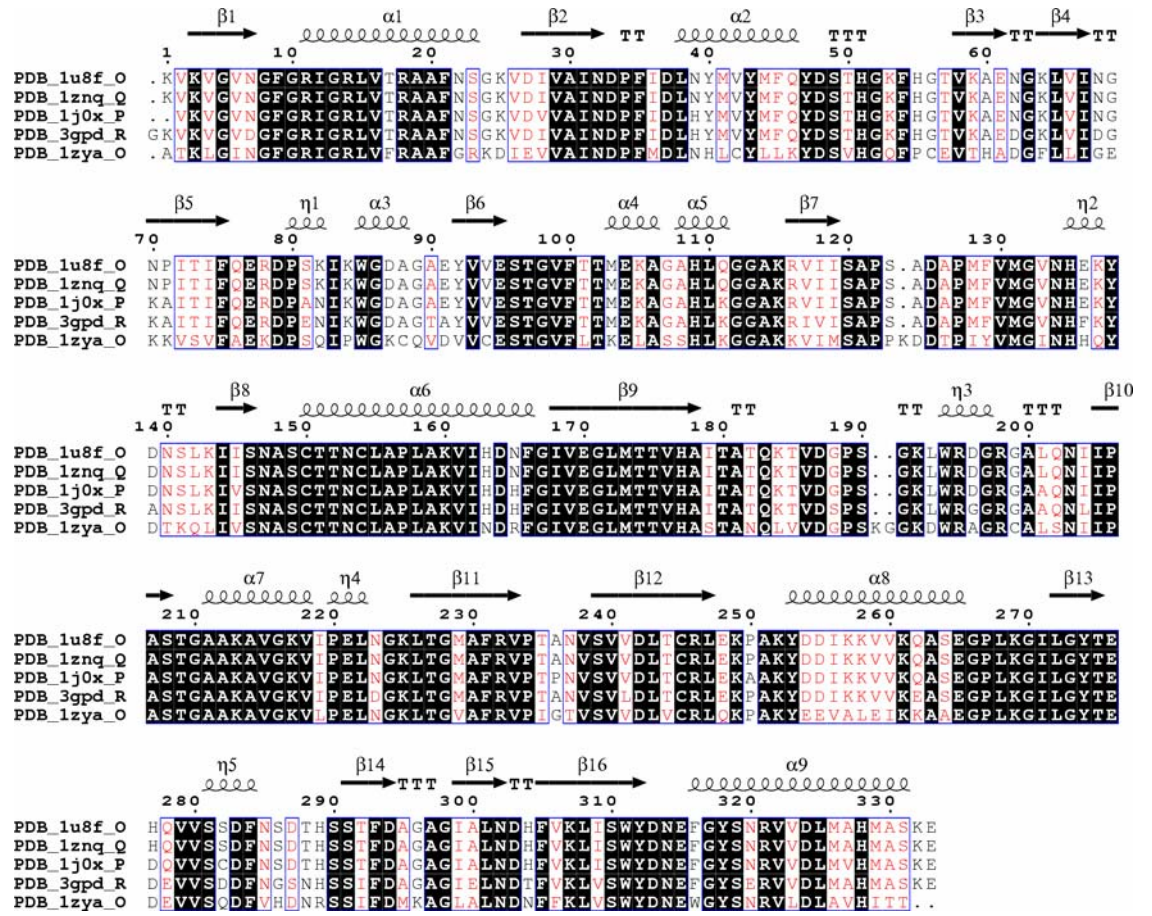
The structure reported here is one of the highest resolution GAPDH structures to date, exceeded only by the 1.7 Å structure of *Alcaligenes xylosoxidans* GAPDH (1OBF, (55)). Our structure is the highest resolution structure of a mammalian GAPDH. For comparison, structures of GAPDH from human liver (PDB code 1ZNQ) and rabbit muscle (PDB code 1J0X, (56)) have been determined at 2.5 Å and 2.4 Å, respectively. A sequence alignment of the available mammalian GAPDH structures and that of *P. falciparum* is included (Figure 1.1)

The refined model consists of one tetramer with subunits labeled O, P, Q and R, containing 10133 atoms, 1332 amino acid residues, 911 water molecules and 3 NAD<sup>+</sup> molecules (Table 1.1). Dual side-chain conformations have been modeled for Glu93 (O, P, Q), MetQ103 and CysO152. Disorder was found only in the two N-terminal residues and in the side-chains of LysO219 and LysQ84.

The refinement statistics are consistent with a well-refined and accurate crystal structure. For example, the R-factor is 0.19 with R<sub>free</sub> = 0.22 for all reflections to 1.75 Å. The average protein *B*-factors for individual subunits are 17.1 Å<sup>2</sup> (O), 15.0 Å<sup>2</sup> (P), 14.3 Å<sup>2</sup> (Q) and 15.3 Å<sup>2</sup> (R). The root mean square deviations of bonds (0.0047 Å) and angles (1.4°) indicated excellent geometry and the stereochemistry meets or exceeds all the main chain and side chain tests of the PROCHECK package (40). Only Ala150 (O, P, R) and Val240 (all four subunits) occupy the generously allowed or disallowed regions of the Ramachandran plot. These residues appear in loops and they adopt similar conformations in other GAPDH structures (7, 13, 56-60).



**Figure 1.1.** *Sequence alignment of mammalian and P. falciparum GAPDHs of known structure. (PDB code 1ZYA). 1U8F (HsGAPDH) has pairwise identities of 100% with 1ZYA, 95% with 1J0X, 89% with 3GPD (Human muscle) and 64% with 1ZYA. Secondary structure was assigned by DSSP from 1U8F. Identical residues are denoted by white type on a black background. Conserved residues are colored red. This alignment was generated by the program ESPript.*

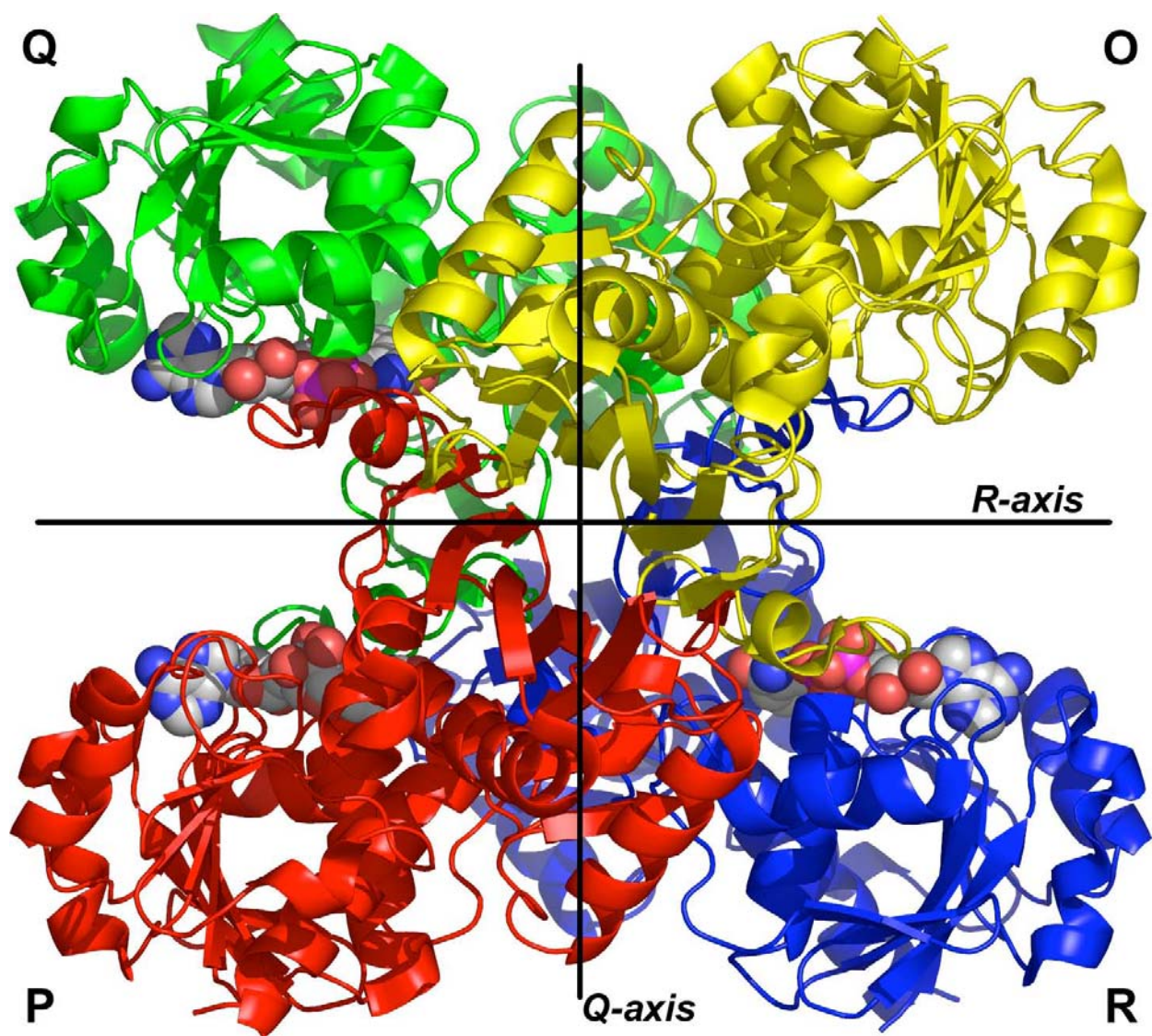


The secondary, tertiary, and quaternary structures of HsGAPDH are very similar to those of other GAPDH structures, as expected (Figure 1.2). The tetramer displays approximate 222 symmetry, with each subunit consisting of an NAD<sup>+</sup>-binding domain (residues 1-151, 315-335) and a catalytic domain (residues 152-314). The NAD<sup>+</sup>-binding domain features the well-known Rossmann dinucleotide-binding fold (61), in which NAD<sup>+</sup> binds in an extended conformation (Figure 1.3A) at the C-terminal edge of a parallel  $\beta$ -sheet that is flanked by helices on either side. The catalytic domain consists of a mixed, twisted  $\beta$ -sheet flanked by three helices on one side. The other side of the  $\beta$ -sheet forms extensive intersubunit contacts with the  $\beta$ -sheet of the catalytic domain of an adjacent subunit. This so-called “P interface” is largest of the three intersubunit interfaces of the tetramer and it buries 7814 Å<sup>2</sup> of surface area in our structure. The Q and R interfaces bury 1824 Å<sup>2</sup> and 5652 Å<sup>2</sup> of surface area, respectively.

The individual subunits of HsGAPDH are very similar to each other, as indicated by RMSD values of 0.50 - 0.66 Å for C $\alpha$  atoms. The catalytic domains superimpose with RMSD values of 0.35 - 0.55 Å and the corresponding values for the NAD<sup>+</sup>-binding domains are 0.55 - 0.67 Å.

The closest structural homologue to our enzyme is human liver GAPDH (1ZNQ, 100 % amino acid sequence identity to HsGAPDH) based on a search of the PDB using SSM (62). The RMSDs between the subunits of our structure and those of 1ZNQ are in the range 0.26 Å – 34 Å. The next closest homologue identified by SSM was another mammalian GAPDH, rabbit GAPDH (1J0X, 95 % amino acid sequence identity to HsGAPDH). Subunits of 1J0X superimpose onto HsGAPDH with RMSD values of 0.28 – 0.68 Å. For reference, the RMSD values between HsGAPDH subunits and those of *T.*

**Figure 1.2.** *Ribbon drawing of the HsGAPDH tetramer viewed down the P-axis.* Subunits are colored as follows: O, yellow; P, red; Q, green; R, blue. Lines indicate locations of the Q and R molecular 2-fold axes. The NAD<sup>+</sup> cofactors are drawn in CPK format.



**Figure 1.3.** *NAD<sup>+</sup> conformation and interactions.* *A*, NAD<sup>+</sup> and selected surrounding water molecules from the P subunit. The map is a simulated annealing  $\sigma$ A-weighted mFo-DFc electron density map contoured at  $3\sigma$ . The simulated annealing calculation was started from the final model with NAD<sup>+</sup> and surrounding residues/water within 3.9 Å of NAD<sup>+</sup> omitted. *B*, schematic diagram of cofactor-protein interactions in the P subunit. The dotted lines indicate electrostatic interactions within 3.2 Å. The thick solid lines denote nonpolar contacts within 3.9 Å.



[illegible]

*aquaticus* GAPDH, which was used for molecular replacement calculations, are 0.90 – 1.1 Å.

### *Active Site*

The active site of each subunit is located in a large cleft between the NAD<sup>+</sup>-binding and catalytic domains. The Cys nucleophile (Cys152) resides at the N-terminus of the first helix of the catalytic domain. Electron density maps clearly indicated the presence of NAD<sup>+</sup> bound with high occupancy in subunits P, Q and R (Figure 1.3A). The average B-factors for NAD<sup>+</sup> were 17 Å<sup>2</sup> (P), 22 Å<sup>2</sup> (Q) and 27 Å<sup>2</sup> (R). On the other hand, electron density for the cofactor in the O chain was very weak. Density representing the pyrophosphate was observed, but density features corresponding to the nicotinamide and adenosine groups were almost completely absent. Thus, NAD<sup>+</sup> was not modeled in the O subunit. Differential cofactor binding to GAPDH subunits has been observed previously, and it is presumably a consequence of the cooperativity of the enzyme (63-65). For example, three NAD<sup>+</sup> molecules are bound to the tetramer of *Trypanosoma cruzi* (PDB code 1ML3, (16)), and only two cofactor molecules are bound to the rabbit muscle GAPDH tetramer (1J0X, (56)).

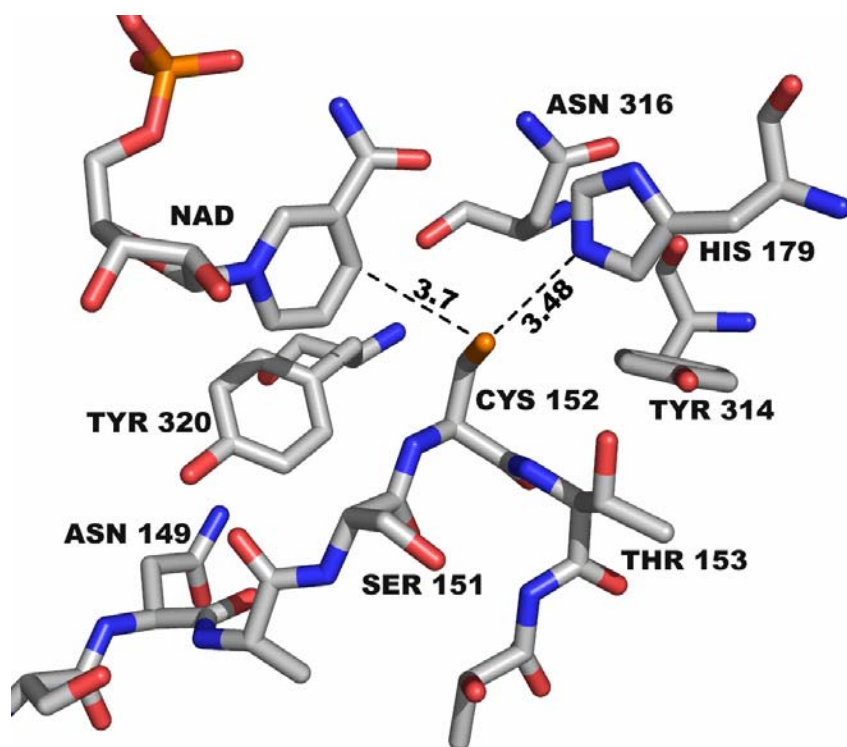
Dual side chain conformations were observed for CysO152, which is presumably due to the absence of NAD<sup>+</sup> in the O subunit. One conformation (Figure 1.4A) matches that found in the other subunits ( $\chi_1 = 52^\circ$ , occupancy = 0.7), and the other conformation (Figure 1.4B) corresponds to a relative rotation of  $105^\circ$  to  $\chi_1 = -53^\circ$  (occupancy = 0.3). Note that the  $\chi_1 = -53^\circ$  conformation for the nucleophilic Cys was also observed in the structure of *Palinurus versicolor* GAPDH (PDB code 1IHX, (66)), which contains a



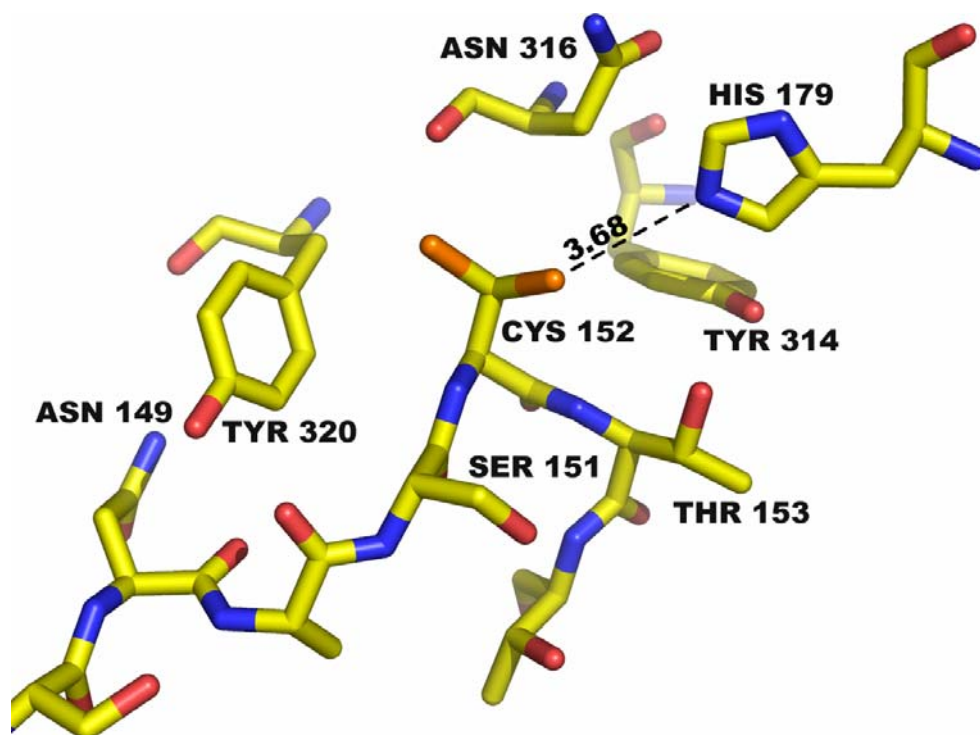
bound NAD<sup>+</sup> analogue with disordered thionicotinamide ring. The authors suggested that movement of the catalytic Cys disrupted the highly conserved salt-bridge triad Arg13-Asp50-Glu317 (HsGAPDH numbering). We note that the triad is formed in all four subunits of our structure.

**Figure 1.4.** *Effect of  $NAD^+$  binding on conformation of catalytic Cys152.* *A*, active site of P subunit, which has bound  $NAD^+$ . *B*, active site of O subunit which does not have bound  $NAD^+$ . Note the dual conformations of Cys152.

A



B



The NAD<sup>+</sup>-binding sites of GAPDHs from pathogenic trypanosomatids are drug design targets (9-13, 67). Knowledge of protein-NAD<sup>+</sup> interactions in human GAPDH is useful as a guide for developing highly specific anti-trypanosomatid drugs that do not inhibit human GAPDH. A schematic diagram of the interactions formed by NAD<sup>+</sup> in our structure is provided (Figure 1.3B). NAD<sup>+</sup> forms several noncovalent interactions that are also observed in other GAPDH structures. For example, the nicotinamide carbonyl forms a hydrogen bond with Asn316, while the nicotinamide amine forms an intramolecular hydrogen bond with the pyrophosphate. As expected, the pyrophosphate binds in the glycine-rich loop of the Rossmann fold, which allows the pyrophosphate to hydrogen bond to the backbone amine groups of the N terminal residues of the first helix of the Rossmann fold. The adenosine ribose forms two hydrogen bonds with highly conserved Asp35 while the pyrophosphate forms hydrogen bonds with the backbone amine groups of highly conserved Arg13 and Ile14. Nonpolar interactions with the cofactor aromatic rings are also observed. The nicotinamide ring contacts Ile14 and Tyr320, while the adenine contacts Pro36, Phe37, Thr99 and Phe102.

The exceptional quality of our electron density maps allowed detailed modeling of solvent (Figure 1.3A). For example, over a dozen water molecules interact with the cofactor (3.2 Å cutoff). Most of these water molecules mediate hydrogen bonds between NAD<sup>+</sup> and the protein (Figure 1.3B). In contrast, the 2.5 Å human liver GAPDH structure has only three water molecules bound to NAD<sup>+</sup>, which are equivalent to HOH15, HOH19 and HOH429 of our structure.

The highly conserved water molecule of the Rossmann fold, which was described in detail by Bottoms *et al.* (68), is present in all four subunits of our structure (HOH425,

HOH429, HOH227, HOH211). This water molecule is important because it bridges the pyrophosphate and the glycine-rich loop. The average B-factor of the conserved water molecule is  $17 \text{ \AA}^2$ , which is significantly lower than the average solvent B-factor of  $23 \text{ \AA}^2$ . Note that the conserved water molecule is present in the O chain, which does not contain a bound  $\text{NAD}^+$ . This result emphasizes that the conserved water molecule is an inherent structural feature of the Rossmann fold itself, as discussed previously (68, 69).

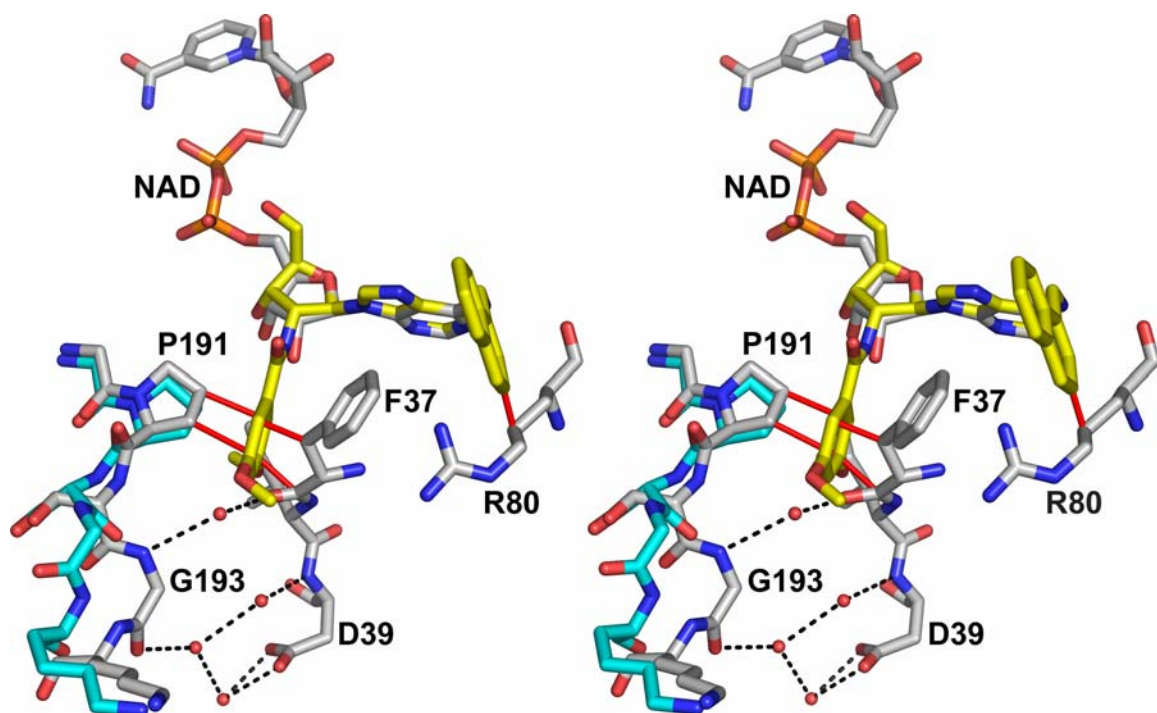
### *Selectivity Cleft*

One strategy for designing potent and selective inhibitors of trypanosomatid GAPDHs is to elaborate an adenosine framework with functional groups that bind in a narrow intersubunit “selectivity cleft” that has been observed in the  $\text{NAD}^+$ -binding sites of GAPDHs from *T. brucei*, *T. cruzi* and *L. mexicana* (9-13, 67). It has been proposed that the selectivity cleft is closed in human GAPDH and thus inhibitors that fill the cleft of trypanosomatid GAPDHs could bind tightly to the desired target enzyme without inhibiting the human enzyme. This strategy has been realized, for example, in the disubstituted adenosine derivative NMDBA (N6-(1-naphthalenemethyl)-2'-deoxy-2'-(3,5-dimethoxybenzamido) adenosine (10-12)). In our structure, Phe37, Ile38 and Asp39 form one side of the cleft, and Pro191, Ser192 and Gly193 from an adjacent subunit form the other side (Figure 1.5). The cleft is only 4 – 5  $\text{\AA}$  wide, compared to 7 – 8  $\text{\AA}$  wide in the structure of *L. mexicana* GAPDH complexed with NMDBA. We superimposed NMDBA onto the adenine moiety of  $\text{NAD}^+$  in our structure in order to examine steric clashes that might prevent binding of the inhibitor to the human enzyme (Figure 1.5). The O-2' substituent of NMDBA is predicted to form steric clashes (contact distances = 1.1 –

2.7 Å) with cleft residues Phe37, Ile38, and Pro191 (Figure 1.5). We also predict steric clash of the inhibitor naphthalene group with Arg80. Although our modeling does not consider the possibility of induced fit binding, the severity of the predicted clashes suggests that the selectivity cleft is effectively closed in HsGAPDH, which supports the strategy of leveraging this cleft in inhibitor design.

Interestingly, our structure shows a large conformational difference with the 2.5 Å human liver GAPDH structure (PDB entry 1ZNQ) in the cleft region. Residues 192 – 194 differ by 1.1 – 6.5 Å in the two structures, with the largest difference occurring at the carbonyl of Gly193 (Figure 1.5). In fact, this section of the polypeptide chain represents the largest backbone difference between the two structures. This difference is significant because Gly193 in our structure forms intersubunit interactions that help keep the cleft closed. For example, Gly193 links to the other side of the cleft via intersubunit water-mediated hydrogen bonds to the carbonyl of Phe37 and to the main chain and side chain of Asp39 (Figure 1.5). These interactions cannot be formed in 1ZNQ because the distance across the cleft at position 193 is too large.

**Figure 1.5.** *Model of inhibitor NMDBA in the HsGAPDH active site (stereoview).* The 1.75 Å HsGAPDH structure is shown in white. NMDBA is shown in yellow. Residues 191 - 193 of the 2.5 Å human liver GAPDH structure are shown in cyan. The dashed lines indicate intersubunit water-mediate hydrogen bonds present in HsGAPDH. The solid red lines indicate predicted steric interference between the inhibitor and HsGAPDH (contact distance  $\leq 2.7$  Å).





### *Gel Filtration Studies*

CGP-3466 is a deprenyl-related compound (Figure 1.6A) that inhibits the proapoptotic activity of GAPDH (22, 24, 25). However, the mechanism of inhibition and molecular-level details of GAPDH/CGP-3466 interaction are not known. It has been suggested that binding of CGP-3466 or (–)-desmethyldprenyl (DES) (metabolized form of deprenyl) may change the oligomeric state of GAPDH from a tetramer to a dimer or monomer thus halting its proapoptotic activity (24). To investigate this behavior we used gel filtration which indicates that HsGAPDH is found as a tetramer in the precrystallization buffer (Figure 1.7) in the absence of DES or DEP. Using DEP (Figure 1.6B) or DES (Figure 1.6C) at a ratio of compound/HsGAPDH dimer of  $\geq 6$  we incubated each sample and put it across the precrystallization buffer equilibrated column. Figure 1.8A shows a gel filtration chromatogram of HsGAPDH incubated with 10 mM DES. The resultant large peak at 12.8 ml corresponds to the tetrameric form of HsGAPDH. This result was also found when we incubated the protein with 0.4 mM DEP. We then added 0.4 mM DEP to the running buffer to insure that the protein was saturated with the compound throughout the column. The resultant chromatogram (Figure 1.8B) has a peak at 12.93 ml which is not a significant change from the 12.8 for the HsGAPDH tetramer.

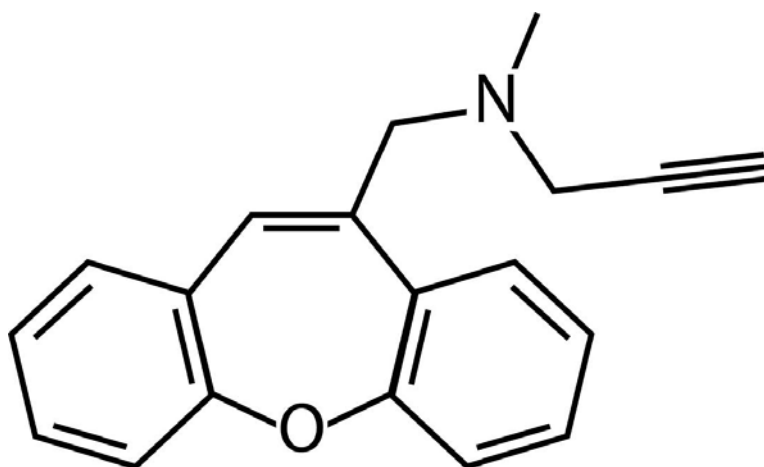
### *Co-crystallization and Soaking Studies*

Co-crystallization of HsGAPDH with DEP was attempted using concentrations of DEP at 10 mM and 2 mM. Also, native crystals were soaked with 15 mM DEP. Diffraction was poor for all crystals grown or soaked with DEP and the unit cell

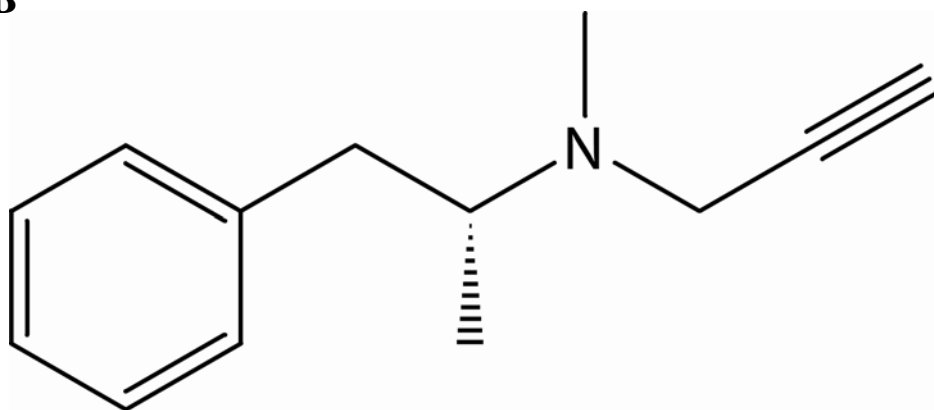
dimensions were different from our crystals as they appeared to be F orthorhombic and not P orthorhombic.

**Figure 1.6.** *Chemical structures of CGP-3466 (A), DEP (B), and DES (C).*

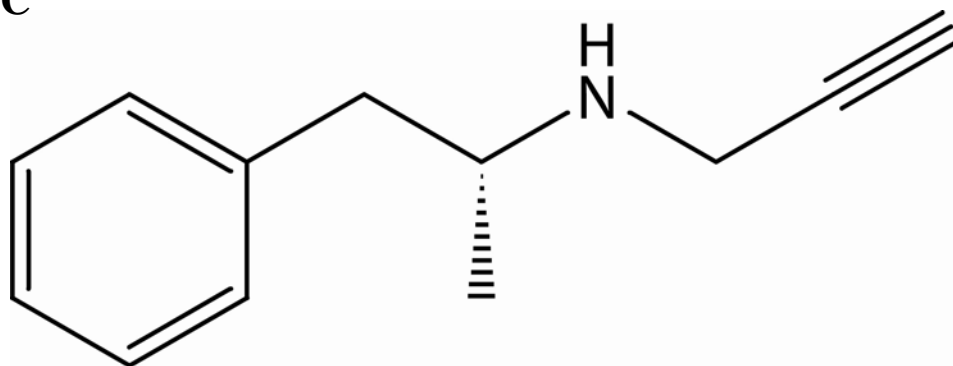
**A**



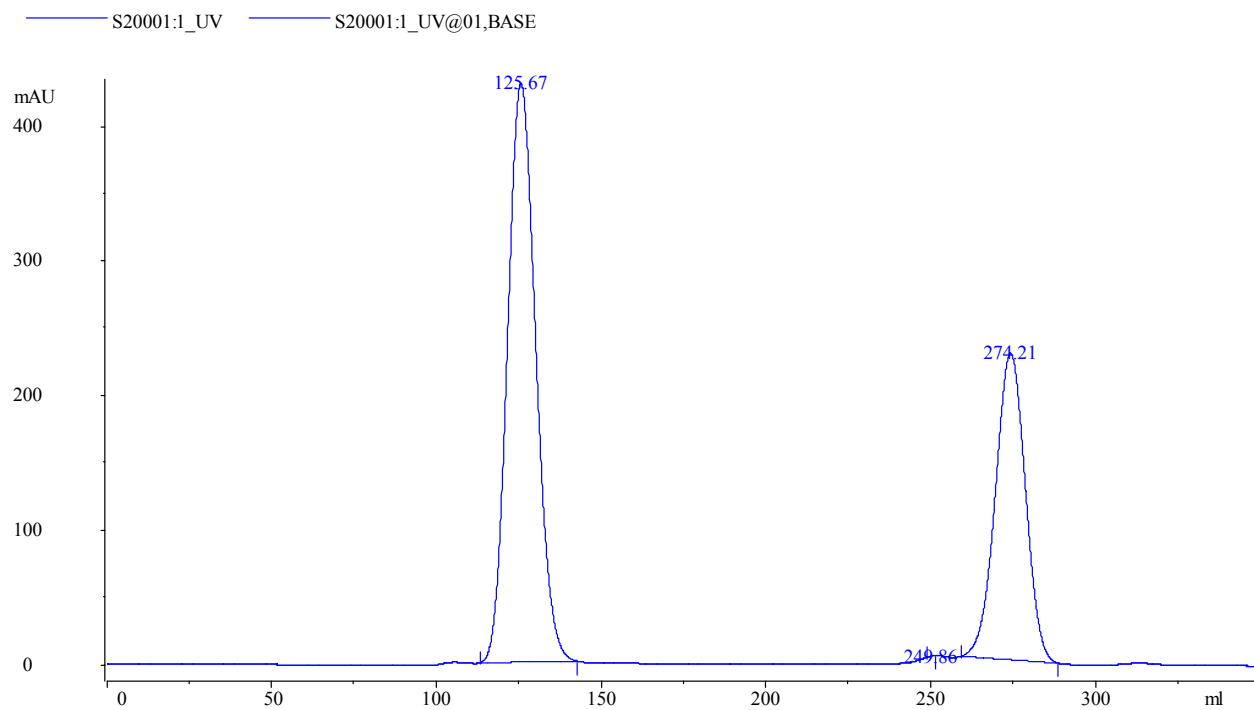
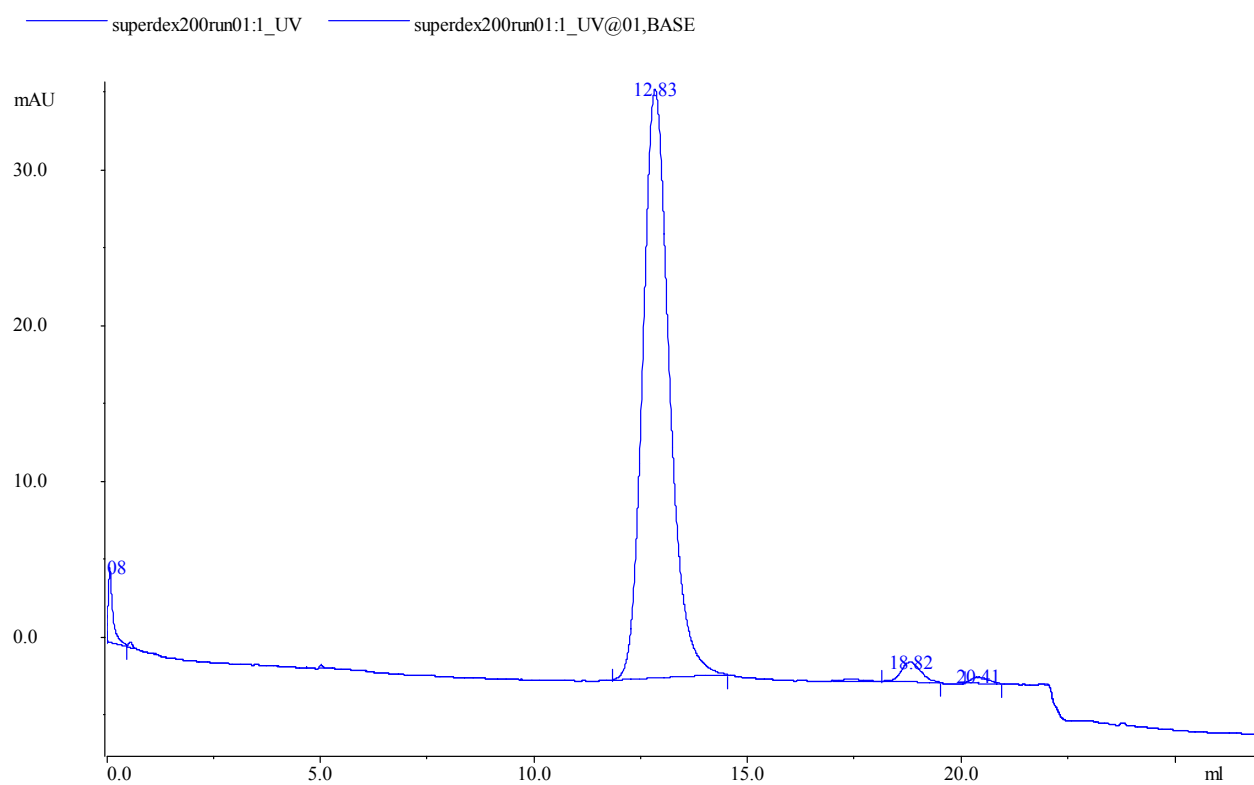
**B**



**C**

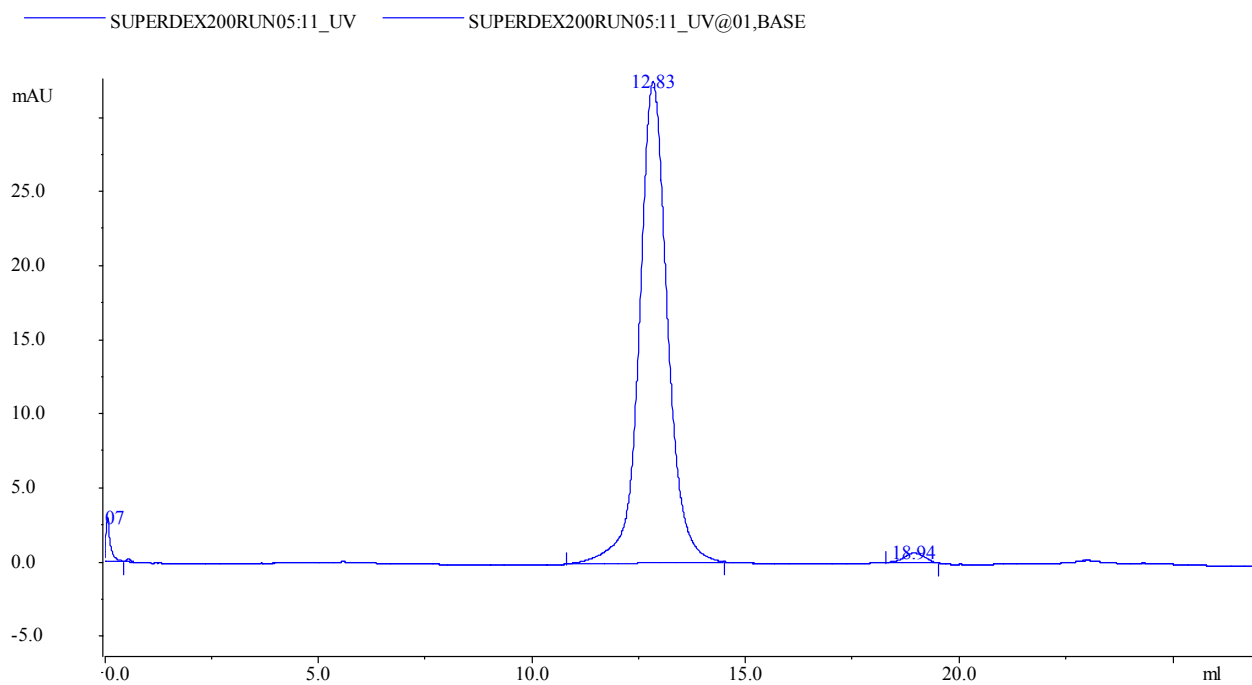


**Figure 1.7.** *Gel filtration of HsGAPDH in precrystallization buffer.* *A.* Chromatogram of HsGAPDH using a S-200 column at 298 K. The first peak, 125.67 ml, represents the HsGAPDH tetramer the second peak is mostly NAD<sup>+</sup>. *B.* Chromatogram of HsGAPDH using a Superdex 200 column at 277 K. The large peak at 12.83 ml represents the HsGAPDH tetramer. For reference Alcohol dehydrogenase, a tetramer of 150 000 kDa, elutes at 13.00 ml on the Superdex 200 column.

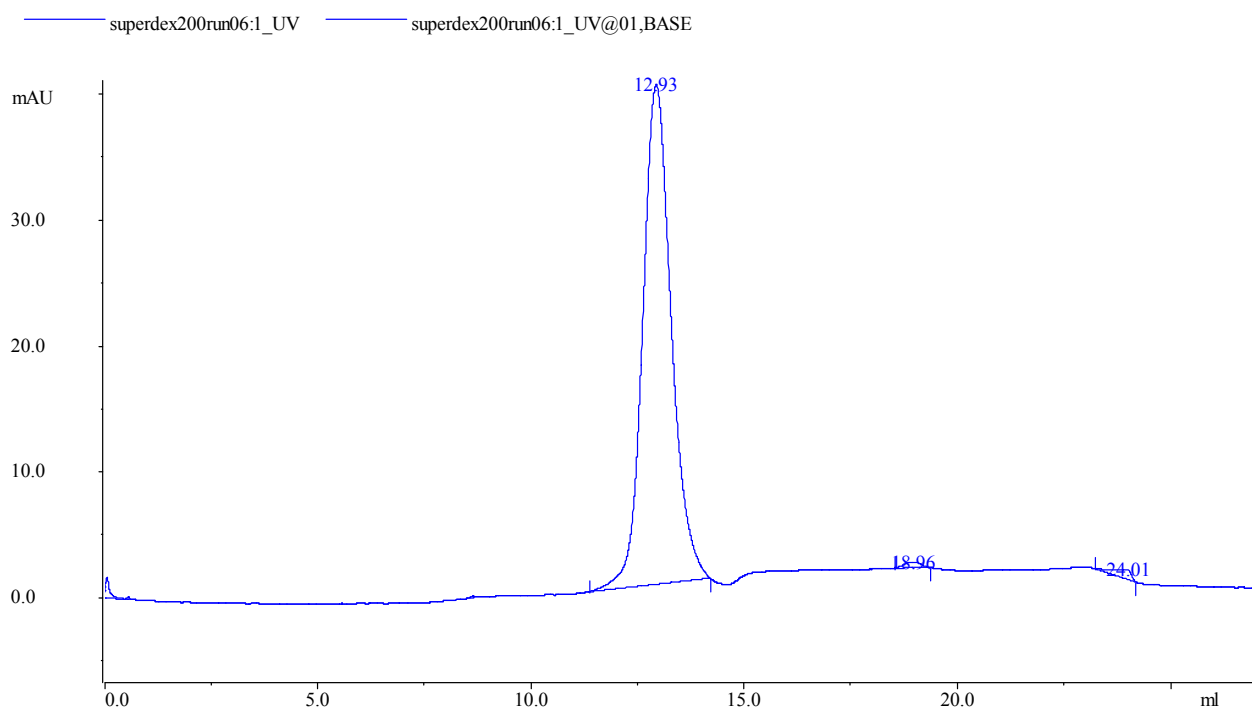
**A****B**

**Figure 1.8.** *Gel filtration of HsGAPDH in the presence of DEP and DES.* *A.* HsGAPDH incubated with 10 mM DES for 60 min at 276 K. *B.* HsGAPDH incubated with 0.4 mM DEP for 30 min at 276 K with an elution buffer of the precrystallization buffer supplemented with 0.4 mM DEP.

**A**



**B**





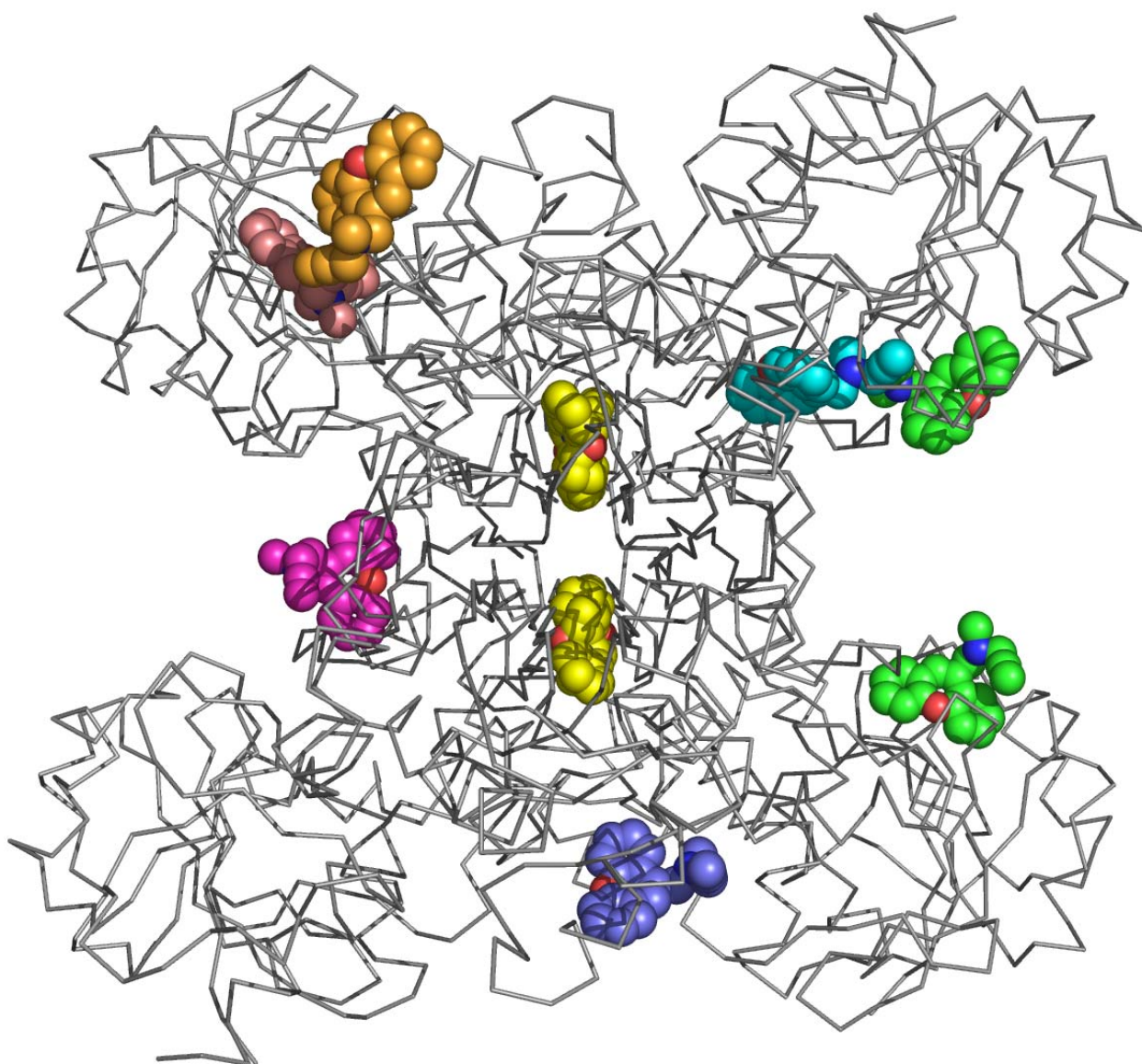
### *Docking Studies of CGP-3466.*

Not having access to CGP-3466, which has a greater antiapoptotic activity than DEP or DES (24), for gel filtration or crystallization studies we instead performed virtual docking calculations using our high resolution HsGAPDH structure and a flexible model of CGP-3466 in order to generate hypotheses regarding plausible binding sites.

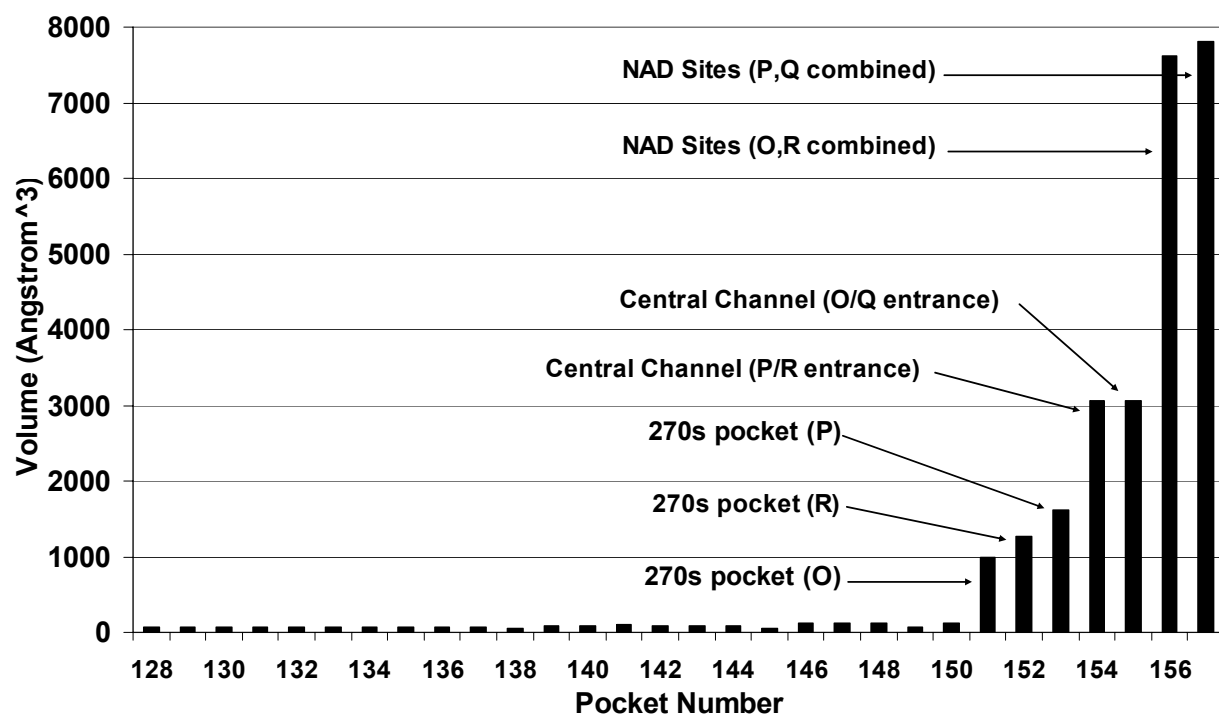
Using the entire GAPDH tetramer as the protein target, AutoDock identified 16 multimember conformational clusters (MMCCs), which had docking scores in the range -10.96 kcal/mol to -8.12 kcal/mol. Two regions of the protein clearly stood out as potential CGP-3466 binding sites: the adenosine pocket (ADE) of the  $\text{NAD}^+$ -binding site and the central channel (CC) located near the intersection of the three molecular twofold axes (Figure. 1.9). Five of the 16 MMCCs, corresponding to 24 of the 100 docking runs, were located in the adenosine pocket of an  $\text{NAD}^+$ -binding site. These docked poses had scores in the range -9.49 to -8.95. Another 5 MMCCs, which represented 18 of the 100 docking runs, were located in the central channel. The central channel poses corresponded to the top 10 docking scores (-10.96 to -9.57 kcal/mol). Thus, the adenosine pocket and central channel together accounted for almost two-thirds of the 16 MMCCs and nearly half of the 100 docking runs. No other region of the protein was implicated with a higher frequency or better docking score by our initial docking calculation.

To complement and validate the global docking results, we used CASTP (70) to survey the surface topography of the tetramer to identify pockets large enough to serve as ligand-binding sites. CASTP found 157 pockets in the GAPDH tetramer. These pockets had surface areas in the range 1 – 3514  $\text{\AA}^2$ , with an average of 120  $\text{\AA}^2$ . The pocket volumes were in the range 3 – 7814  $\text{\AA}^3$ , with an average volume of 191  $\text{\AA}^3$  (Figure 1.10).

**Figure 1.9.** *HsGAPDH tetramer with space fill molecules representing the seven preferred clustering sites of CGP 3466.* This is a ribbon structure of HsGAPDH showing the most popular clustering sites of CGP 3466 the sites: Central channel (yellow), NAD<sup>+</sup> adenine (light green), NAD<sup>+</sup> nicotinamide (cyan), NAD<sup>+</sup> grove (purple), Intersubunit (rust), and NES (blue and gold).



**Figure 1.10.** *CASTP analysis of HsGAPDH pockets.* The plot shows the volumes of the 30 largest pockets found by CASTP. Locations of pockets with significant volume are indicated.



Note that the two largest pockets each correspond to two NAD binding sites, which explains their large CASTP volumes ( $>7500 \text{ \AA}^3$ ). Only three pockets identified by CASTP are large enough to accommodate CGP-3466, which has a molecular volume of  $255 \text{ \AA}^3$ : stood out in terms of surface area and volume: central channel,  $\text{NAD}^+$ -binding site and a surface indentation located near the 274 – 277  $\beta$ -strand (Figure 1.11).

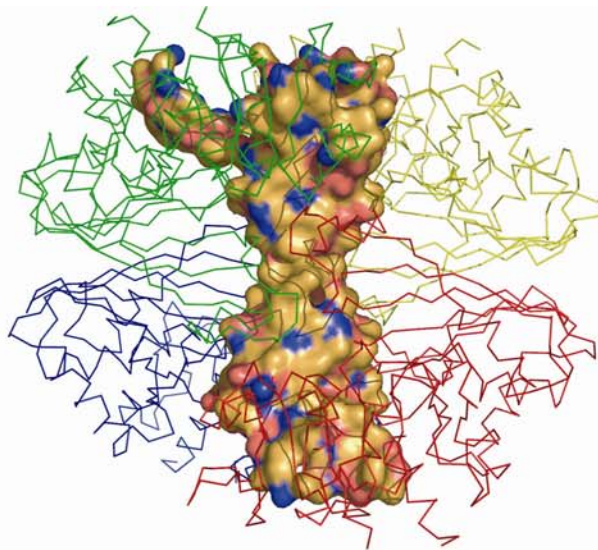
The 270s-strand pocket identified by CASTP is a shallow bowl-shaped indentation near a region of the protein that contains a novel CRM1-mediated nuclear export signal (71). The  $\beta$ -strand consisting of Leu274, Gly275, Tyr276 and Thr277 forms the bottom of the bowl. The rim of the bowl is formed by Lys263, Glu267, Lys271, Glu278, His279, Phe286, His291, and Thr294 of one subunit, and Tyr49 and Lys55 from an NCS-related (Q-axis) subunit (Figure 1.12). In our global docking calculation, this site corresponded to two MMCCs, representing 5 out of 100 docking runs with docking scores in the range -8.71 to -8.29 kcal/mol. Focused docking calculations on this region revealed two prevalent conformations with a combined 85 of the 100 docking runs (Figure 1.12B). Although this site is large enough to accommodate CGP-3466, it is quite hydrophilic due to the many charged and polar residues of the rim, which seems incompatible with binding a very hydrophobic molecule like CGP-3466. Since the adenosine pocket and central channel were implicated by both the global docking calculation and by CASTP, and these sites are appropriately hydrophobic, they were explored further using focused docking calculations. The focused calculations each consisted of 100 docking runs and employed smaller grids centered on the regions of interest. Results of these calculations are summarized in Table 1.3.

**Figure 1.11.** *Surface view of the top seven CASTP predicted pockets for HsGAPDH.*

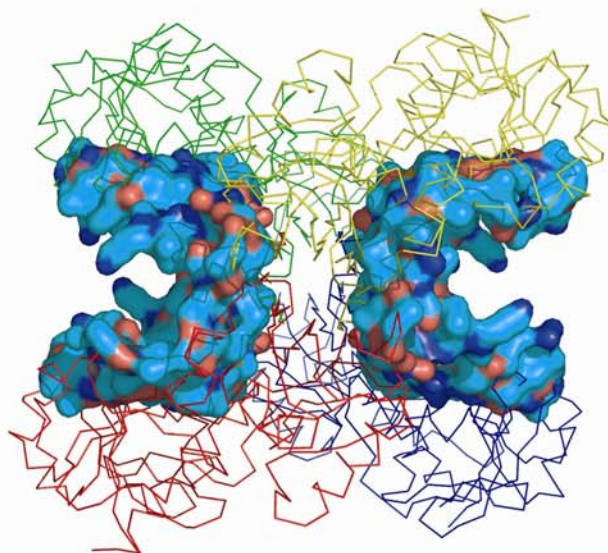
The subunits in ribbon form are colored as follows: O, yellow; P, red; Q, green; R, blue.

*A*, a view down the R axis of the two central channel pockets, one of which includes the fourth 270s pocket. *B*, detailed view of the NAD sites as viewed down the P axis. *C*, the three 270s pockets as viewed down the R axis.

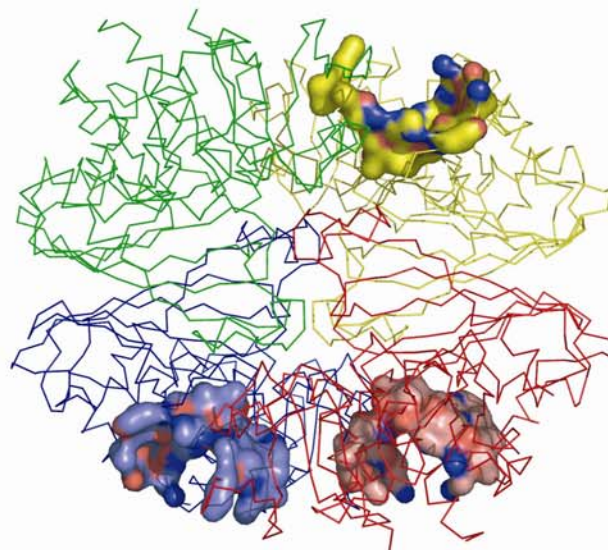
**A**



**B**



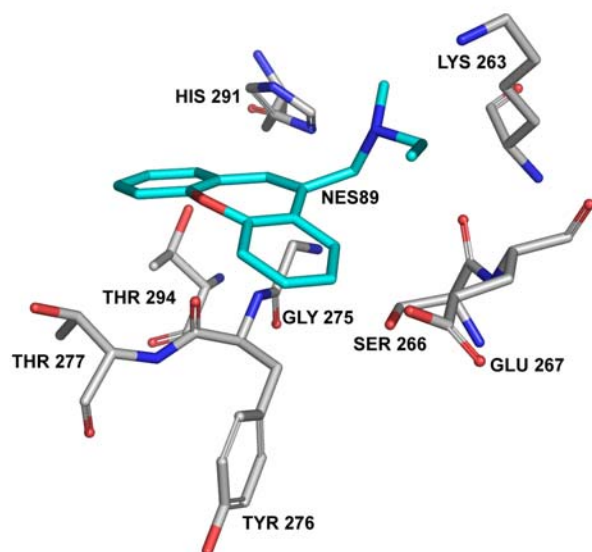
**C**



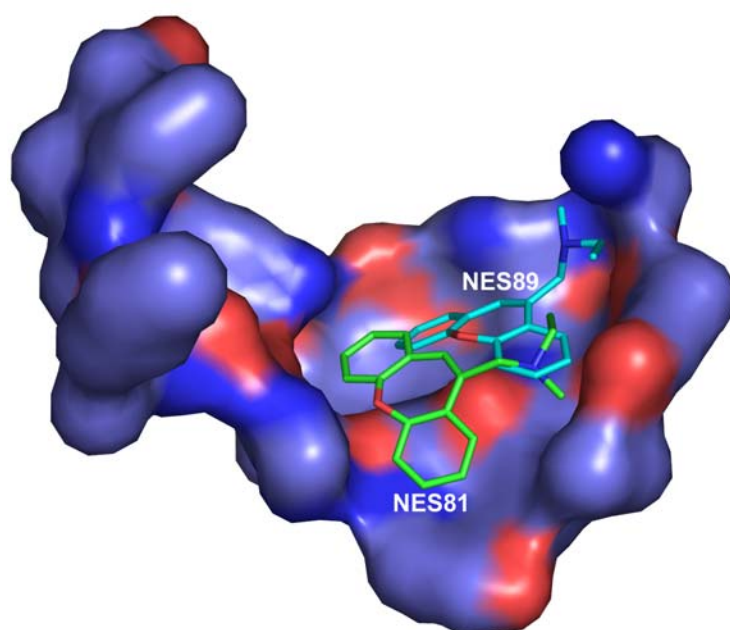


**Figure 1.12.** *Predicted binding sites for CGP-3466 in 270s binding pocket. A, docked CGP-3466 conformation NES89. Residues shown in grey are within 3.9 Å. B, surface view of the 270s pocket. Bound in the pocket are the two highest scoring MMCCs NES89 (60 individuals) and NES81 (25 individuals) from our 270s strand-pocket focused docking run.*

**A**



**B**



A docking calculation focused on the adenosine pocket revealed three major clusters of poses, which had energies in the range -9.42 to -8.92 kcal/mol (Table 1.3). The docked ligands are 76 – 82 % buried and the CGP-3466 ring system occupies the same location as the adenine ring system of NAD<sup>+</sup> (Figure 1.13A). The overlap between adenine and the CGP-3466 ring system is particularly good for the two lowest energy conformations, ADE79 and ADE69 (Figure 1.13A). Residues forming major contacts with all three docked ligands, based on hydrophobic ligand-protein contact surface area calculated with LPC (50), include Pro36, Phe37, Thr99, Val101 and Phe102. Note that residues 36, 37, 99 and 102 also form hydrophobic contacts with the NAD<sup>+</sup> adenine (Figure 1.3B).

The docking calculation focused on the central channel produced the top docking scores of our study. This calculation revealed two major poses, which together accounted for 90 of the 100 docking runs (Table 1.3, CCH46 and CCH60). The docked ligands are almost completely buried (94 %) and contact several residues (Table 1.3, Figure 1.13B). The binding site exhibits symmetry due to its location near the intersection of the molecular 2-fold axes and thus all four chains contribute to the site. The major contact residues for CC46 are Leu203 (P, R), Gln204 (P, R), Thr237 (O, Q), Asn239 (O) and Ala238 (O) (Figure 1.13B). Major contact residues for CCH60 are Leu203 (P, R), Gln204 (P, R), Ala238 (O, Q), Asn239 (O, Q) and Thr52 (Q) (Figure 1.13B). In addition, the heteroatom oxygen atoms of CCH46 and CCH60 are predicted to form hydrogen bonds with Ala238 and Gln204, respectively (Figure 1.13B).

Note that three additional, equivalent central channel binding sites are related by 2-fold rotations around the P, Q and R axes of the tetramer. In fact, the CASTP analysis

**Table 1.3.** *Summary of CGP-3466 docking results.*

Ligand pose	Location	Docking Score (kcal/mol)	Cluster Size <sup>a</sup>	Buried Surface Area (%) <sup>b</sup>	Contact Residues
CCH46 <sup>c</sup>	Central channel	-10.96	80	94	L203, Q204, P236, T237, A238 <sup>d</sup> , N239, S283, S284, N287
CCH60	Central channel	-10.34	10	94	S51, T52, L203, Q204 <sup>d</sup> , P236, T237, A238, N239, N287
ADE79 <sup>c</sup>	Adenosine pocket	-9.42	38	82	N9, G10, G12, N34, D35, P36, F37, P191, R80, T99, V101, F102
ADE69	Adenosine pocket	-9.21	25	76	N9, G10, G12, N34, D35, P36, F37, R80, T99, V101, F102
ADE42	Adenosine pocket	-8.92	15	80	N9, G10, G12, N34, D35, P36, F37, T99, V101, F102

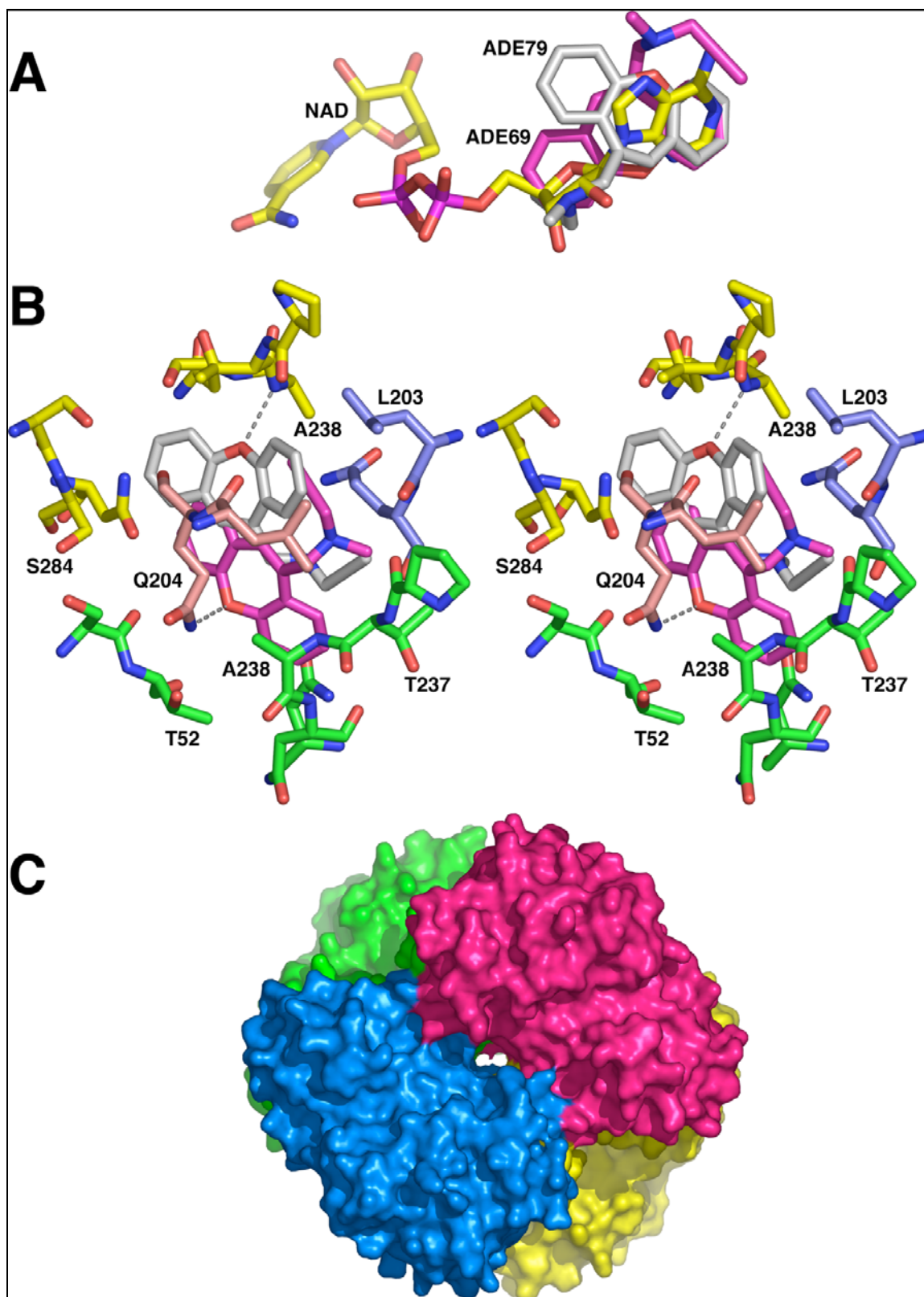
<sup>a</sup> Refers to a 100-run docking calculation focused on the indicated location.

<sup>b</sup> Expressed as a percentage of the solvent accessible surface area of the uncomplexed ligand.

<sup>c</sup> Denotes lowest energy conformation of a 100-run docking calculation focused on the indicated location.

<sup>d</sup> Forms hydrogen bond with docked ligand.

**Figure 1.13.** *Predicted binding sites for CGP-3466.* *A*, adenosine pocket. Docked CGP-3466 poses ADE79 and ADE69 are shown in white and magenta, respectively. For reference, NAD<sup>+</sup> is included in yellow. *B*, detailed view of the predicted CGP-3466 binding site in the central channel. GAPDH subunits are colored as follows: O, yellow; P, pink; Q, green; R, blue. Docked CGP-3466 poses CC46 and CC60 are shown in white and magenta, respectively. *C*, empty central channel viewed through the P/R entrance (down the Q axis). GAPDH subunits are colored as follows: O, yellow; P, red; Q, green; R, blue.



revealed that the central channel actually consists of two distinct, symmetry-related chambers having separate entrances located at opposite ends of the Q-axis (Figures 1.10 and 1.11A). Division of the central channel into two chambers is caused by a constriction of the channel at the center of the tetramer (Figure 1.13C).

#### *Structural insights into GAPDH-Siah1 interaction*

Truncation mutagenesis data suggest that GAPDH residues 222 - 240 and Siah1 residues 270 - 282 are essential for GAPDH-Siah1 association (27). Moreover, mutagenesis of GAPDH Lys227 abrogates GAPDH-Siah1 complex formation (27), which suggests that Lys227 directly interacts with Siah1 in the interface between the two proteins.

GAPDH residues 222 - 240 begin with a solvent-exposed loop (residues 222 - 227) and continue through a  $\beta$ -strand (228 - 234) that travels toward the central channel, finishing in a loop (residues 235 - 240) inside the channel (Figure 1.14A). This  $\beta$ -strand is part of the mixed, twisted  $\beta$ -sheet of the catalytic domain that forms the extensive P interface and thus it makes critical intersubunit interactions. Note that Lys227 is located on a solvent-exposed loop (Figure 1.14A), which is consistent with its predicted involvement in complex formation. On the other hand, the  $\beta$ -strand (residues 228 - 234) and the central channel would be inaccessible to Siah1; assuming GAPDH retains the tetrameric form that we observe in the crystal structure.

Siah1 residues 270 - 282 form the last  $\beta$ -strand ( $\beta$ 8) of a curved, anti-parallel 4-stranded  $\beta$ -sheet located in the C-terminus of the protein (Figure 1.14B and (54)). This  $\beta$ -sheet joins with the analogous sheet from the opposite subunit of the Siah1 dimer to form

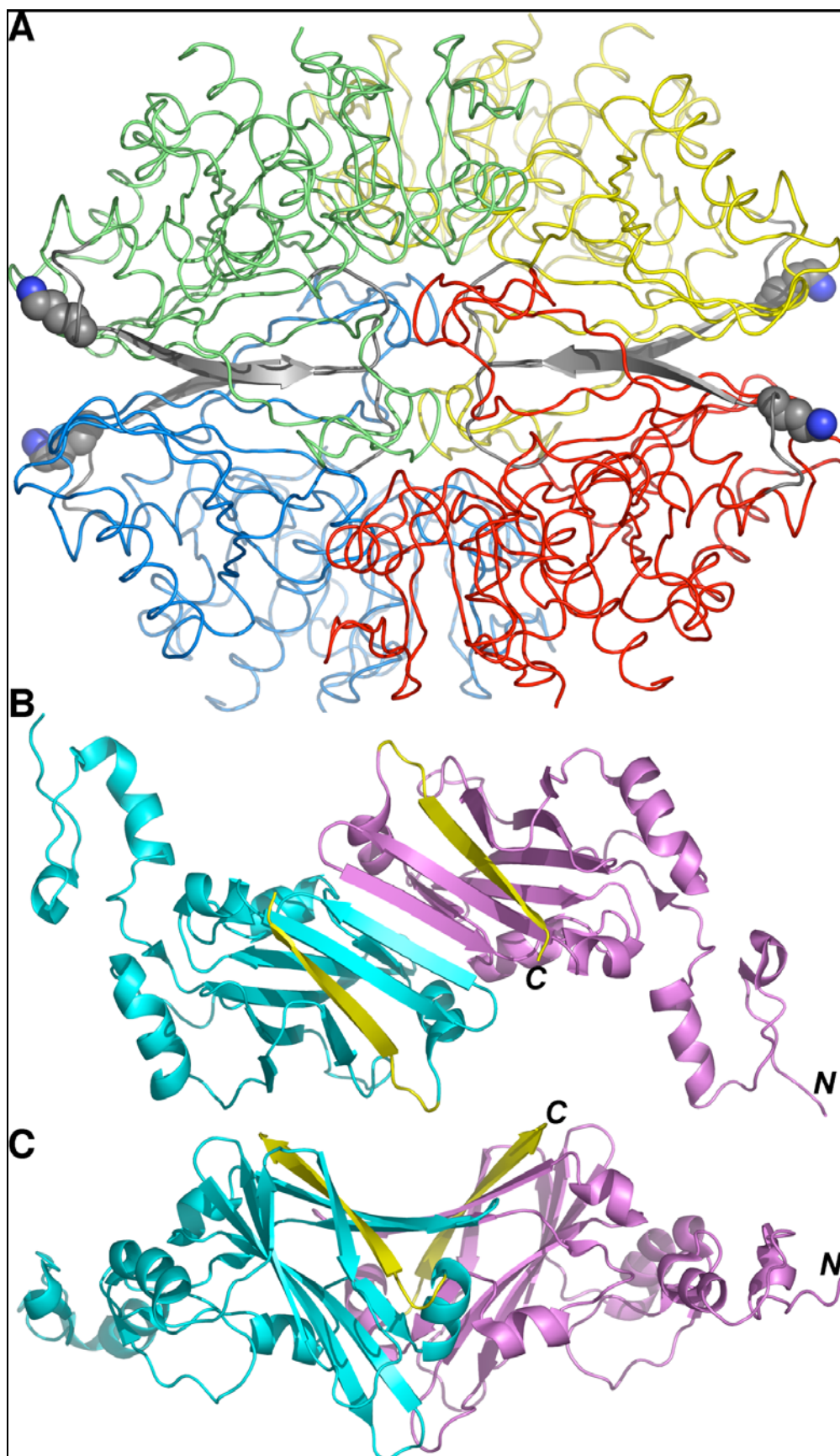
an intermolecular  $\beta$ -sheet, which results in a large (30 Å wide), shallow groove (Figure 1.14C).

We used the protein-protein docking program PatchDock to build a qualitative model of the GAPDH-Siah1 complex, subject to the constraint that GAPDH Lys277 is in the binding interface. Interestingly, the complex having the top score from PatchDock featured the large, shallow groove of Siah1 contacting the convex surface of GAPDH near Lys227 (Figure 1.15A). In this model, GAPDH Lys227 interacts with Siah1  $\beta$ 8 via a hydrogen bond with the hydroxyl of Siah1 Ser280 (Figure 1.15B). Thus, our model is consistent with the experimental data of Hara et al., which implicates GAPDH Lys227 and Siah1  $\beta$ 8 as being essential for GAPDH-Siah1 association. Moreover, the model is consistent with the assertion of Polekhina et al. that the large, shallow groove of Siah1 plays a role in recognizing and interacting with other proteins (54).

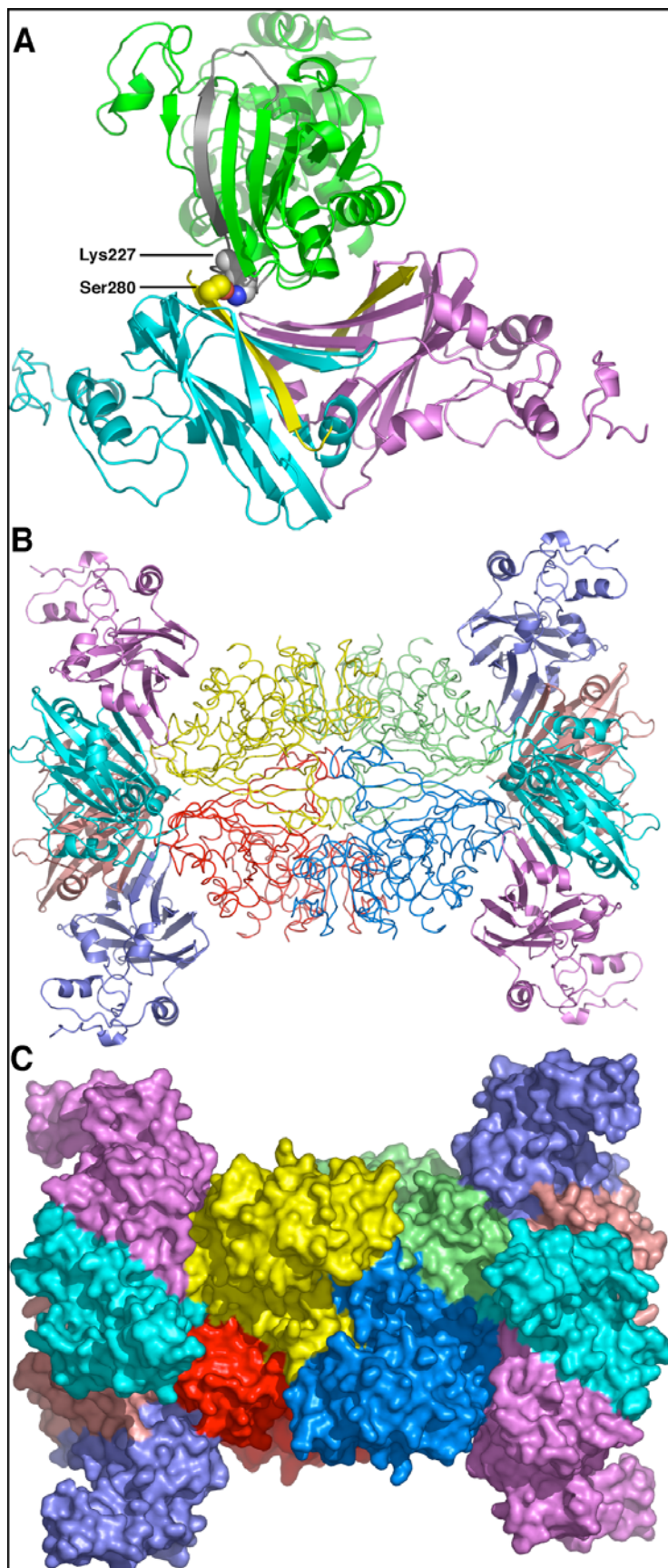
We expanded the GAPDH-Siah1 model using the symmetry of the GAPDH tetramer and found that no steric clashes were introduced (Figures 1.15B, C). The expanded model contains one GAPDH tetramer interacting with 4 Siah1 dimers. Note that our model implies a binding stoichiometry of 1:2 (GAPDH: Siah1), which could be tested experimentally.



**Figure 1.14.** *Structures of GAPDH and Siah1 highlighting regions that are important for GAPDH-Siah1 association.* *A*, HsGAPDH tetramer viewed down the R-axis. Subunits are colored as follows: O, yellow; P, red; Q, green; R, blue. Residues that are essential for complex formation with Siah1 are colored gray (residues 222 - 240) and atoms of Lys227 are drawn as spheres. *B*, homodimer of Siah1 from PDB entry 1K2F (51). The two subunits of Siah1 are colored cyan and violet. Residues that are essential for interaction with GAPDH are colored yellow (residues 270- 282). *C*, homodimer of Siah1 rotated 90° from orientation in panel *B*. The coloring scheme of panel *B* is used.



**Figure 1.15.** *Model of GAPDH-Siah1 complex.* *A*, docked complex having the top score from PatchDock. A GAPDH subunit is shown in green and the Siah1 dimer is shown in cyan/violet. Only one subunit of the GAPDH dimer used for docking is shown for clarity. GAPDH residues 222 – 240 are colored gray. Siah1 residues 270 – 282 are colored yellow. GAPDH Lys227 and Siah1 Ser280 are drawn as spheres. *B*, model of a GAPDH tetramer interacting with four Siah1 dimers. This model was generated from the model in panel *A* using the symmetry of the GAPDH tetramer. The view is looking down the GAPDH R axis. GAPDH subunits are colored as follows: O, yellow; P, red; Q, green; R, blue. Siah1 dimers are colored cyan/violet and salmon/slate. *C*, surface representation of the model shown in panel *B*.



## DISCUSSION

Our main motivation for engineering a recombinant HsGAPDH and determining a high resolution crystal structure was to provide a better structural foundation for drug design and biophysical studies related to GAPDH's moonlighting functions. When we started this work, the only human GAPDH structure in the PDB was Watson's 3.5 Å structure determined from twinned crystals (72). Since then, a 2.5 Å structure of human liver GAPDH has been deposited in the PDB (35). However, the crystallographic resolution of 1ZNQ is only 2.5 Å and the structure was determined from a data set with modest completeness (82 % overall, 58 % in the high resolution bin). Thus, our structure is the highest resolution structure of a human (or mammalian) GAPDH to date and represents a significant improvement over the previously available models.

### *Selectivity Cleft*

Our structure provides the first high resolution view of a human GAPDH cofactor binding site, which is relevant to design of inhibitors that specifically target trypanosomatid GAPDHs. In particular, we examined the proposed "selectivity cleft" near the adenosine binding site and found it to be 3 – 4 Å narrower than that of *L. mexicana* GAPDH, that is, effectively closed.

Moreover, our structure revealed a new feature of the cleft: water-mediated intersubunit hydrogen bonds. These interactions appear to help maintain the cleft in the closed orientation and thus they represent a structural determinant of selectivity of NMDBA and similar drugs toward trypanosomatid GAPDHs. These interactions are not present in the human liver GAPDH structure due to large conformational differences at

residues 192 -195 (Figure 1.5). We note that the conformation of residues 192 – 195 of rabbit GAPDH (1J0X) matches that of our structure, although the intricate water-mediated hydrogen bond network was not observed in 1J0X due, perhaps, to its moderate crystallographic resolution of 2.4 Å.

The underlying reason for the large conformational difference between the two human GAPDH structures is unclear. Neither structure is compromised by strong crystal contacts in this particular region. Perhaps the difference may reflect an inherent flexibility in this region of the protein. It remains to be seen whether such flexibility might allow certain anti-trypanosomatid GAPDH inhibitors to bind to the human enzyme.

#### *Docking Studies of CGP-3466*

CGP-3466 inhibits the proapoptotic activity of GAPDH by preventing its nuclear localization by an unknown mechanism. Since the molecular mechanism of this compound is poorly understood, we performed *in silico* docking studies to gain hypotheses about potential binding sites.

Our calculations predicted the adenosine pocket as a possible binding site, which is consistent with the observation that binding of rabbit GAPDH to immobilized CGP-3466 is inhibited by NAD<sup>+</sup> as measured by surface plasmon resonance (25). Also, Cowan-Jacob *et al.* (56) reported crystallographic evidence for binding of CGP-3466 to the adenosine pocket of rabbit GAPDH, however, the electron density representing the inhibitor was extremely weak and consequently CGP-3466 was not included in PDB entry 1J0X. Interestingly, one of the ligand orientations identified by our study (ADE69)

is very similar to that reported by Cowan-Jacob *et al.* (56), which provides additional support for the adenosine pocket as a potential binding site for CGP-3466. Site-directed mutagenesis of Pro36, Phe37, Thr99, Val101 and Phe102 could be used to test the importance of this proposed binding site.

Our docking calculations also suggested the central channel as a plausible binding site for CGP-3466. In fact, this site produced the best docking scores. This prediction is consistent with the experimental work of Carlile *et al.*, which showed that an antibody raised against residues that block the channel entrance prevent binding of BODIPY-labeled CGP-3466 (24). Primary candidates for site-directed mutagenesis studies to test the importance of this binding site include Leu203 and Gln204.

#### *NO-S-nitrosylation-GAPDH-Siah1 Cascade*

Hara and coworkers showed that nitrosylation of the GAPDH active site Cys (-SNO) enhances GAPDH-Siah1 association (27). Hydrolysis of S-nitrosylated GAPDH produces various oxidized forms such as the sulfenic (-SOH), sulfinic (-SO<sub>2</sub>H) and sulfonic (-SO<sub>3</sub>H) acids. The latter form of GAPDH was isolated from apoptotic HEK293 cells and identified by mass spectrometry (27). Presumably, modification of the active site Cys causes conformational changes in GAPDH that enhance GAPDH-Siah1 interaction. However, it is unlikely that this rather small chemical modification produces large conformational changes in GAPDH. In fact, the sulfonic acid form of the active site Cys is present in the crystal structure of *A. xylosoxidans* GAPDH (55). This protein displays the classic GAPDH tetrameric form with no significant global structural changes

(55). Thus, it is not clear why S-nitrosylation and subsequent oxidation of GAPDH promotes association with Siah1.

Discovery of the NO-GAPDH-Siah1 cascade forces rethinking of the mechanisms of antiapoptosis compounds such as CGP-3466. Experimental work and molecular modeling, including our docking study, predict CGP-3466 binding sites in the adenosine pocket and central channel. There are at least two plausible mechanisms by which CGP-3466 binding in the cofactor site could disrupt the NO-GAPDH-Siah1 cascade. Firstly, CGP-3466 could protect the active site Cys from reaction with NO. Second, binding of CGP-3466 in the cofactor site could cause a change in the tertiary and/or quaternary structure of GAPDH that reduces GAPDH-Siah1 affinity. This scenario is plausible because binding of  $\text{NAD}^+$  is known to cause tertiary and quaternary structural changes that underlie cooperativity (63, 73).

Considering the proposed central channel binding site for CGP-3466, the channel itself is too narrow for Siah1 to enter and thus it is unlikely that CGP-3466 competes with Siah1 for a central channel binding site. Given that the central channel is located within intersubunit interfaces, it seems more likely that binding of CGP-3466 in the central channel causes quaternary structural adjustments that decrease GAPDH-Siah1 affinity. We note that our model of GAPDH-Siah1 shows Siah1 dimers packed closely together on the surface of GAPDH (Figures 1.15B, C). Thus, subtle changes in GAPDH quaternary structure could cause steric interference between the bound Siah1 dimers. Ultimately, biophysical and structural studies will be needed to understand the details of GAPDH-Siah1 association and to design drugs that intersect the NO-GAPDH-Siah1 cascade.



## **Acknowledgements**

Predoctoral support for J.L.J. was provided, in part, by the National Science Foundation and the Missouri Alliance for Graduate Education and the Professorate (MAGEP). We thank M.A. Sirover for his generous gift of pChug 20.2 and Ming-Yi Zhou of the DNA Core Facility at the University of Missouri-Columbia for preparation of the HsGAPDH plasmid. We thank the personnel of APS beamline 19-ID for assistance with data collection, especially Stephen L. Ginell. Use of the Argonne National Laboratory Structural Biology Center beamlines at the APS was supported by the U. S. Department of Energy, Office of Energy Research, under Contract No. W-31-109-ENG-38.

## REFERENCES

1. Sherman, I. W. (1998) pp 135-143, ASM Press, Washington, DC.
2. de Marchi, A. A., Castilho, M. S., Nascimento, P. G., Archanjo, F. C., del Ponte, G., Oliva, G., and Pupo, M. T. (2004) *Bioorg Med Chem* 12, 4823-33.
3. Ladame, S., Castilho, M. S., Silva, C. H., Denier, C., Hannaert, V., Perie, J., Oliva, G., and Willson, M. (2003) *Eur J Biochem* 270, 4574-86.
4. Leitao, A., Andricopulo, A. D., Oliva, G., Pupo, M. T., de Marchi, A. A., Vieira, P. C., da Silva, M. F., Ferreira, V. F., de Souza, M. C., Sa, M. M., Moraes, V. R., and Montanari, C. A. (2004) *Bioorg Med Chem Lett* 14, 2199-204.
5. Menezes, I. R., Lopes, J. C., Montanari, C. A., Oliva, G., Pavao, F., Castilho, M. S., Vieira, P. C., and Pupo, M. T. (2003) *J Comput Aided Mol Des* 17, 277-90.
6. Callens, M., and Hannaert, V. (1995) *Annals of Tropical Medicine & Parasitology* 89 Suppl 1, 23-30.
7. Kim, H., Feil, I. K., Verlinde, C. L., Petra, P. H., and Hol, W. G. (1995) *Biochemistry* 34, 14975-86.
8. Kim, H., and Hol, W. G. (1998) *Journal of Molecular Biology* 278, 5-11.
9. Aronov, A. M., Verlinde, C. L., Hol, W. G., and Gelb, M. H. (1998) *Journal of Medicinal Chemistry* 41, 4790-9.
10. Verlinde, C. L., Callens, M., Van Calenbergh, S., Van Aerschot, A., Herdewijn, P., Hannaert, V., Michels, P. A., Opperdoes, F. R., and Hol, W. G. (1994) *J Med Chem* 37, 3605-13.
11. Bressi, J. C., Verlinde, C. L., Aronov, A. M., Shaw, M. L., Shin, S. S., Nguyen, L. N., Suresh, S., Buckner, F. S., Van Voorhis, W. C., Kuntz, I. D., Hol, W. G., and Gelb, M. H. (2001) *Journal of Medicinal Chemistry* 44, 2080-93.
12. Suresh, S., Bressi, J. C., Kennedy, K. J., Verlinde, C. L., Gelb, M. H., and Hol, W. G. (2001) *Journal of Molecular Biology* 309, 423-35.
13. Aronov, A. M., Suresh, S., Buckner, F. S., Van Voorhis, W. C., Verlinde, C. L., Opperdoes, F. R., Hol, W. G., and Gelb, M. H. (1999) *Proc Natl Acad Sci U S A* 96, 4273-8.
14. Ladame, S., Bardet, M., Perie, J., and Willson, M. (2001) *Bioorganic & Medicinal Chemistry* 9, 773-83.
15. Daubenberger, C. A., Poltl-Frank, F., Jiang, G., Lipp, J., Certa, U., and Pluschke, G. (2000) *Gene* 246, 255-64.
16. Castilho, M. S., Pavao, F., Oliva, G., Ladame, S., Willson, M., and Perie, J. (2003) *Biochemistry* 42, 7143-51.
17. Vellieux, F. M. (1995) *Acta Crystallogr D Biol Crystallogr* 51, 575-89.
18. Jeffery, C. J. (1999) *Trends Biochem Sci* 24, 8-11.
19. Sirover, M. A. (1996) *Life Sciences* 58, 2271-7.
20. Sirover, M. A. (1997) *Journal of Cellular Biochemistry* 66, 133-40.
21. Sirover, M. A. (1999) *Biochimica et Biophysica Acta* 1432, 159-84.
22. Tatton, W., Chalmers-Redman, R., and Tatton, N. (2003) *J Neural Transm* 110, 509-15.
23. Tatton, W. G., Chalmers-Redman, R., Brown, D., and Tatton, N. (2003) *Ann Neurol* 53 Suppl 3, S61-70; discussion S70-2.

24. Carlile, G. W., Chalmers-Redman, R. M., Tatton, N. A., Pong, A., Borden, K. E., and Tatton, W. G. (2000) *Mol Pharmacol* 57, 2-12.
25. Kragten, E., Lalande, I., Zimmermann, K., Roggo, S., Schindler, P., Muller, D., van Oostrum, J., Waldmeier, P., and Furst, P. (1998) *J Biol Chem* 273, 5821-8.
26. Burke, J. R., Enghild, J. J., Martin, M. E., Jou, Y. S., Myers, R. M., Roses, A. D., Vance, J. M., and Strittmatter, W. J. (1996) *Nat Med* 2, 347-50.
27. Hara, M. R., Agrawal, N., Kim, S. F., Cascio, M. B., Fujimuro, M., Ozeki, Y., Takahashi, M., Cheah, J. H., Tankou, S. K., Hester, L. D., Ferris, C. D., Hayward, S. D., Snyder, S. H., and Sawa, A. (2005) *Nat Cell Biol* 7, 665-74.
28. Saunders, P. A., Chen, R. W., and Chuang, D. M. (1999) *J Neurochem* 72, 925-32.
29. Berry, M. D., and Boulton, A. A. (2000) *J Neurosci Res* 60, 150-4.
30. Ishitani, R., Tajima, H., Takata, H., Tsuchiya, K., Kuwae, T., Yamada, M., Takahashi, H., Tatton, N. A., and Katsube, N. (2003) *Prog Neuropsychopharmacol Biol Psychiatry* 27, 291-301.
31. Maruyama, W., Akao, Y., Youdim, M. B., Davis, B. A., and Naoi, M. (2001) *J Neurochem* 78, 727-35.
32. Dastoor, Z., and Dreyer, J. L. (2001) *Journal of Cell Science* 114, 1643-53.
33. Katsube, N., Sunaga, K., Aishita, H., Chuang, D. M., and Ishitani, R. (1999) *J Pharmacol Exp Ther* 288, 6-13.
34. Warizaya, M., Kinoshita, T., Kato, A., Nakajima, H., and Fujii, T. (2004) *Acta Crystallogr D Biol Crystallogr* 60, 567-8.
35. Ismail, S. A., and Park, H. W. (2005) *Acta Crystallogr D Biol Crystallogr* 61, 1508-13.
36. Vollberg, T. M., Siegler, K. M., Cool, B. L., and Sirover, M. A. (1989) *Proc. Natl Acad. Sci. USA*, 86, 8693-8697.
37. Matthews, B. W. (1968) *J Mol Biol* 33, 491-7.
38. Otwinoski, Z., and Minor, W. (1997) *Methods Enzymol.* 276, 307-326.
39. Engh, R. A., and Huber, R. (1991) in *Acta Crystallogr A* pp 392-400.
40. Laskowski, R. A., Moss, D. S., and Thornton, J. M. (1993) *J Mol Biol* 231, 1049-67.
41. Navaza, J. (1994) *Acta Cryst.* A50, 157-163.
42. Tanner, J. J., Hecht, R. M., and Krause, K. L. (1996) *Biochemistry* 35, 2597-609.
43. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr D Biol Crystallogr* 54 ( Pt 5), 905-21.
44. Morris, R. J., Perrakis, A., and Lamzin, V. S. (2002) *Acta Crystallogr D Biol Crystallogr* 58, 968-75.
45. Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) *Acta Crystallogr D Biol Crystallogr* 59, 1131-7.
46. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard. (1991) *Acta Crystallogr A* 47 ( Pt 2), 110-9.
47. DeLano, W. L. (2002), DeLano Scientific, San Carlos, CA.
48. Shindyalov, I. N., and Bourne, P. E. (1998) *Protein Eng* 11, 739-47.

49. Dror, O., Benyamini, H., Nussinov, R., and Wolfson, H. (2003) *Bioinformatics* 19 Suppl 1, i95-104.
50. Sali, A., and Blundell, T. L. (1990) *J Mol Biol* 212, 403-28.
51. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998) *J Comput Chem* 19, 1639-1662.
52. van Aalten, D. M., Bywater, R., Findlay, J. B., Hendlich, M., Hooft, R. W., and Vriend, G. (1996) *J Comput Aided Mol Des* 10, 255-62.
53. Sobolev, V., Sorokine, A., Prilusky, J., Abola, E. E., and Edelman, M. (1999) *Bioinformatics* 15, 327-32.
54. Polekhina, G., House, C. M., Traficante, N., Mackay, J. P., Relaix, F., Sassoon, D. A., Parker, M. W., and Bowtell, D. D. (2002) *Nat Struct Biol* 9, 68-75.
55. Antonyuk, S. V., Eady, R. R., Strange, R. W., and Hasnain, S. S. (2003) *Acta Crystallogr D Biol Crystallogr* 59, 835-42.
56. Cowan-Jacob, S. W., Kaufmann, M., Anselmo, A. N., Stark, W., and Grutter, M. G. (2003) *Acta Crystallogr D Biol Crystallogr* 59, 2218-27.
57. Shen, Y. Q., Li, J., Song, S. Y., and Lin, Z. J. (2000) *J Struct Biol* 130, 1-9.
58. Yun, M., Park, C. G., Kim, J. Y., and Park, H. W. (2000) *Biochemistry* 39, 10702-10.
59. Didierjean, C., Rahuel-Clermont, S., Vitoux, B., Dideberg, O., Branlant, G., and Aubry, A. (1997) *Journal of Molecular Biology* 268, 739-59.
60. Skarzynski, T., and Wonacott, A. J. (1988) *Journal of Molecular Biology* 203, 1097-118.
61. Rossmann, M. G., Moras, D., and Olsen, K. W. (1974) *Nature* 250, 194-9.
62. Krissinel, E., and Henrick, K. (2004) *Acta Crystallogr D Biol Crystallogr* 60, 2256-68.
63. Duee, E., Olivier-Deyris, L., Fanchon, E., Corbier, C., Branlant, G., and Dideberg, O. (1996) *Journal of Molecular Biology* 257, 814-38.
64. Roitel, O., Vachette, P., Azza, S., and Branlant, G. (2003) *J Mol Biol* 326, 1513-22.
65. Talfournier, F., Colloc'h, N., Mornon, J. P., and Branlant, G. (1999) *Eur J Biochem* 265, 93-104.
66. Shen, Y. Q., Song, S. Y., and Lin, Z. J. (2002) *Acta Crystallogr D Biol Crystallogr* 58, 1287-97.
67. Van Calenbergh, S., Verlinde, C. L., Soenens, J., De Bruyn, A., Callens, M., Blaton, N. M., Peeters, O. M., Rozenski, J., Hol, W. G., and Herdewijn, P. (1995) *J Med Chem* 38, 3838-49.
68. Bottoms, C. A., Smith, P. E., and Tanner, J. J. (2002) *Protein Sci* 11, 2125-37.
69. Bottoms, C. A., White, T. A., and Tanner, J. J. (2005) *Proteins*.
70. Binkowski, T. A., Naghibzadeh, S., and Liang, J. (2003) *Nucleic Acids Res* 31, 3352-5.
71. Brown, V. M., Krynetski, E. Y., Krynetskaia, N. F., Grieger, D., Mukatira, S. T., Murti, K. G., Slaughter, C. A., Park, H. W., and Evans, W. E. (2004) *J Biol Chem* 279, 5984-92.
72. Mercer, W. D., I., W. S., and Watson, H. C. (1976) *Journal of Molecular Biology* 104, 277-283.
73. Leslie, A. G., and Wonacott, A. J. (1984) *J Mol Biol* 178, 743-72.

## CHAPTER 2

# HIGH RESOLUTION STRUCTURES OF *THERMUS AQUATICUS* GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE: CONFORMATIONAL CHANGES INDUCED BY PHOSPHATE BINDING

Abbreviations: GAPDH, Glyceraldehyde-3-phosphate dehydrogenase; BPGA, 1,3-bisphosphoglycerate; GAP, glyceraldehyde-3-phosphate;  $\text{NAD}^+$ , nicotinamide adenine dinucleotide;  $\text{P}_i$ , inorganic phosphate;  $\text{P}_s$ , substrate phosphate;  $\text{P}_i$ -binding loop, phosphate-binding loop; BsGAPDH, *Bacillus stearothermophilus* GAPDH; LmGAPDH, *Leishmania mexicana*; EcGAPDH, *E. coli* GAPDH; TaGAPDH, *Thermus aquaticus* GAPDH; TcGAPDH, *Trypanosoma cruzi* GAPDH; TbGAPDH, *Trypanosoma brucei* GAPDH; RMSD, root mean square difference; PDB, Protein Data Bank; TBD, to be determined

## INTRODUCTION

Glyceraldehyde-3-phosphate dehydrogenase (1) is a glycolytic enzyme that catalyzes the formation of 1,3-bisphosphoglycerate (BPGA) from glyceraldehyde-3-phosphate (GAP). This reversible reaction requires the presence of nicotinamide adenine dinucleotide ( $\text{NAD}^+$ ) and inorganic phosphate ( $\text{Pi}$ ). GAPDH is a tetrameric protein made up of identical 37 kDa subunits. The enzyme appears to be ubiquitous in nature, cellularly abundant, and is relatively easy to purify, which has led to its widespread use as a model enzyme for the study/comparison of gene expression.

The generally accepted mechanism of GAPDH activity (Figure 2.1) is the initial formation of a covalent hemithioacetal intermediate between GAP and an essential cysteine. The hemithioacetal is then oxidized to a thioester, with simultaneous reduction of  $\text{NAD}^+$  to NADH. BPGA is then released after phosphorolytic attack on the thioester (2).

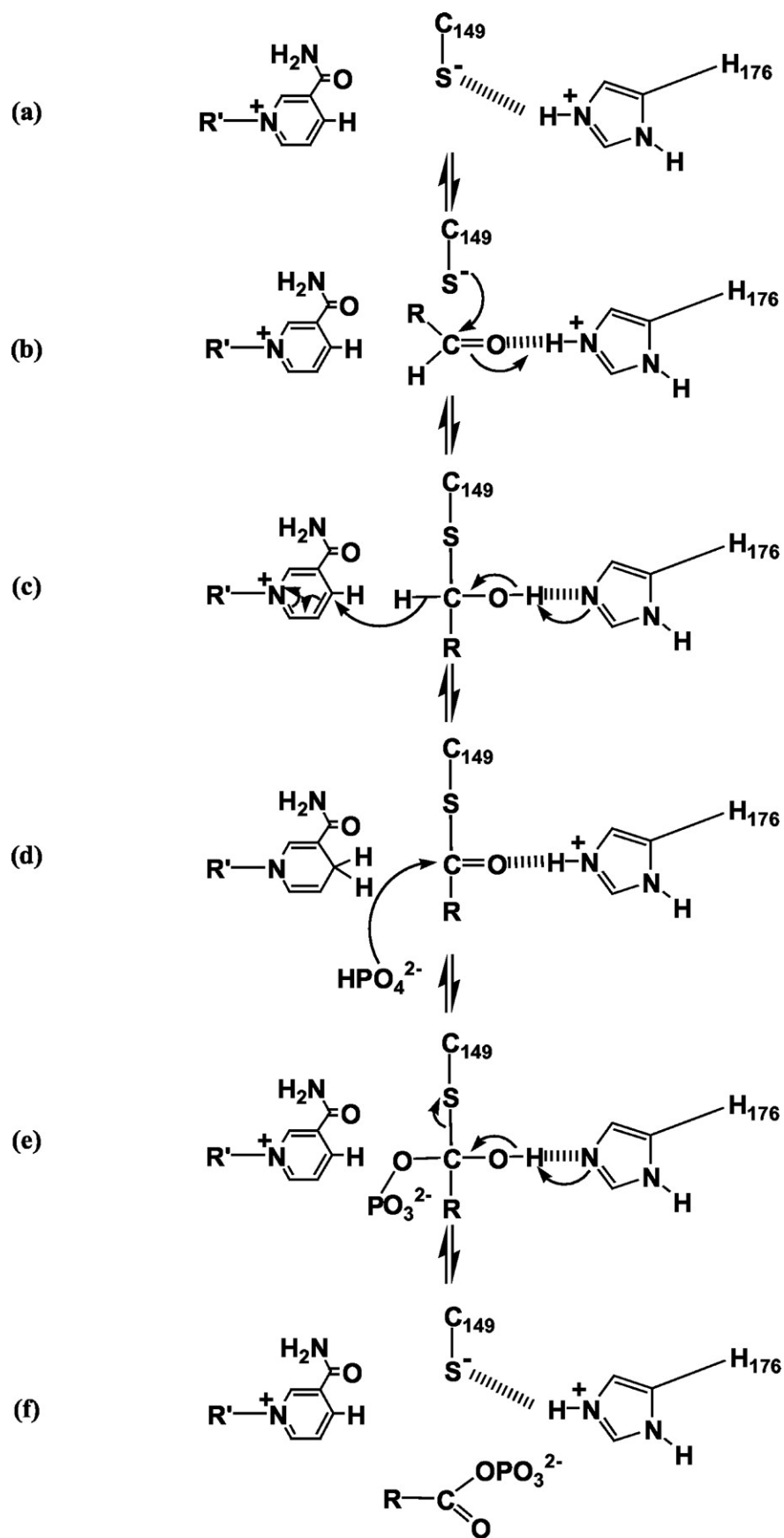
Although intensely studied for decades, some aspects of catalysis remain unresolved. For example, questions about the locations of the substrate and inorganic phosphate groups during catalysis have not been definitively answered. And the role of dynamics of the 220s phosphate-binding loop ( $\text{P}_i$ -binding loop) in promoting catalysis is still an open question.

Two models for phosphate binding have been proposed. The two models differ in their placement of the substrate phosphate during the reaction. The first model asserts that two sites exist where the GAP phosphoryl (substrate phosphate) and inorganic phosphate could reside (3). The site for the substrate phosphate is proposed to be near the active site Cys149 and His176, where it could take place in the formation of a hemithioacetal

intermediate. This site is known as the classical  $P_s$ -site. The second site is farther away from the catalytic Cys149 but close to the nicotinamide ring, making it a good choice for the inorganic phosphate-binding site ( $P_i$ -site). This model was proposed based on crystals grown in high concentrations of sulfate, and thus sulfate ions were modeled into the two sites (4) .



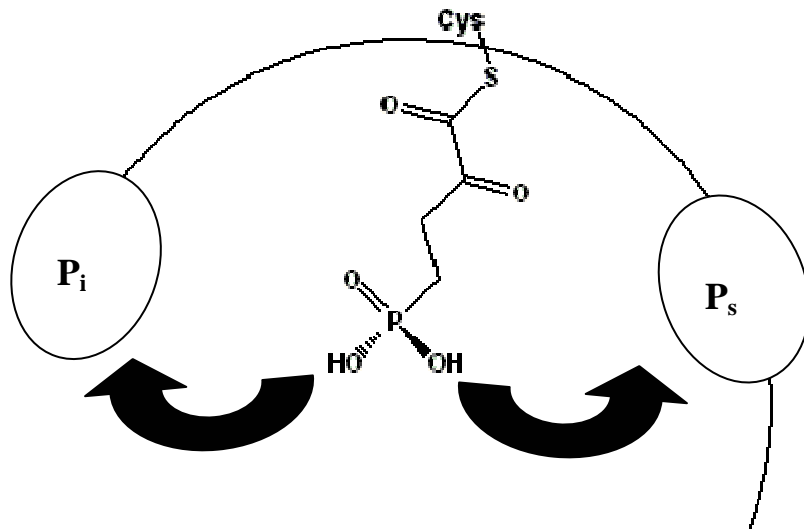
**Figure 2.1.** *The postulated catalytic mechanism for GAPDH.* This schematic and description are from Didierjean *et al.* 2003 (5). In the acylation step, interaction of Cys149 with His176 (*a*) decreases the  $pK_{app}$  of Cys149, thus facilitating the thiolate attack toward the C1 of GAP. The role of His176 is also to stabilize the binding of the substrate in the Michaelis complex GAPDH-NAD-D-GAP (*b*), the thiohemiacetal intermediate (*c*), and the thioacylenzyme intermediate (*d*). His176 also plays a role as a base catalyst, facilitating the hydride transfer from the thiohemiacetal to the nicotinamide of NAD (*c*). In the phosphorylation step, the binding of inorganic phosphate to the thioacylenzyme is followed by its nucleophilic attack on the thioacyl intermediate (*d*), which leads *via* an  $sp^3$ -phosphorylated intermediate (*e*) to the formation and release of BPGA (*f*). His176 is postulated to stabilize the tetrahedral intermediate (*e*) and to facilitate, as an acid (*d*) or base (*e*) catalyst, BPGA formation. The cofactor exchange step (*d* and *e*), which consists of NADH release prior to NAD and inorganic phosphate binding, remains controversial.  $R'$  represents the adenine-ribose-phosphate-phosphate-ribose part of the cofactor, NAD.  $R$  represents the  $CH(OH)COPO_3$  part of the substrate, D-G3P.



An alternative model maintains that the substrate phosphate and inorganic phosphate bind to the same site at different stages of the reaction (1). This so-called “flip-flop model”, which was based in part on the 1.8 Å structure (1GD1) of *Bacillus stearothermophilus* GAPDH (BsGAPDH) (6), proposes that the C3 phosphate of GAP occupies both phosphate sites at different times (Figure 2.2). The P<sub>i</sub>-site is occupied first by the C3 phosphate of GAP before the formation of the acyl-enzyme intermediate, and then, upon coenzyme exchange, a conformational change would lead to the C3 phosphate of GAP flipping to the P<sub>s</sub> site. This model still allows for the P<sub>i</sub>-site to be occupied by inorganic phosphate during the phosphorylation step. The flip-flop model was likewise based on structures grown under high sulfate conditions.

The issue of phosphate binding was complicated by the discovery of a new conformation of a strand-loop-helix region, which results in a new P<sub>i</sub>-site in some GAPDH structures. Kim et al. determined a structure of *Leishmania mexicana* GAPDH (1GYF, LmGAPDH) that displayed a new conformation of the P<sub>i</sub>-binding loop, making relocation of the classical P<sub>i</sub>-site necessary (4). The authors suggested that this new P<sub>i</sub>-site was the result of the P<sub>i</sub>-binding loop constricting the active site, thereby blocking binding to the classical P<sub>i</sub>-site. The overall effect is that the residues that make up the new P<sub>i</sub>-site are now closer to the catalytic Cys149. The crystals were grown in the presence of 200 mM phosphate, which was thought to be the cause of this loop movement.

**Figure 2.2.** *Cartoon representation of the flip-flop model.* Glyceraldehyde-3-phosphate (GAP) is covalently attached to the catalytic cysteine (thiohemiacetal intermediate) with the arrows marking the flipping of the C3 phosphate from the inorganic site ( $P_i$ ) to the substrate site ( $P_s$ ). This cartoon is from Castilho *et al.* 2003 (7).



A second LmGAPDH structure (1A7K) was determined from crystals grown in 100 mM phosphate and it displayed the same loop movement (8). Other structures that also display the new P<sub>i</sub>-site include GAPDHs from *Thermatoga maritima* (1HDG), (9) and Chinese lobster (1CRW)(10). These structures were determined from crystals grown in ammonium sulfate, rather than phosphate.

The 2.5 Å resolution structure of a hemiacetal intermediate of *E. coli* GAPDH (EcGAPDH, 1DC4) provides additional evidence supporting the flip-flop model (11). This structure shows the C3 phosphate of D-GAP binding to the new P<sub>i</sub>-site during formation of the hemiacetal intermediate, and the model does appear to show that GAP can occupy both sites. However, the crystals were grown in the absence of phosphate, sulfate, or NAD<sup>+</sup>, leading the authors to question whether they were actually observing the true binding site of GAP.

The structures of two BsGAPDH ternary complexes (GAPDH/NAD<sup>+</sup>/GAP) showed the substrate phosphate bound to the P<sub>s</sub>-site (1NQA and 1NQO). Interestingly, a rotation around the C1-C2 bond would move the phosphate to the “new P<sub>i</sub>-site” (5), which supports an updated flip-flop model in which the new P<sub>i</sub>-site holds the inorganic phosphate.

To study the role of conformational changes in the P<sub>i</sub>-binding loop induced by phosphate binding, we report three new crystal structures of *Thermus aquaticus* GAPDH (TaGAPDH) corresponding to phosphate concentrations of 0, 50, and 100 mM. The structure of the 50 mM crystal form is particularly relevant because the physiological concentration of phosphate found in cells is 50 mM (12). The 1.65 Å phosphate-free structure reveals a single conformation for the main chain of the P<sub>i</sub>-binding loop but

alternative conformations for P<sub>s</sub>-site side-chains. In the 50 mM and 100 mM structures, the binding of phosphate resulted in multiple conformations of the P<sub>i</sub>-binding loop and the P<sub>i</sub> phosphate. Our structures show that the dynamic P<sub>i</sub>-binding loop plays a more important role than previously thought in delivering GAPDH substrates to their appropriate binding sites for catalysis. Our structures and the structures of others (5,7) clearly show that in the presence of high and low phosphate concentrations, the fully formed P<sub>s</sub>-site (NAD<sup>+</sup> bound) represents the first choice for phosphoryl binding.

## EXPERIMENTAL PROCEDURES

### *Purification and Crystallization*

Recombinant TaGAPDH protein was expressed and purified as reported by Tanner et al., 1994 (12). Briefly the *Thermus aquaticus* GAPDH gene was isolated from the YT1 genomic library and cloned into the *E. coli* W3CG strain (GAPDH null mutant) for overnight expression in M9 minimal media. Purification involved heating the cell lysate to 363K for 30 minutes to remove thermolabile proteins, followed by chromatofocusing chromatography with polybuffer exchanger 94 (Pharmacia). All fractions containing significant amounts of protein were analyzed by denaturing SDS-PAGE gels, and those fractions containing TaGAPDH were combined and dialyzed overnight into 4 M ammonium sulfate. The precipitated protein was resuspended with the pre-crystallization buffer (2.5 mM Tris, 100 mM NaCl, 1mM  $\beta$ ME, 1mM EDTA, pH 7.7), and dialyzed overnight into the same buffer. The protein was concentrated to 25 mg ml<sup>-1</sup> using centrifugal filters & tubes (Millipore) after which NAD<sup>+</sup> was added to a concentration of 2 mM. Protein concentration was estimated using Coomassie Plus<sup>®</sup> protein assay (Pierce).

All crystallization experiments were performed at 295 K using the sitting-drop method of vapor diffusion. Crystallization trials were initiated from conditions used previously to determine the structure of TaGAPDH at 2.5 Å resolution (13). The best crystals were grown using a reservoir solution of 23% PEG 3000, 0.1 M Hepes buffer at pH 8.3, 15% isopropanol and 5% glycerol.

Two crystal forms of TaGAPDH with bound phosphate were obtained. The 50 mM phosphate form was obtained by soaking crystals in 50 mM sodium phosphate



buffer. The 100 mM phosphate form was obtained by co-crystallization using a reservoir of 21% PEG 3000, 0.1 M Hepes buffer at pH 7.6, 10% isopropanol, 100 mM phosphate buffer and 5% glycerol. All crystals were brick shaped, and cryoprotection was achieved by soaking crystals in the harvest buffer supplemented with 10% glycerol.

#### *Data Collection, Model Building, and Refinement*

X-ray diffraction data sets for the phosphate-free and 50 mM phosphate forms were obtained at beamline X8C of the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory (BNL) using an ADSC Quantum 4R detector. Diffraction data for the 100 mM phosphate form were obtained with an in-house Rigaku Cu rotating anode system equipped with an R-Axis IV detector, Osmic optics, and X-stream low-temperature system. The *HKL* suite of programs (14) was used to process the data. The high-resolution limit for processing each data set was chosen such that more than half of the reflections in the highest resolution shell of data have  $I/\sigma(I) > 2$  (15). Refer to Table 2.1 for data collection statistics.

All three crystal forms belong to space group  $P2_12_12_1$  with two tetramers per asymmetric unit. There are four  $\text{NAD}^+$  molecules bound to each tetramer of all three structures. The unit cell dimensions are in the ranges  $a = 141\text{-}142 \text{ \AA}$ ,  $b = 147\text{-}148 \text{ \AA}$ ,  $c = 148 \text{ \AA}$ . Thus, neither soaking crystals with phosphate nor co-crystallization with phosphate caused significant unit cell changes.

Structure refinement was done with CNS (16) and REFMAC5 (17) via the program CCP4i (18). Model building was done with O (19). Refinement of the 1.65  $\text{\AA}$  phosphate-free structure was initiated from the coordinates of the previously determined

2.5 Å resolution TaGAPDH structure (1CER), (20). The refined 1.65 Å phosphate-free structure (without solvent) served as the starting model for refinement of the two phosphate-bound structures. Electron density maps clearly indicated that the active site Cys149 was oxidized in all three structures. The oxidized Cys149 was modeled as cysteine-s-dioxide (sulfinic acid) using the CSW entry from the HIC-Up database (21). See Table 2.1 for refinement statistics.

The programs CNS, O, and Pymol (22) were used for structural analysis. Superimpositions of structures and RMSD calculations were done using the websites of Mass (23), CE (24), and COMPARER (25).

**Table 2.1.** Data collection and refinement statistics<sup>a</sup>

[Phosphate] (mM)	0	50	100
PDB accession code	2G82	TBD	TBD
Wavelength (Å)	0.97900	0.97900	1.542
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Unit cell dimensions (Å)	<i>a</i> = 141.5 <i>b</i> = 147.4 <i>c</i> = 147.6	<i>a</i> = 141.1 <i>b</i> = 147.4 <i>c</i> = 147.6	<i>a</i> = 142.3 <i>b</i> = 147.9 <i>c</i> = 148.2
Resolution (Å)	99-1.65	99-1.85	100 - 2.23
Total observations	2 736 709	1 810 736	524 069
Unique reflections	390 958	259 119	144 853
Redundancy	7.0	7.0	3.6
Completeness (%)	98 (85)	99 (98)	95(90)
Average I/σ(I)	18.3 (3.9)	23.9 (3.3)	16.8 (3.0)
R <sub>merge</sub>	0.098 (0.397)	0.080 (0.463)	0.057 (0.322)
No. of protein atoms	22023	22460	21185
No. of NAD <sup>+</sup> molecules	8	8	8
No. of water molecules	1586	1581	450
R <sub>cryst</sub>	15.8	15.9	18.8
R <sub>free</sub> <sup>b</sup>	17.2	18.4	22.6
RMSD bond lengths (Å) <sup>c</sup>	0.011	0.010	0.018
RMSD bond angles (deg.)	1.24	1.28	1.65
Ramachandran plot <sup>d</sup>			
Favored (no. residues)	2528	2532	2522
Allowed (no. residues)	72	66	79
Outlier (no. residues)	0	1	1
Average <i>B</i> -factors (Å <sup>2</sup> )			
Protein	13.8	12.8	23.9
NAD <sup>+</sup>	10.9	12.6	26.8
Water	29.4	29.1	31.3

<sup>a</sup>Values for the outer resolution shell of data are given in parenthesis.<sup>b</sup>5% R<sub>free</sub> test sets.

<sup>c</sup>Compared to the Engh and Huber force field.

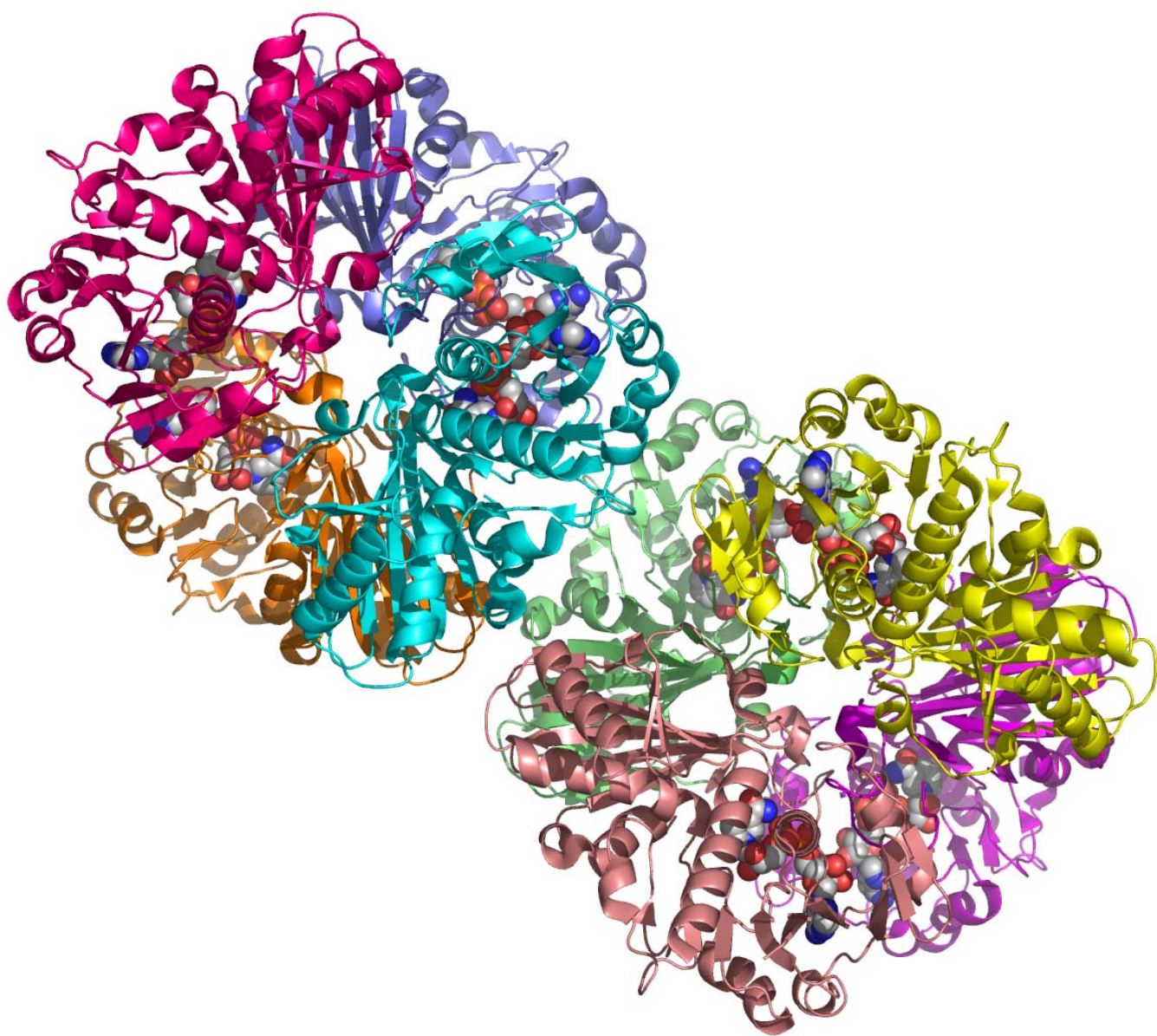
<sup>d</sup>The Ramachandran plot was generated with RAMPAGE.

## RESULTS AND DISCUSSION

### *Quality of all Models*

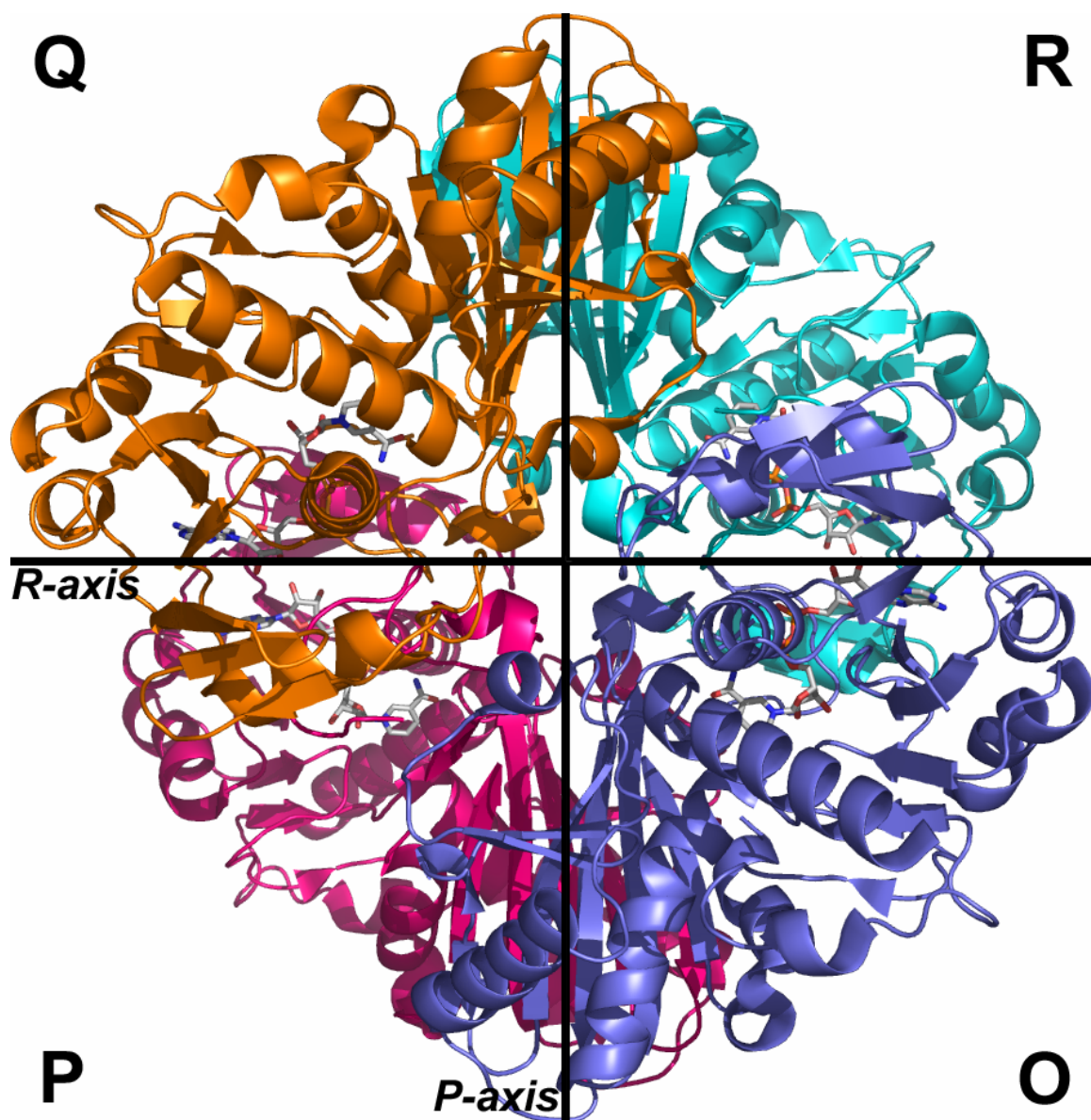
The asymmetric unit consists of two tetramers, OPQR and ABCD, and eight  $\text{NAD}^+$  molecules (Figure 2.3). The tetramers display the expected approximate 222 symmetry described by 3 mutually perpendicular 2-fold axes (Figure 2.4). Each subunit consists of an  $\text{NAD}^+$ -binding domain (residues 1-147, 314-331) exhibiting the classical Rossmann dinucleotide-binding fold (26) and a catalytic domain (residues 148-313) consisting of the dihydrodipicolinate reductase domain 2 fold (27). The subunits were labeled as in the previously solved 1CER structure. The structures with resolutions 2.23 - 1.65 Å have  $R_{\text{factor}} = 0.188 - 0.158$  and  $R_{\text{free}} = 0.226 - 0.172$  for all reflections (Table 2.1). The 1.65 Å phosphate-free structure is the highest resolution GAPDH structure deposited in the PDB (28) to date, and contains 1586 water molecules, which represents a significant improvement from the 1CER structure that was built without water. The 1.85 Å (50 mM phosphate) structure shares many of these water molecules but also has unique waters. In the 2.23 Å (100 mM phosphate) model, only the strongest waters were built due to its lower resolution. The phosphate-free and the 50 mM structures each include one polyethylene glycol (PEG) fragment located, between Pro C52 and the symmetry related Pro Q52. The PEG fragment interacts with a  $\text{Na}^+$  ion, which in turn interacts with 3 water molecules (Figure 2.5). As a result, the  $\text{Na}^+$  ion has octahedral coordination. The 100 mM and 50 mM structures both have 23 and 24 phosphate molecules ( $\text{PO}_4$ ), respectively, and their relative positions in the active site will be discussed below.

**Figure 2.3.** *Cartoon representation of the two tetramers found in the asymmetric unit of the 1.65 Å resolution TAGAPDH structure. Each subunit contains one NAD<sup>+</sup> molecule (grey). Subunits are colored as follows: O, blue; P, red; Q, orange; R, cyan; A, yellow; B, purple; C, salmon; D, green.*

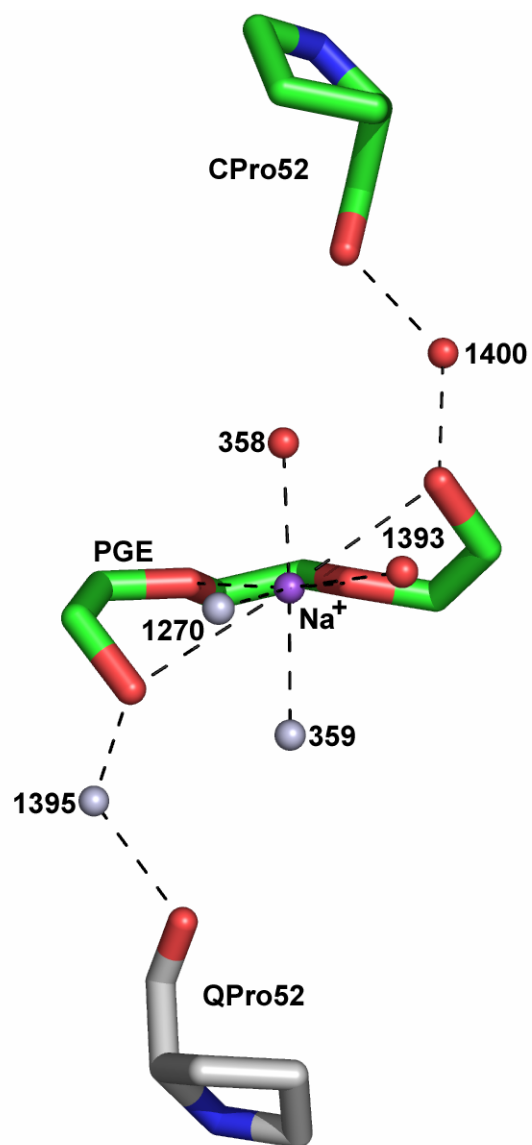


**Figure 2.4.** *Ribbon drawing of a TaGAPDH tetramer viewed down the Q-axis.* Subunits are colored as follows: O, blue; P, red; Q, orange; R, cyan. Lines indicate locations of the P and R molecular 2-fold axes. The NAD<sup>+</sup> cofactors are drawn in stick mode.





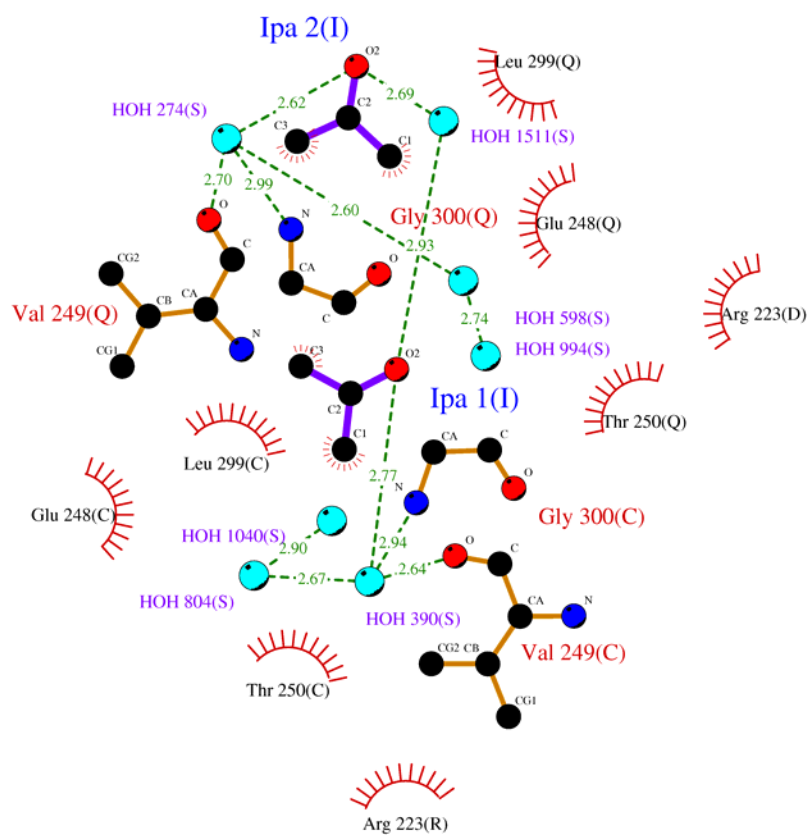
**Figure 2.5.** *A view of the interactions between PEG and a  $\text{Na}^+$  ion with the surrounding solvent. The C subunit Pro 52 is colored green with  $\text{H}_2\text{O}$  in red. Symmetry related Q Pro52 is gray as are the symmetry related waters. The dashed lines represent interaction distances  $\leq 3.0 \text{ \AA}$ .*



The phosphate-free structure also includes two isopropyl alcohol molecules, IPA1 and IPA2, (Figure 2.6). Both are within 2.9 Å of H<sub>2</sub>O 1511 which is positioned between the two tetramers near subunits C, D, Q, and R. Only IPA1 could be modeled in the 50 mM structure where there is no water equivalent to H<sub>2</sub>O 1511, however, IPA1 in the 50 mM structure does form one hydrogen bond to H<sub>2</sub>O 390 as in the 1.65 Å structure. There are also three glycerol molecules (GOL) in the 1.65 Å structure and all three are found within hydrogen bonding distance of Glu276 and Arg193 from neighboring subunits and Thr206 of the P<sub>i</sub>-binding loop (Figure 2.7). Density reminiscent of GOL was found in the same position in six of the 8 subunits but only these three could be modeled with confidence in the 1.65 Å structure. The 1.85 Å structure shares only one of the three GOLs. The stereochemistry for each model meets or exceeds all the main chain and side chain tests of the PROCHECK package (29). Only two residues, Ala210 subunit D and Ile80 subunit B in the 50 mM and 100 mM structures, respectively, are found in the outlier region of the Ramachandran plot (30). Ala210 is part of the P<sub>i</sub>-binding loop (8) and is modeled in dual conformations in subunit D and other subunits. Residue Ile80 in subunit B is found in a region with poor density but the residue adopts similar conformations in all subunits in the structures.

The electron density maps were of sufficient quality to allow for clear modeling of protein, NAD<sup>+</sup>, and solvent in all but a few regions. The three structures together have some 53 missing side chain residues, due to poor density. The average *B*-factor for all protein atoms is 13.8 Å<sup>2</sup>, 12.8 Å<sup>2</sup>, and 23.9 Å<sup>2</sup> for the phosphate-free, 50 mM, and the 100 mM structures respectively.

**Figure 2.6.** *Ligplot of the two isopropanol molecules (IPA) found in the 1.65 Å structure.*



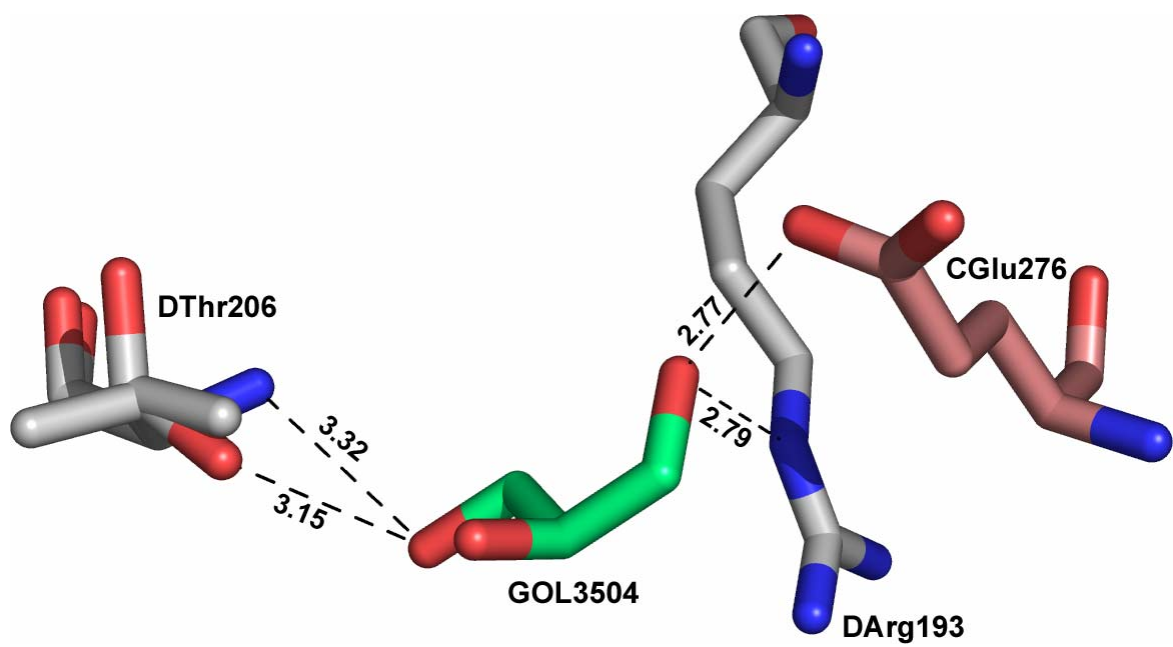
## Key

- ● Ligand bond
- ● Non-ligand bond
- — ● Hydrogen bond and its length
- His 53 Non-ligand residues involved in hydrophobic contact(s)
- Corresponding atoms involved in hydrophobic contact(s)

**Figure 2.7.** *View of one of the three glycerol molecules (GOL) of the 1.65 Å structure.*

Residues from subunits D and C are colored grey and salmon, respectively. The glycerol molecule is shown in green. Residue D Thr206 has been modeled in two conformations.

Distances are listed in Angstroms.



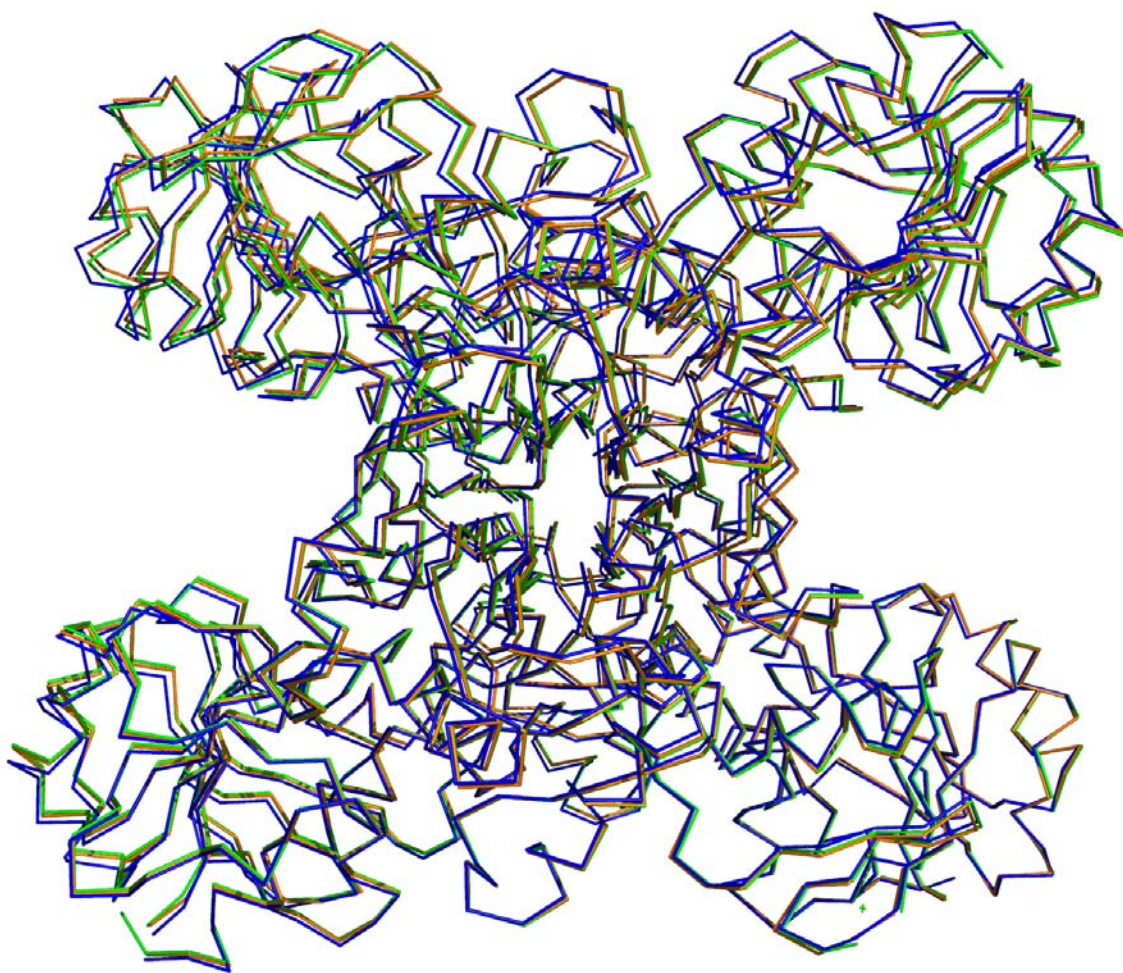


Superposition of the TaGAPDH tetramers (Figure 2.8) in the three structures returns RMSD values of 0.20–0.29 Å for the C<sub>α</sub> atoms (Table 2.2A). The individual subunits from each structure were also superimposed and display RMSD values of 0.12–0.35 Å, 0.13–0.31 Å, and 0.20–0.33 Å for the phosphate-free, 50 mM, and 100 mM respectively (Table 2.2B). Taken together, the superpositions show that the TaGAPDH subunits are structurally very similar to each other. The closest solved structural homologue of TaGAPDH is *Thermus thermophilus* GAPDH (PDB code 1VC2), which has 90% sequence identity to TaGAPDH and 96% structural similarity as indicated by SSM (31). The RMSD values between the subunits from the TaGAPDH structures and the A subunit of 1VC2 are 0.39–0.5 Å.

*Structural comparison of the 1.65 Å structure with 1CER.*

The 1.65 Å structure represents a significant improvement in resolution over the previously deposited 2.5 Å 1CER. The two structures share missing side chains Met1, Glu25, Lys75, and Lys78. The missing side chains Lys104 and Arg330 from 1CER are present in the 1.65 Å structure in at least three of the eight subunits. Val333 was successfully built in several subunits of the 50 mM and 100 mM structures. There are 4 residues (Asp181, Arg194, Thr206, and Arg230) that have been modeled into two conformations in each of the 1.65 Å subunits. In the previous structure, it was noted that there was a bad contact between Asp138 in subunit P and the symmetry related Glu253 in subunit A, which was attributed to the use of strict NCS symmetry restraints during refinement. This bad contact is not present in any of the three new TaGAPDH structures, most likely due to our removal of NCS symmetry restraints during the late stages of refinement.

**Figure 2.8.** *Ribbon drawing of the 3 TaGAPDH OPQR tetramers superimposed and viewed down the P-axis.* Subunits are colored as follows: phosphate-free, orange; 50 mM, green; 100 mM, blue.



**Table 2.2.**  $C_a$  Superposition statistics.*(a) Superposition of tetramers*

<b>Tetramer</b>	Phosphate-free (OPQR)	50 mM (OPQR)	100 mM (OPQR)	Phosphate-free (ABCD)	50 mM (ABCD)	100 mM (ABCD)
Phosphate-free (OPQR)		0.20	0.26	0.31	0.29	0.44
50 mM (OPQR)			0.24	0.39	0.27	0.39
Phosphate-free (ABCD)					0.20	0.29
50 mM (ABCD)						0.24

*(b) Superposition of subunits*

<b>Phosphate-free</b>	O	P	Q	R	A	B	C	D
O		0.33	0.17	0.16	0.13	0.17	0.20	0.17
P			0.27	0.28	0.32	0.29	0.35	0.31
Q				0.16	0.21	0.19	0.22	0.20
R					0.13	0.12	0.20	0.12
A						0.14	0.19	0.14
B							0.19	0.11
C								0.16

<b>50 mM</b>	O	P	Q	R	A	B	C	D
O		0.30	0.17	0.16	0.13	0.17	0.20	0.17
P			0.28	0.26	0.29	0.28	0.31	0.28
Q				0.17	0.23	0.21	0.20	0.19
R					0.16	0.13	0.18	0.11
A						0.19	0.19	0.17
B							0.18	0.13
C								0.15

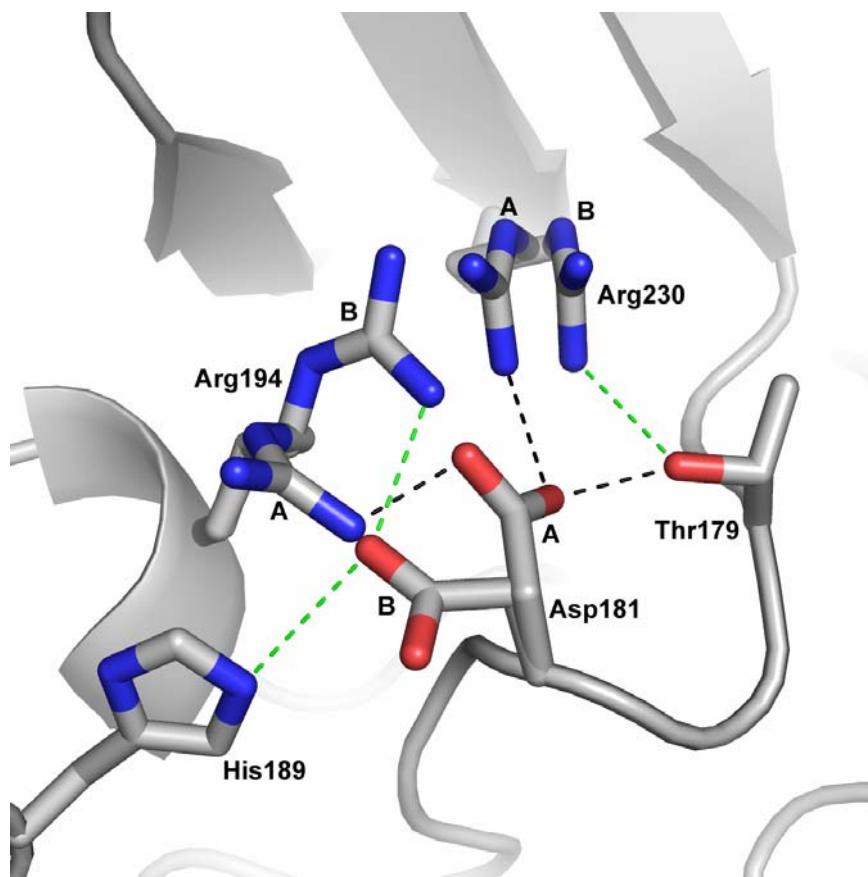
<b>100 mM</b>	O	P	Q	R	A	B	C	D
O		0.31	0.29	0.21	0.20	0.26	0.25	0.24
P			0.26	0.26	0.32	0.33	0.30	0.28
Q				0.23	0.31	0.30	0.22	0.24
R					0.22	0.24	0.21	0.19
A						0.28	0.24	0.20
B							0.28	0.23
C								0.23

*Substrate Phosphate binding site and movement of Asp181, Arg194, and Arg230.*

Typically GAPDH crystals grown in the presence of sulfate or phosphate have two sulfate/phosphate ions per subunit occupying the aforementioned inorganic phosphate ( $P_i$ ) site and substrate phosphate ( $P_s$ ) site. The previously solved TaGAPDH structure (1CER) was crystallized in PEG and isopropanol and did not display any movement of the  $P_i$ -binding loop but did show small rearrangements of Asp181, Arg194, Arg230 residues near the  $P_s$ -site. In the phosphate-free structure, these three residues are all modeled in dual conformations (Figure 2.9). Conformation A is very similar to that found in the 1CER structure. In conformation A (Table 2.3), Asp181 forms hydrogen bonds with Arg194 (2.2–3.0 Å), Arg230 (2.85–3.2 Å), and Thr179 (2.0–2.4 Å). The major difference from 1CER is that interaction distances between Arg230 and Thr179 have increased to 3.5–3.7 Å.

In conformation B (Table 2.3), Asp181 is torsioned  $\sim 5.8$  Å away from the OG1 of Thr179 and now Asp181 is capable of forming a weak hydrogen bond (3.3–3.5 Å) with His189 in half of the subunits. Asp181 is still capable of hydrogen bonding (2.6–3.0 Å) to Arg194 but not Arg230. Also, the interaction distance between Arg230 and Thr179 has now decreased to 2.7–2.9 Å. Asp 181, Arg194 and Arg230 adopt conformation B in the 50 mM and 100mM phosphate-bound structures. The only difference from conformation B of the phosphate-free structure is that the hydrogen-bonding distance between Asp181 and His189 is shorter (3.0–3.1 Å) in the phosphate-bound structures.

Figure 2.9. *The two conformations modeled in the phosphate-free  $P_s$ -site.* The black dashed lines indicate electrostatic interactions found in conformation A between with Arg230, Arg194, Thr1791 and His189. The green dashed lines indicate electrostatic interactions found in conformation B.



**Table 2.3** *Hydrogen bonding interactions found in the phosphate-free P<sub>s</sub>-site residues.*

Conformation A

<b>Interactions</b>	<b>O</b>	<b>P</b>	<b>Q</b>	<b>R</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
Thr179-Asp181	2.01	2.26	2.15	2.13	2.24	2.36	2.09	2.01
Thr179-Arg230	3.69	3.62	3.60	3.61	3.47	3.67	3.72	3.60
Asp181-Arg194	3.06	2.19	3.03	2.73	2.51	2.99	2.95	2.83
Asp181-Arg230	2.70	2.85	2.87	2.92	2.93	2.93	3.11	3.09

Conformation B

<b>Interactions</b>	<b>O</b>	<b>P</b>	<b>Q</b>	<b>R</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
Thr179-Asp181	5.79	5.69	5.71	5.78	5.72	5.71	5.76	5.89
Thr179-Arg230	2.79	2.94	2.85	2.76	2.73	2.82	2.76	2.77
Asp181-Arg194	2.74	2.82	2.70	2.70	2.96	2.75	2.72	2.59
Asp181-Arg230	6.06	5.93	6.01	5.99	5.97	5.92	5.97	6.08
Asp181-His189	3.29	3.29	3.43	3.46	3.59	3.52	3.45	3.53



In the phosphate containing LmGAPDH structures, Arg249, Thr199, and Thr197 are structurally equivalent to residues Arg230, Asp181 and Thr179. In fact, residues Thr179 and Arg230 are found in all deposited eukaryote and bacterial GAPDH structures (5). In the phosphate-containing LmGAPDH structures, only conformation B is observed.

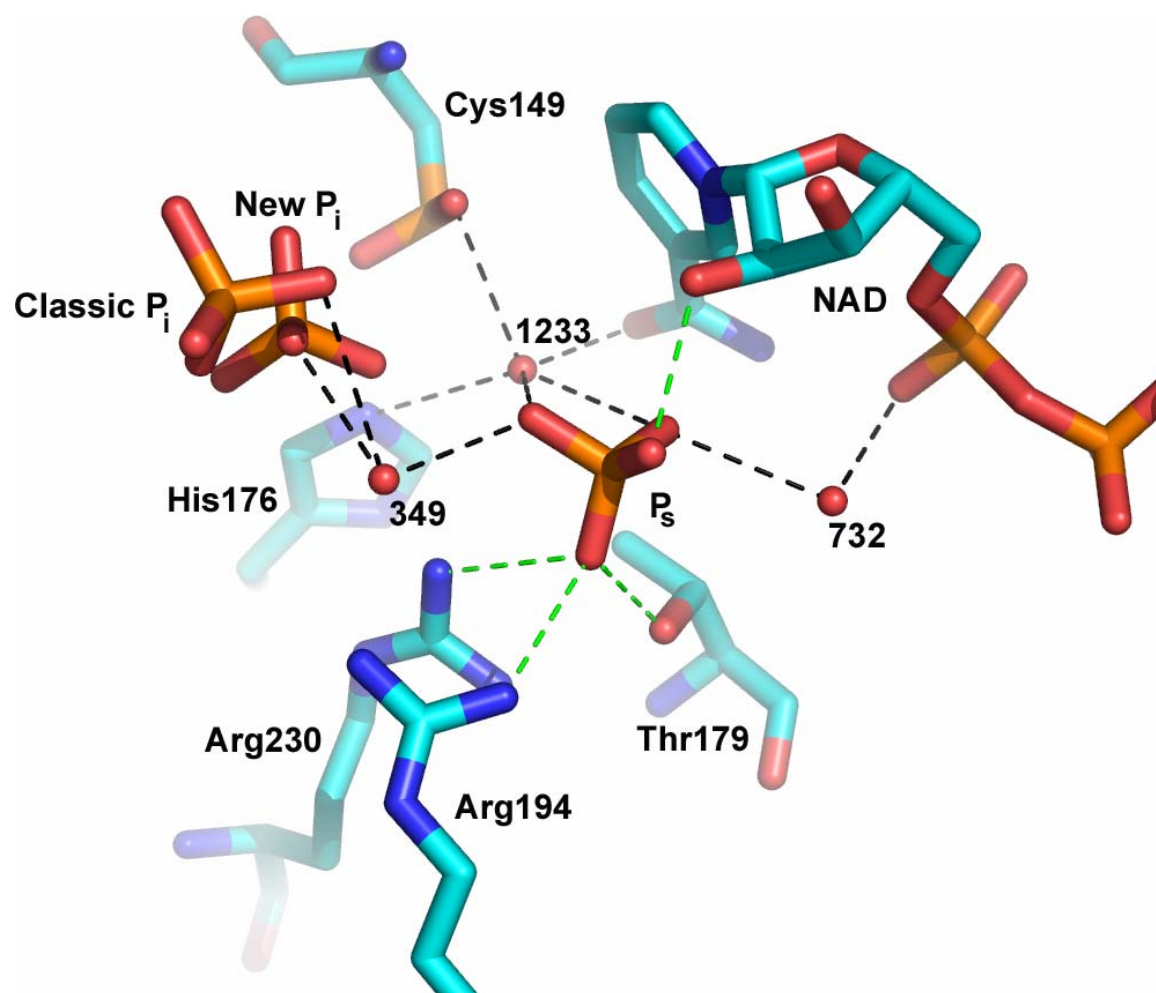
Conformation A and B are thought to represent the phosphate-free and phosphate-bound states, respectively. This raises the question of why both conformations are seen in our high-resolution phosphate-free structure. It is possible that we are seeing two conformations of these residues in the zero phosphate structure due to partial occupancy of the  $P_s$ -site by sulfate. There is unmodeled density in the  $P_s$ -site of some of the subunits due perhaps from the failure to completely remove sulfate after an early ammonium sulfate precipitation step, despite extensive dialysis afterwards. In the structures 1NPT and 1NQ5, they had a similar result when they tried to grow crystals of BsGAPDH with an empty  $P_s$ -site (5).

In the 50 mM and 100 mM structures,  $P_s$  interacts (Table 2.4) with Thr179, Arg194, Arg230 and the O2' of the nicotinamide ribose (Figure 2.10). These interactions are found in other GAPDH structures with phosphate/sulfate bound in the  $P_s$ -site, such as the above mentioned 1NPT and 1NQ5 (5). In the 50 mM model, there are at least two water molecules interacting with the bound  $PO_4$  in each subunit (Table 2.4) and some water-mediated interactions with the  $P_i$ -site (Figure 2.10).

Comparison of our 50 mM and 100 mM phosphate structures with 1A7K, 1ML3, 1QXS (32), 1NQA, and 1NQO (ternary complexes (GAPDH/ $NAD^+$ /GAP)) reveals little new information about the well characterized  $P_s$ -site. In fact, the  $P_s$  sites of different

GAPDH phosphate/sulfate-bound structures superimpose almost perfectly. Didierjean and colleagues offer the explanation that the  $P_s$ -site has a higher positive charge than the  $P_i$ -site according to their unpublished GAPDH monomer electrostatic potential calculations (5). That is why the ternary BsGAPDH structures have GAP occupying the  $P_s$ -site only, rather than the  $P_i$  site as proposed in the flip-flop model. Also, in structure 1DC4 (11), with GAP covalently bound to Cys149 forming the hemiacetal intermediate, the GAP phosphoryl group binds the new  $P_i$ -site only because the empty  $NAD^+$  binding site forms a less attractive  $P_s$ -site.

**Figure 2.10.** *A view of the 50 mM Q subunit P<sub>s</sub>-site interactions.* All of the TaGAPDH phosphate subunits share the conserved interactions (green dashed lines) with Arg230, Arg194, Thr1791 and NAD<sup>+</sup>. The subunits in 50 mM phosphate structure have additional hydrogen bonds with solvent (black dashed lines) and water-mediated interactions with the P<sub>i</sub>-site phosphates, NAD<sup>+</sup>, and side chains. The dashed lines indicate hydrogen bonds  $\leq 3.3$  Å.



**Table 2.4** *Interactions between the substrate phosphate group and active site residues in the 50 and 100 mM phosphate structures.*

<b>50 MM</b>	<b>O</b>	<b>P</b>	<b>Q</b>	<b>R</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
Thr179	2.70	2.68	2.71	2.69	2.75	2.68	2.71	2.80
Arg194	2.95	2.93	3.00	2.91	2.82	2.86	2.93	2.96
Arg230	2.97	2.95	2.96	2.91	2.98	2.95	2.96	2.93
NAD-NO'2	2.93	2.67	2.80	2.90	2.97	3.07	2.74	2.82
H <sub>2</sub> O(#)	1460 791	1246 370	349 1233 732	666 499	1059 846	399 981	1091 1352 765	635 423
<b>100 mM</b>	<b>O</b>	<b>P</b>	<b>Q</b>	<b>R</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
Thr179	2.84	2.54	2.90	2.66	2.58	2.80	2.55	2.80
Arg194	2.83	3.10	2.95	3.17	3.35	2.95	3.28	2.95
Arg230	2.90	3.30	2.78	2.91	2.91	3.09	3.08	3.06
NAD-NO'2	3.48	2.59	2.82	2.99	3.10	2.81	2.80	2.84
H <sub>2</sub> O(#)	NA	99	156 183	218	299	319	374	435

Distances are in Å.

*Movement of the inorganic phosphate binding loop.*

A strand-loop-helix substructure is critical for binding phosphate ion in the P<sub>i</sub>-site in all GAPDHs. This region corresponds to residues 205-213 of TaGAPDH (Figures 2.11, 2.12). As described above, there are generally two conformations for the P<sub>i</sub>-site, classical and new, and the same terms can be used to describe the location of the P<sub>i</sub>-binding loop whose residues interact with phosphate helping orient it into the P<sub>i</sub>-site. In the majority of the NAD<sup>+</sup>-specific GAPDH structures, the P<sub>i</sub>-binding loop is found in the classical conformation, which places the highly conserved Gly208 ~10 Å from the catalytic cysteine. The new conformation of the P<sub>i</sub>-binding loop, now ~3 Å closer to the active site cysteine, pushes the P<sub>i</sub>-site ~3.0 Å closer to the catalytic cysteine (4,8). This new conformation, found in few GAPDH structures, does not always correlate with the phosphate or sulfate ion concentration of the crystals.

We thought it would be of interest to see if we could induce movement of the P<sub>i</sub>-binding loop by addition of phosphate buffer. We chose 50 mM phosphate because it represents the cellular phosphate concentration and 100 mM phosphate because it has been shown to induce loop movement in LmGAPDH (8).

Surprisingly, electron density maps for the phosphate-bound structures showed evidence for both the classic and new conformations of the P<sub>i</sub>-binding loop. The 50 mM and 100 mM structures presented here are the first structures to have two conformations of the P<sub>i</sub>-binding loop (residues 205-213) and phosphate modeled into both the classical and new P<sub>i</sub>-site in the same subunit. The electron-density maps (Figure 2.13) were of sufficient quality to build alternative conformations for residues in the P<sub>i</sub>-binding loop in all of the 16 subunits in the two structures. Met227 also adopts two conformations in the

phosphate-bound structures, and thus movement of the  $P_i$ -binding loop seems to induce movement of Met227.

**Figure 2.11.** *Sequence alignment of GAPDH structures that show movement of the  $P_i$ -binding loop.* TaGAPDH (185) has pairwise identities of 48% with 1A7K, 63% with 1HDG, 50% with 1CRW (Chinese lobster) and 49% with 1DC4. Secondary structure was assigned by DSSP from 2G82. Identical residues are denoted by white type on a black background. Black stars denote the flexible strand-loop-helix substructure that binds  $P_i$  (8). This alignment was generated by the program ESPript.



185.pdb:Q

185.pdb:Q ..MKVGVINGFGRIGRQVFRILHSR.G...V...EVALINDL.T.DNKTLAHLKVDSTIYHRFP.GEVAAY  
PDB:1a7k:A APIKVGINGFGRIGRMVFQALCDQGLIGTEI...DVVAVVDM.STNAEYFAYQMKHDFVHGRPKYT.VEA  
PDB:1hdg:O ..ARVAINGFGRIGRLVYRIIYER.K...NFDIEVVAINDL.T.DTKTLAHLKVDSTVHKKFP.GKVEY  
PDB:1crw:R ..SKIGINGFGRIGRLVLRRALEM.G...A...QVVAVNDPFI.ALEYMVYMPKYDSTHGMFK.GBVKA  
PDB:1dc4:A .TIKVGINGFGRIGRIVFRAAQKR.SD...I...EIVAINDL.L.DADYMAVMLKYDSTHGRFD.GTVEV

185.pdb:Q

185.pdb:Q DD.....QYLYVDGKATIRATAVK.DPKSEIPWAEAGVGVVIESTCGVFTDADKAKAHLEGGAKKVHITA  
PDB:1a7k:A VKSSPSVETADVLLVNGHRIKCVKAQRNPADLPWGKICVDYVIESTCGVFTDKLKAEGHFKGGAKKVVISA  
PDB:1hdg:O TE.....NSLIVDGEIKVFAEP.DPSKLPWKDLCVDFVIESTCGVFRNREKAEHHLQAGAKKVHITA  
PDB:1crw:R ED.....GALVVDGKKITVFNEM.KPENIPWSKAGAEYIVESTCGVFTTIEKASAEFKGGAKKVHISA  
PDB:1dc4:A KD.....GHLIVNGKKIRVTAER.DPANLKWDEVGVVVAEATCGVFTLDETARRHITAGAKKVVMTG

185.pdb:Q

185.pdb:Q PA.K.GEDITIVMGVNHAEYDPSRHHISNASCTTNCLAPVMKVLEEA.FGVEKALMTTVHSTYNDQRLL  
PDB:1a7k:A PA.SGGA.KTIVMGVNHAEYSPASHHVSNASCTTNCLAPIVHVLTKENFGIETGLMTTTHSYTATQKTV  
PDB:1hdg:O PA.K.GEDITIVMGVNHAEYDQLKPE.HTIIISGASCTTNCLAPIVKVLHEK.FGIVSGMITTVHSTYNDQRVL  
PDB:1crw:R PS.A.DA.PMFVCGVNLEKYSKD.MKVVSNASCTTNCLAPVAKVLHEN.FEIVEGLMTTVHVAVTATQKTV  
PDB:1dc4:A PSKD.NT.PMFVKGANFDKYAG..QDFVSNASCTTNCLAPLAKVINDN.FGIIEGLMTTVHATATQKTV

185.pdb:Q

185.pdb:Q DLPH.KDLRRARRAAAINIIPSTTGAAKAATVLPVSLKCRFDGMALRVPTATGSISDHTALLKREVTAEV  
PDB:1a7k:A DGVSLKDWRGGRRAAVNIIPSTTGAAKAVGMVIPSTKGLTGMSFRVPTFDVSVVDHTFRATRDTSIQEI  
PDB:1hdg:O DLPH.KDLRRARRAAAVNIIPSTTGAAKAVLVPEVKGKLDGMAIRVPTFDGSITDHTVLVEKETTVVEV  
PDB:1crw:R DGFSAKDWRGGRGAAQNIIPSTTGAAKAVGVVPELDGKLTGMARVPTFNVSVVDHTVRLGKECSYDDI  
PDB:1dc4:A DGFSSHKDWRGGRGASQNIIPSTTGAAKAVGVVPELDGKLTGMARVPTFNVSVVDHTVRLKAATYEQI

\*\*\*\*\*

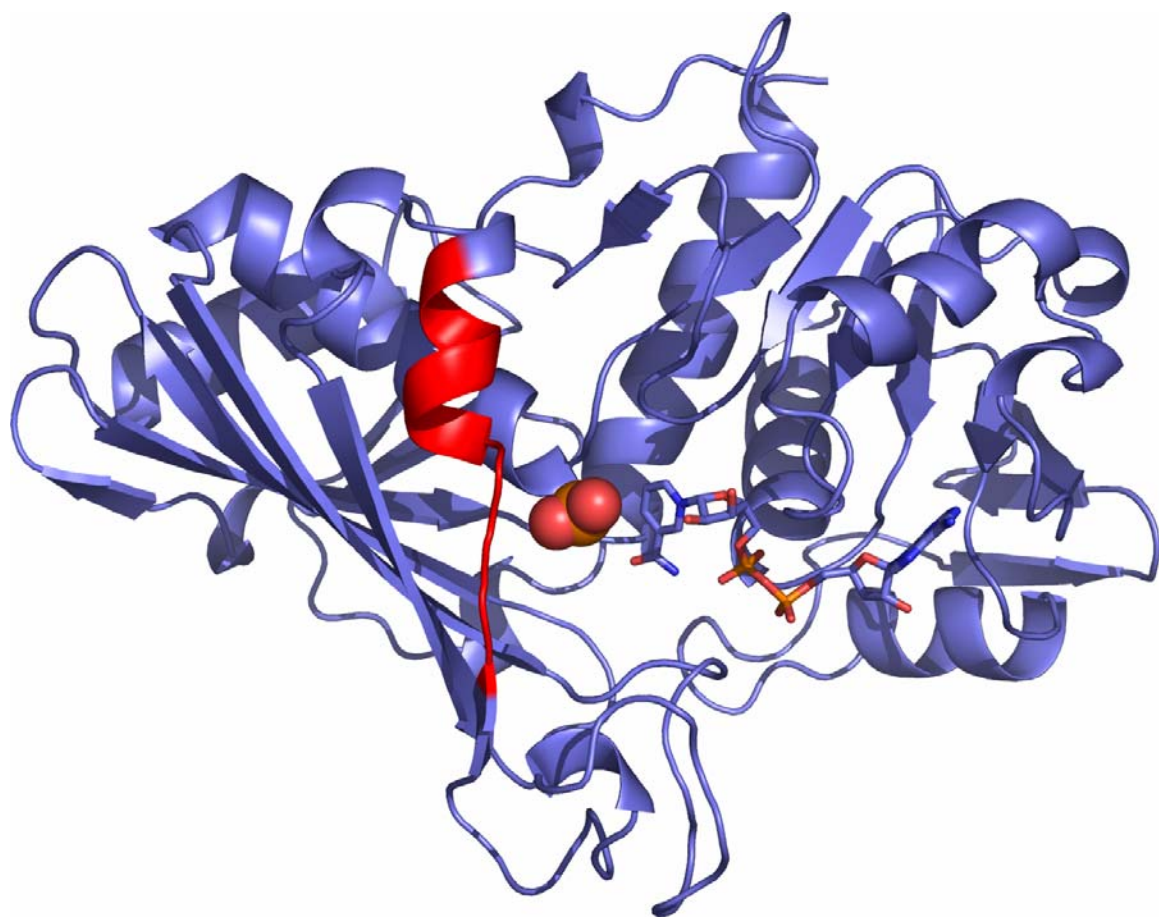
185.pdb:Q

185.pdb:Q NAALKAAAEGLPKGILAYTEDEIVLQDIVMDPHSSIVDAKLT...KA...L.GNMVVKVFAWYDNEWGYANR  
PDB:1a7k:A DKAIKKAAQTYMKGILGFTDEELVSAADFINDNRSSVYDSKATLQNNLPGE.KRFFKVVSWYDNEWGYSHR  
PDB:1hdg:O NAVMKAEATEGRLLKGIIGYNDEPIVSSDIIGTTFSGIFDATIT...NV...IGGKLVKVASWYDNEWGYSNR  
PDB:1crw:R KAAMKAASEGFLQGVLYTEDDVVSCDFTGDNRSSIFDAKAG...IQ...LSKTFVKVSWYDNEFCYSQR  
PDB:1dc4:A KAAVKAAAEGEMKGVLYTEDDVVSTDFNGEVCTSVFDAKAG...IA...LNDNFVKVSWYDNETGYSNK

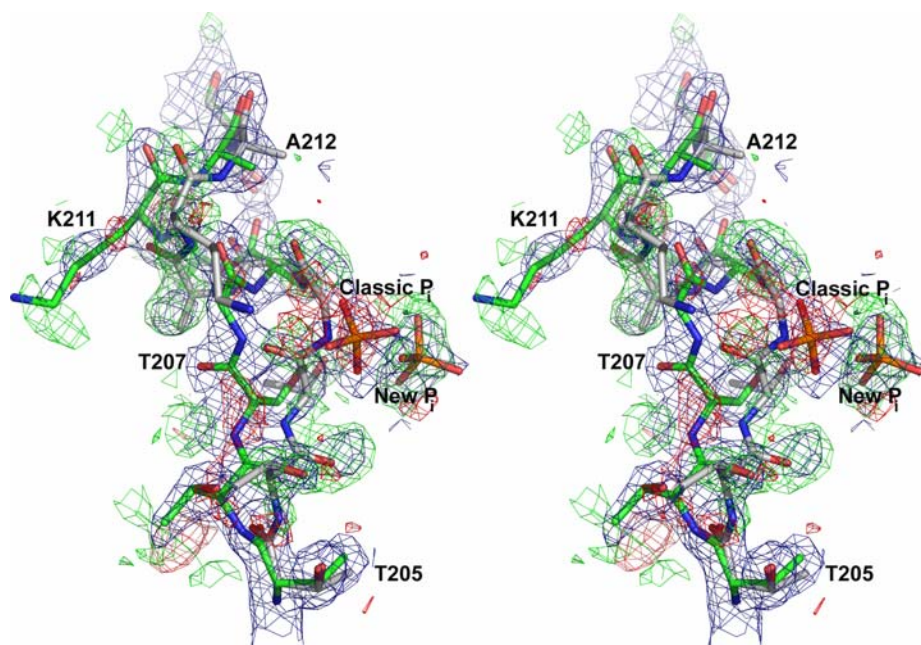
185.pdb:Q

185.pdb:Q VADLVLELVLRKGV...  
PDB:1a7k:A VVDLVRYMAAKDAASS  
PDB:1hdg:O VVDLLELLLLKM...  
PDB:1crw:R VVDLIKHMVKVDSAA...  
PDB:1dc4:A VVDLIAHISK.....

**Figure 2.12.** *Cartoon of the  $P_i$ -binding loop (red) in TaGAPDH. The catalytic cysteine is shown in spheres and  $\text{NAD}^+$  as sticks.*



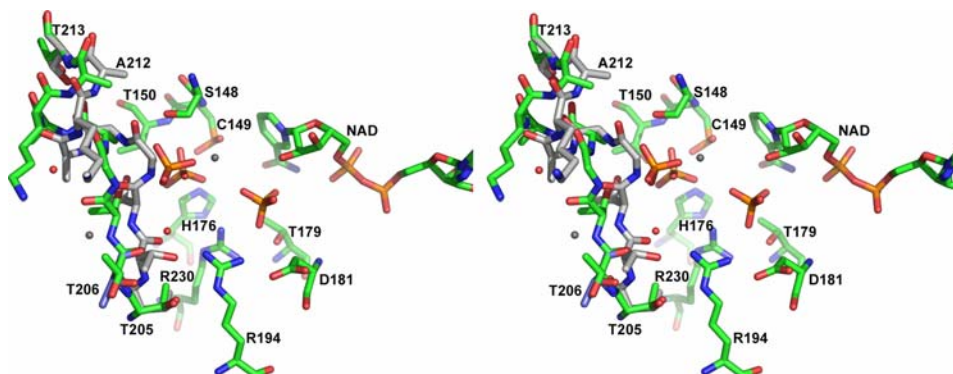
**Figure 2.13.** *The  $P_i$ -binding loop in 50 mM phosphate structure.* The maps were made by omitting the new  $P_i$ -binding loop conformation (205-213), new  $P_i$  phosphate, and surrounding water molecules and running refinement with only the classic loop at full occupancy. The  $2F_o-F_c$  electron density map (blue) is contoured at  $1\sigma$  and the negative (red) and the positive (green)  $F_o-F_c$  maps are contoured at  $-2.5\sigma$  and  $2.5\sigma$ .



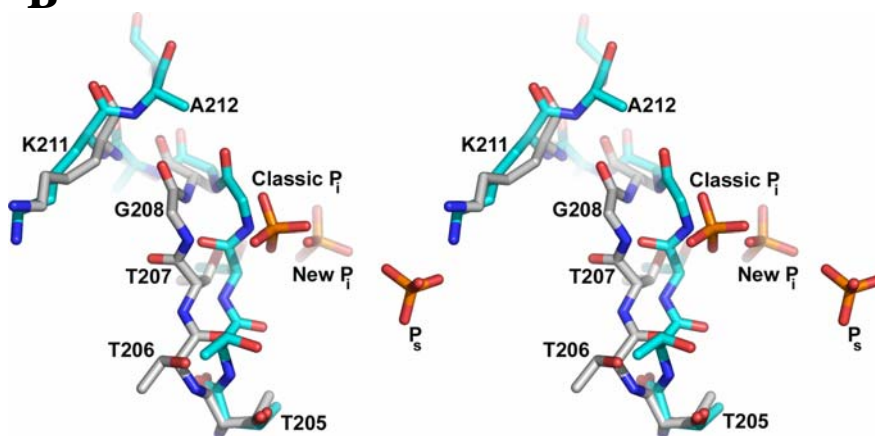
In each subunit of the phosphate-bound structures, the P<sub>i</sub>-binding loop adopts the classic conformation plus one of two other conformations, which we refer to as new1 and new2. Conformation new1 (Figure 2.14A) is equivalent to the previously described conformation found in 1a7k and includes movement of the whole loop 205-213. In new1, the loop moves as much as 3.8 Å from the classical location toward the active site, and the P<sub>i</sub> phosphate is repositioned by ~4.1 Å towards Cys149. Conformation new2 (Figure 2.14B) displays movement of only residues 205-209. As with conformation new1, the new2 conformation is shifted toward the active site relative to the classic conformation.

**Figure 2.14.** *Conformations of the  $P_i$ -binding loops.* The classic conformation of the  $P_i$ -binding loop is colored green and grey in A, and B, respectively. A, New1 conformation of the new  $P_i$ -binding loop displayed by the 50 mM Q subunit. The new1 conformation is shown in grey. Water molecules, colored red and grey, correspond to classic and new1 conformations, respectively. B, The new2 conformation displayed by the D subunit. The new2 loop is colored cyan, and residues 205-213 and the phosphates are labeled.

**A**



**B**



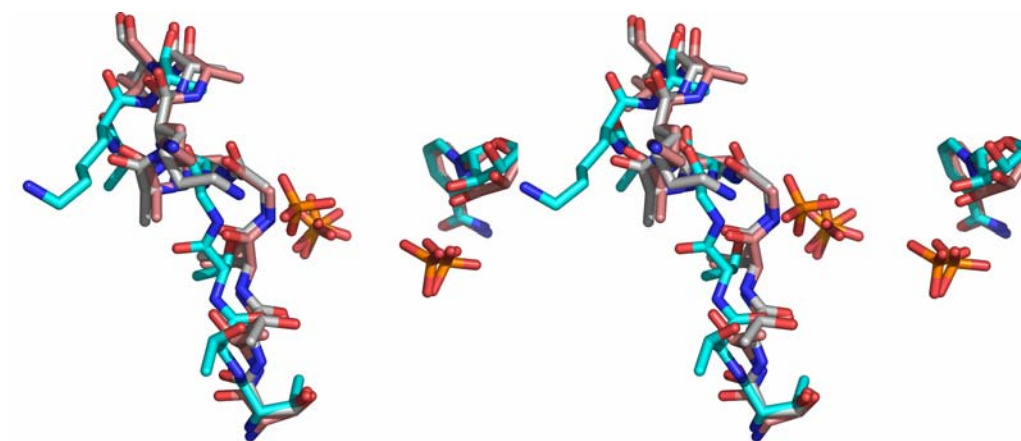


Comparing conformation new1 found in our phosphate-bound structures to other available GAPDH structures shows that it superimposes well (Figure 2.15) with 1A7K and 1GYP that have an overall RMSD of 1.2 Å for all the C<sub>α</sub> atoms. The phosphates occupying the new P<sub>i</sub>-site and the P<sub>s</sub>-site superimpose within 0.4-1.2 Å when compared to those of the 1A7K with the 100 mM structure showing the best fit, as is expected since 1A7K was grown in 100 mM sodium phosphate (8). Due to the resolution of our structures, we made no attempt to refine occupancies for the two P<sub>i</sub>-sites from that of 50/50, but it is possible that they exist at a different ratio in some of the subunits. There might be a preference for one site over the other when comparing the maps between the 50 and 100 mM phosphate structures, with the new P<sub>i</sub>-site phosphate having stronger density, but due to the 2.23 Å resolution, we cannot say this definitively.

When comparing our phosphate containing structures to structures in complex with GAP or GAP analogues (1NQA, 1ML3, and 1QXS), the structures support the flip-flop model of phosphate binding. 1NQA, 1NQO, 1ML3, and 1QXS have substrates with their phosphate moiety bound in the P<sub>s</sub>-site (Figure 2.16). Interestingly, none of the structures have an occupied new or classic P<sub>i</sub>-site, and the authors of each paper offer different interpretations as to why. As explained above, the P<sub>s</sub>-site appears to be the higher affinity phosphate site, so GAP in 1NQA binds but does not induce movement of the P<sub>i</sub>-binding loop (Figure 2.16A). However, the binary (GAPDH/NAD<sup>+</sup>) BsGAPDH structures (1NPT, 1NQ5) show slight movement, 0.2–1.1 Å, of the loop 206-210 (TTTGA), even though only the P<sub>s</sub>-site contains sulfate. The EcGAPDH (11), LmGAPDH, trypanosomatid, *Palinurus versicolor* (33), and plasmodium (34) GAPDH structures all have either an Ala or Gly at the residue 194 instead of the more

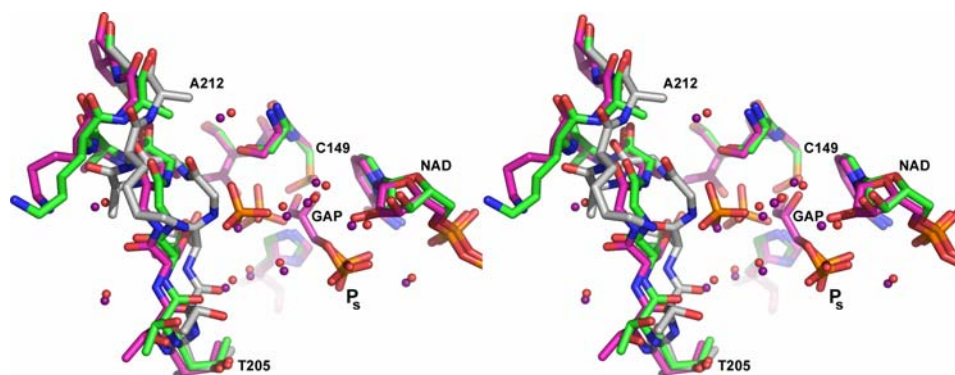
**Figure 2.15.** *Superimposed conformations of the  $P_i$ -binding loops from different GAPDHs.* 1A7K new  $P_i$ -binding loop (salmon) superposed onto the 50 mM B subunit, with the classic and the new  $P_i$ -binding loops colored cyan and grey, respectively.

**A**

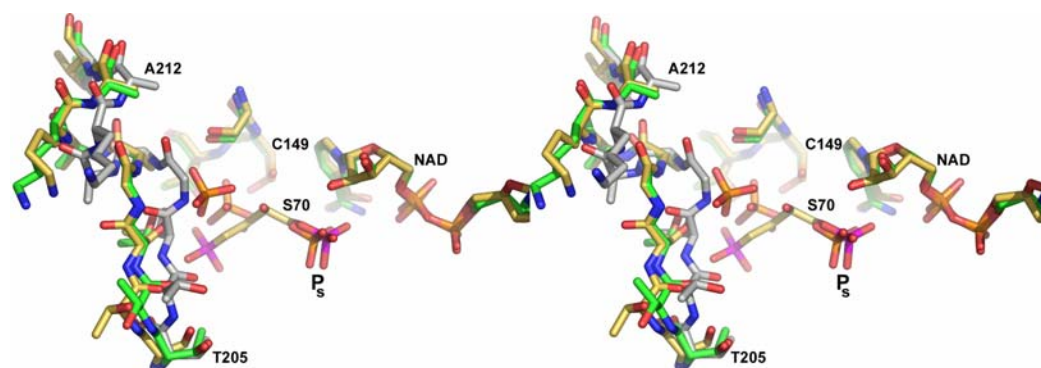


**Figure 2.16.** *Stereoview of superimposed conformations of the  $P_i$ -binding loops from different GAPDHs.* These cartoons show the flexibility of the  $P_i$ -site and the rigidity of the  $P_s$ -site and the lack of movement in the  $P_i$ -binding loop when the  $P_s$ -site is occupied in these structures. *A*, the BsGAPDH ternary complex superposed onto 50 mM B subunit displays the location of the GAP phosphoryl in the  $P_s$ -site and the  $P_i$ -binding loop. 1NQA is colored purple, with the classic and new  $P_i$ -binding loop colored green and grey, respectively. Water molecules colored red and purple are conserved between the two structures and interact with the active site residues and phosphates. *B*, 1QXS superposed onto 50 mM Q subunit displays the location of the two HOP (S70) phosphoryl groups. Subunit Q is colored as in *A*, and 1QXS is colored yellow, with one phosphoryl group in the  $P_s$ -site and the other positioned between the classic and new  $P_i$  sites. The superimposition shows there is no movement of the 1QXS  $P_i$ -binding loop toward the active site. *C*, 1ML3 superposed onto 50 mM B subunit displays the location of the two covalently bound CYX phosphoryl groups. 1ML3 is colored green with the 50 mM classic and new  $P_i$ -binding loop colored cyan and grey respectively. One CYX compound has its phosphoryl group squarely in the  $P_s$  whereas the other has it positioned  $\geq 3$  Å away from the classic and new  $P_i$  sites. 1ML3 displays no  $P_i$ -binding loop movement toward the active site when compared the classic  $P_i$ -binding loop of the TaGAPDH phosphate bound structures.

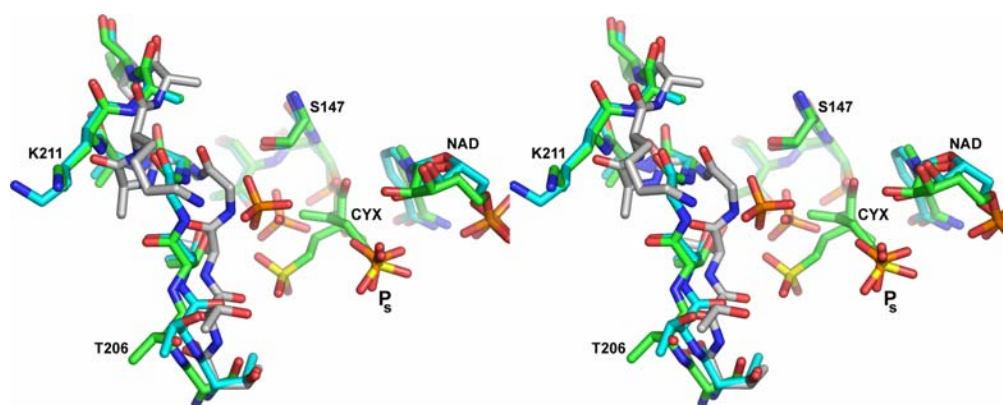
**A**



**B**



**C**



prevalent Arg, thus removing a residue capable of further stabilizing the P<sub>s</sub> phosphate (Figure 2.10) and reducing the positive charge.

Based on the above information, the parasitic GAPDHs would be more likely to bind GAP or a GAP analogue in the P<sub>i</sub>-site and possibly induce loop movement. However, this is not the case, as the TcGAPDH structure 1QXS has the [3(R)-hydroxy-2-oxo-4-phosphonoxybutyl]phosphonic acid (HOP) phosphoryl bound in the P<sub>s</sub>-site (32). The structure (Figure 2.16B) clearly shows that one of the two phosphoryl groups of HOP, a structural analogue of BPGA (IC<sub>50</sub> = 2.0 mM), binds to the P<sub>s</sub>-site, and the other binds  $\geq 4.0$  Å from the classic and  $\geq 2.0$  Å from the new P<sub>i</sub>-sites found in our 50 mM phosphate structure.

The GAP analogue 3-(*p*-nitrophenoxycarboxyl)-3-ethylene propyl dihydroxyphosphonate (CYX) covalently attached to the catalytic cysteine (thioester) in 1ML3 superposes to very similar positions (Figure 2.16C) (7). In both subunits B and C, there is a covalently bound Cys166 and CYX with dual conformations at half occupancy. Each subunit also has NAD<sup>+</sup> at half occupancy. In the D subunit of 1ML3, the NAD<sup>+</sup> is at full occupancy and the CYX phosphoryl moiety is found only at the P<sub>s</sub>-site, whereas subunit A does not have an inhibitor bound, likely due to its lack of NAD<sup>+</sup> which results in low affinity P<sub>s</sub>-site (7). Inspection of the  $2F_o - F_c$  maps for 1ML3 reveals poor density for the majority of the half-occupancy NAD<sup>+</sup> and the dual CYX ligands in subunit B and C. Neither structure, when compared to TaGAPDHs phosphate structures, shows much movement in the P<sub>i</sub>-binding loop. The RMSD for all the C<sub>α</sub> atoms was 1.2 Å. We saw similar poor density for NAD<sup>+</sup> in the structure of human placental GAPDH (1U8F) and chose not to include NAD<sup>+</sup> in the O subunit (35). Castilho et al., questionable refinement

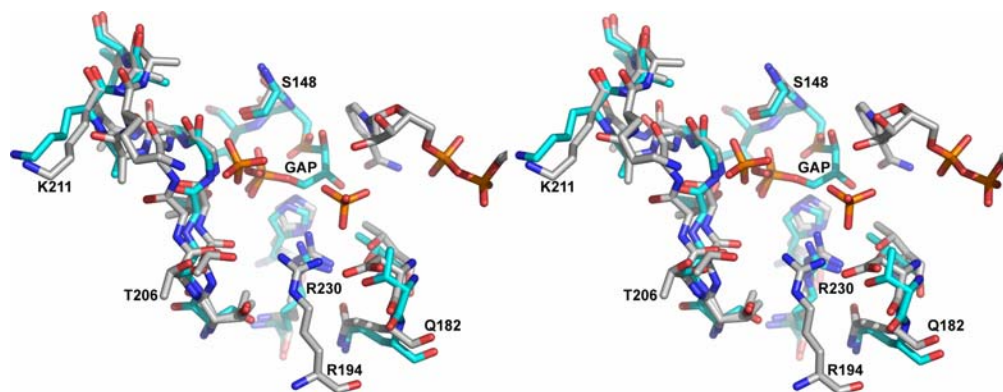
of the occupancy for both  $\text{NAD}^+$  and CYX at 2.5 Å resolution and the poor electron density suggests that the sites should have been left empty, leaving a structure with one  $\text{NAD}^+$  molecule and one CYX molecule with the phosphoryl group in the  $\text{P}_s$ -site only.

The CYX ligand also lacks a hydroxyl group at the C2 position that others (5,11) have found important for binding of D-GAP in the *R* configuration and for interactions with Cys149, Thr179,  $\text{NAD}^+$ , and solvent in the active. The GAPDH-hemithioacetal structure 1DC4, 1NQO, and 1NQA clearly show the importance of the hydroxyl for stabilizing each GAP conformation. It can be clearly seen that GAP in the 1DC4 structure has its phosphoryl group in the  $\text{P}_i$ -site, and that the  $\text{P}_i$ -binding loop is in its new position (Figure 2.17). The structure 1DC4 was solved from crystals of the apo-EcGAPDH soaked in GAP (11). Thus,  $\text{NAD}^+$  was not present in the structure, and with no positively charged Arg230-like residue there is a weaker  $\text{P}_s$ -site than that of BsGAPDH, possibly leading to the phosphoryl binding to the  $\text{P}_i$ -site.

Overall, the 1NQA, 1NQO, 1DC4, 1ML3, and 1QXS structures support the flip-flop model that states that the C3 phosphate of GAP occupies both phosphate sites at different times during the reaction. However, according to the trypanosomatid structures (7,32,36), the rotation occurs between the Cys149 S-C1 bond of GAP and not the C1-C2 bond, as generally accepted (1,5,11,37). The remaining question is when does this flip-flopping occur during the catalytic cycle? According to Castilho et al. (7), the trypanosomatid structures show flip-flopping is related to the cofactor exchange process, whereas Yun et al. (11) suggest it occurs before hydride exchange. However, if the questionable 1QXS B and C subunits displaying a covalently bound GAP analog phosphoryl group in the  $\text{P}_s$ -site are disregarded.

**Figure 2.17.** *A view of the active site of 50 mM structure superposed on to the GAPDH hemiacetal structure. The 50 mM phosphate B subunit superposed onto EcGAPDH hemiacetal complex (1DC4) displays the location of the covalently bound GAP phosphoryl in the P<sub>i</sub>-site and the movement P<sub>i</sub>-binding loop. 50 mM phosphate B subunit (grey) superposed onto 1DC4 (cyan).*





The BsGAPDH ternary structures clearly advocate the initial binding of the GAP phosphoryl to the  $P_s$ -site before formation of a covalent bond with cysteine. The EcGAPDH 1DC4 structure demonstrates that the GAP phosphoryl can bind to the new  $P_i$ -site after the formation of the hemiacetal. We therefore propose a revised flip-flop model consisting of the following steps:

(a) GAP initially binds to GAPDH with the phosphoryl group in the  $P_s$ -site (5,7). Formation of the covalent bond between GAP and the catalytic cysteine (hemithioacetal) causes the phosphoryl group to flip to the new  $P_i$ -site where it is stabilized by interactions with His176 (11), the  $P_i$ -binding loop, the conserved residues Thr150, and Ser148.

(b) After cofactor exchange, the thioester phosphoryl group flips to the  $P_s$ -site allowing the  $P_i$ -site to be occupied by inorganic phosphate during the phosphorylation step. The new  $NAD^+$  is stabilized by hydrogen bonding between the O2' of the nicotinamide ribose and the thioester phosphoryl, and the  $NAD^+$  in turn enhances (4) the phosphorolysis of the thioester.

We propose renaming the  $P_i$ -sites given that all the characterized GAP and GAP analogue structures to date have failed to show an occupied classic  $P_i$ -site. The “new  $P_i$ -site” should be called the “high-affinity  $P_i$ -site” and the “classic  $P_i$ -site” renamed to “low affinity  $P_i$ -site” as it seems to represent a weak binding site that is occupied only at high phosphate/sulfate concentrations.

## CONCLUSION

In conclusion, we have determined three TaGAPDH crystal structures, corresponding to phosphate concentrations zero, 50 mM, and 100 mM. Our structures show that the  $P_i$ -binding loop is quite flexible as illustrated by multiple conformations of residues 205-213 in 50 mM and 100 mM structures. The structures presented herein and the structures of others (5,7) clearly show that, at high and low phosphate concentrations, the fully formed  $P_s$ -site ( $NAD^+$  bound) represents the first choice for phosphoryl binding. The  $P_i$ -sites, which we propose have lower affinity than the  $P_s$ -site, are occupied in subsequent steps during the catalytic cycle. The determination of more GAPDH structures having  $NAD^+$  and covalent reaction intermediates bound to Cys149 might help further our understanding of the GAPDH reaction. They would also be helpful in the structure-based design of inhibitors to trypanosomatid and plasmodium GAPDHs.

## References

1. Corbier, C., Michels, S., Wonacott, A. J., and Branlant, G. (1994) *Biochemistry* **33**, 3260-3265
2. Harris, J. I., and Waters, M. (ed) (1976). *The Enzymes*. Edited by Boyer, P. D., Academic Press, NY
3. Moras, D., Olsen, K. W., Sabesan, M. N., Buehner, M., Ford, G. C., and Rossmann, M. G. (1975) *J Biol Chem* **250**, 9137-9162
4. Kim, H., Feil, I. K., Verlinde, C. L., Petra, P. H., and Hol, W. G. (1995) *Biochemistry* **34**, 14975-14986
5. Didierjean, C., Corbier, C., Fatih, M., Favier, F., Boschi-Muller, S., Branlant, G., and Aubry, A. (2003) *J Biol Chem* **278**, 12968-12976
6. Skarzynski, T., Moody, P. C., and Wonacott, A. J. (1987) *J Mol Biol* **193**, 171-187
7. Castilho, M. S., Pavao, F., Oliva, G., Ladame, S., Willson, M., and Perie, J. (2003) *Biochemistry* **42**, 7143-7151
8. Kim, H., and Hol, W. G. (1998) *Journal of Molecular Biology* **278**, 5-11
9. Korndorfer, I., Steipe, B., Huber, R., Tomschy, A., and Jaenicke, R. (1995) *J Mol Biol* **246**, 511-521
10. Song, S., Li, J., and Lin, Z. (1998) *Acta Crystallographica Section D-Biological Crystallography* **54**, 558-569
11. Yun, M., Park, C. G., Kim, J. Y., and Park, H. W. (2000) *Biochemistry* **39**, 10702-10710
12. Brock, T. D., Smith, D. W., and Madigan, M. T. (1984) *Biology of microorganisms*, 4th Ed., Prentice-Hall, Englewood Cliffs, N.J.
13. Tanner, J., Hecht, R. M., Pisegna, M., Seth, D. M., and Krause, K. L. (1994) *Acta Cryst.* **D50**, 744
14. Otwinoski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307-326
15. Dauter, Z. (1997) *Methods Enzymol.* **276**, 326-344
16. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr D Biol Crystallogr* **54** ( Pt 5), 905-921
17. Collaborative Computational Project, N. (1994) *Acta Crystallogr D Biol Crystallogr* **50**, 760-763
18. Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) *Acta Crystallogr D Biol Crystallogr* **59**, 1131-1137
19. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard. (1991) *Acta Crystallogr A* **47** ( Pt 2), 110-119
20. Tanner, J. J., Hecht, R. M., and Krause, K. L. (1996) *Biochemistry* **35**, 2597-2609
21. Kleywegt, G. J., and Jones, T. A. (1998) *Acta Crystallogr D Biol Crystallogr* **54**, 1119-1131
22. DeLano, W. L. (2002), DeLano Scientific, San Carlos, CA
23. Dror, O., Benyamini, H., Nussinov, R., and Wolfson, H. (2003) *Bioinformatics* **19** Suppl 1, i95-104
24. Shindyalov, I. N., and Bourne, P. E. (1998) *Protein Eng* **11**, 739-747

25. Sali, A., and Blundell, T. L. (1990) *J Mol Biol* **212**, 403-428
26. Rossmann, M. G., Moras, D., and Olsen, K. W. (1974) *Nature* **250**, 194-199
27. Richardson, J. S. (1981) *Adv Protein Chem* **34**, 167-339
28. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) *Nucleic Acids Res* **28**, 235-242
29. Laskowski, R. A., Moss, D. S., and Thornton, J. M. (1993) *J Mol Biol* **231**, 1049-1067
30. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) *Proteins* **50**, 437-450
31. Krissinel, E., and Henrick, K. (2004) *Acta Crystallogr D Biol Crystallogr* **60**, 2256-2268
32. Ladame, S., Castilho, M. S., Silva, C. H., Denier, C., Hannaert, V., Perie, J., Oliva, G., and Willson, M. (2003) *Eur J Biochem* **270**, 4574-4586
33. Shen, Y. Q., Li, J., Song, S. Y., and Lin, Z. J. (2000) *J Struct Biol* **130**, 1-9
34. Robien, M. A., Bosch, J., Buckner, F. S., Van Voorhis, W. C., Worthey, E. A., Myler, P., Mehlin, C., Boni, E. E., Kalyuzhniy, O., Anderson, L., Lauricella, A., Gulde, S., Luft, J. R., DeTitta, G., Caruthers, J. M., Hodgson, K. O., Soltis, M., Zucker, F., Verlinde, C. L., Merritt, E. A., Schoenfeld, L. W., and Hol, W. G. (2006) *Proteins* **62**, 570-577
35. Jenkins, J. L., and Tanner, J. J. (2006) *Acta Crystallogr D Biol Crystallogr* **62**, 290-301
36. Souza, D. H., Garratt, R. C., Araujo, A. P., Guimaraes, B. G., Jesus, W. D., Michels, P. A., Hannaert, V., and Oliva, G. (1998) *FEBS Letters* **424**, 131-135
37. Skarzynski, T., and Wonacott, A. J. (1988) *Journal of Molecular Biology* **203**, 1097-1118

## Chapter 3

# CRYSTAL STRUCTURES OF THE DNA-BINDING DOMAIN OF THE MULTIFUNCTIONAL PUTA FLAVOPROTEIN<sup>†,‡</sup>

*John D. Larson,<sup>§</sup> Jermaine L. Jenkins,<sup>⊥</sup> Jonathan P. Schuermann,<sup>§</sup> Yuzhen  
Zhou,<sup>||</sup> Donald F. Becker,<sup>||</sup> and John J. Tanner<sup>§,⊥,\*</sup>*

Departments of Chemistry and Biochemistry, University of Missouri-Columbia,  
Columbia, MO 65211, and Department of Biochemistry, Redox Biology Center,  
University of Nebraska, Lincoln, NE 68588

AUTHOR EMAIL ADDRESS: [tannerjj@missouri.edu](mailto:tannerjj@missouri.edu)

---

<sup>†</sup> The work was supported by NIH Grants GM065546 (to J.J.T.) and GM061068 (to D.F.B.), NSF Grant MCB0340912 (to D.F.B) and the Nebraska Agricultural Research Division.

<sup>‡</sup> Coordinates and structure factors have been deposited in the Protein Data Bank under Accession Numbers 2AY0 and 2GPE.

<sup>§</sup> University of Missouri-Columbia Department of Chemistry

<sup>⊥</sup> University of Missouri-Columbia Department of Biochemistry

<sup>||</sup> University of Nebraska

\* Address correspondence to this author. Tel.: 573-884-1280; Fax: 573-882-2754; E-mail: [tannerjj@missouri.edu](mailto:tannerjj@missouri.edu)

## ABSTRACT

PutA (Proline Utilization A) from *Escherichia coli* is a 1320 amino acid residue protein that functions as both a bifunctional proline catabolic enzyme and transcriptional regulator of *put* genes. As a regulatory protein, *E. coli* PutA functions as an autogenous transcriptional repressor by virtue of a DNA-binding domain located in the N-terminus of the protein. Here we report the first crystal structure of a PutA DNA-binding domain. Crystals were grown using a polypeptide corresponding to residues 1-52 of *E. coli* PutA (PutA52). The structure of native PutA52 was determined to 1.9 Å resolution and the structure of PutA52 mutant Lys9Met was solved at 2.1 Å resolution. Residues 3-46 form a ribbon-helix-helix (RHH) substructure, which establishes PutA as the largest example of a RHH protein to date. The PutA RHH domain forms the intertwined dimer that is characteristic of the RHH family. Comparison of PutA52K9M to other RHH proteins reveals structural features unique to the PutA RHH subfamily. These subfamily-defining features include a glycine (Gly7 in PutA) in a critical position of the  $\beta$ -sheet, arginine on the surface of the first helix (Arg15 in PutA), absence of Gly in the inter-helix turn (Asp26 in PutA), a proline (Pro29) at the N-terminus of the second helix, and a solvent exposed tryptophan (Trp31) in the second helix. Homology models of the PutA RHH domain complexed with DNA predict that Lys9 and Thr5 contact DNA bases through the major groove, while Arg15, Thr28 and His30 interact with the phosphate backbone. We demonstrate that mutation of Lys9 to Met in full-length *E. coli* PutA eliminates binding to *put* control intergenic DNA, which shows that Lys9 is essential for specific DNA binding. These results provide new insights into the structural basis of PutA dimerization and DNA recognition.

## INTRODUCTION

Proline utilization A (PutA) is a large, membrane-associated bi-functional enzyme that catalyzes the sequential two-step oxidation of proline to glutamate (1-6). In the first step, proline is oxidized to  $\Delta^1$ -pyrroline-5-carboxylate (P5C) by the FAD-dependent PutA proline dehydrogenase (PRODH) domain. P5C is hydrolyzed nonenzymatically to glutamic semialdehyde, and the semialdehyde is oxidized to glutamate by the NAD<sup>+</sup>-dependent PutA P5C dehydrogenase (P5CDH) domain. PutAs typically contain 1000-1300 amino acid residues, with the PRODH domain located in the N-terminal half of the polypeptide chain and the P5CDH domain located in the C-terminal half. PutA is almost exclusively found in gram-negative bacteria. In contrast, PRODH and P5CDH appear as separate enzymes encoded by distinct genes in most gram-positive bacteria (7) and eukaryotes (8).

In addition to their dual catalytic functions, some PutAs are responsible for transcriptional regulation of the proline utilization genes, *putA* and *putP* (encodes the high affinity Na<sup>+</sup>-proline transporter) (9-14). Examples of this PutA subfamily include PutA from *Escherichia coli*, *Salmonella typhimurium* and *Pseudomonas putida*. These “trifunctional” PutAs repress expression of the *put* genes, which are transcribed in opposite directions, by binding to the *put* intergenic DNA region. PutA switches from a transcriptional repressor to a membrane bound enzyme by translocating from the cytoplasm to a peripheral position on the membrane, where proline is efficiently converted to glutamate. Reduction of the PutA FAD cofactor by the substrate proline triggers PutA to associate with the membrane, thus disrupting PutA-DNA association and activating transcription of the *put* genes (2, 6, 14, 15).



*E. coli* PutA is a polypeptide of 1320 amino acids and purifies mainly as a dimer with a molecular mass of ~293 kDa (10). We have previously determined crystal structures of the PRODH domain bound to various inhibitors using a truncated protein corresponding to residues 86-669 of *E. coli* PutA (16, 17). The structures showed that the core of the PRODH domain is a unique  $\beta_8\alpha_8$  barrel (residues 263 - 561) with  $\alpha_8$ , which contains essential catalytic residues Arg555 and Arg556, located above the carboxyl face of the barrel rather than alongside the barrel. Prior to the work described below, the PRODH domain was the only PutA domain to be structurally characterized.

We recently showed that the DNA-binding domain of *E. coli* PutA is contained in residues 1 – 47, based on a combination of molecular dissection experiments and amino acid sequence similarity analysis (18). The former approach established that a polypeptide corresponding to PutA residues 1 – 47 (PutA1-47) exhibited *in vivo* transcriptional repressor activity and bound specifically to *put* intergenic DNA *in vitro* with  $K_d = 15$  nM. The latter approach suggested that PutA belongs to the ribbon-helix-helix (RHH) family of transcription factors.

As part of our ongoing studies of PutA structure and function, we report the first crystal structures of a PutA DNA-binding domain. Crystals of this domain were obtained from a polypeptide corresponding to *E. coli* PutA residues 1-52 (PutA52). The PutA52 structures show that PutA belongs to the RHH family, which establishes PutA as the largest RHH family member. The PutA RHH domain forms the tight, intertwined dimer that is characteristic of the RHH family. Based on analysis of the structure and comparison to other RHH proteins, we propose several residues that may be important for dimerization and DNA recognition. We tested one of these predictions by mutating Lys9

to Met in full-length PutA, which abrogated binding to *put* intergenic DNA. These structures provide key information for understanding structure-function relationships in PutA and new insights into the functional domain arrangement of PutA.

## EXPERIMENTAL PROCEDURES

### *Materials.*

Unless stated otherwise, all chemicals and buffers were purchased from Fisher Scientific and Sigma-Aldrich, Inc. Restriction endonucleases and T4 DNA ligase were purchased from Fermentas and Promega, respectively. BCA reagents used for protein quantitation were obtained from Pierce. All experiments used 18 MΩ water. *E. coli* strains XL-Blue and BL21 (DE3) pLysS were purchased from Novagen. The *put* intergenic DNA (419 base pairs) used for DNA binding assays was prepared as described previously using genomic DNA from *E. coli* strain JT31 (18).

### *Subcloning and Site-directed Mutagenesis of PutA and PutA52.*

The region of the *putA* gene that codes for the N-terminal 52 amino acids of *E. coli* PutA was PCR-amplified using the construct PutA-pET23b (19) as a template. The T7 promoter primer and the primer 5'-GCGATACTCTGCTCGAGCTACCTGCGC-3', which introduced a *Xho*I site, were used for the PCR reaction. The PCR-amplified *putA52* gene product was subsequently subcloned into pET23b using *Nde*I and *Xho*I to produce the plasmid PutA52-pET23b. The encoded protein includes residues 1-52 and a noncleavable C-terminal His<sub>6</sub> tag.

A PutA52 construct containing a cleavable N-terminal His<sub>8</sub> tag was generated by mutagenesis of the plasmid PutA-pKA8H. Using QuickChange, the codon for Leu53 was mutated to an ochre (UAA) stop codon using the following oligonucleotide and its complement: 5'- GCGATACTCTGCCGGAGTAACCTGCGCTG CTTTCTG-3'. The

encoded protein includes a His8 tag, followed by a tobacco etch virus protease (TEVP) cleavage site and PutA residues 1-52.

Plasmids coding for *E. coli* PutA mutant Lys9Met (PutAK9M) and PutA52 mutant Lys9Met (PutA52K9M) were prepared using the QuikChange (Stratagene) site-directed mutagenesis kit with PutA-pET23b and PutA52-pET23b serving as the templates. The following oligonucleotide and its complement were used for Lys9Met mutagenesis: 5'- ACCATGGGGGTTATGCTGGACGACGCG-3'. The clones and site-directed mutants were confirmed by nucleic acid sequencing.

*Preparation of PutA Proteins for DNA-binding Studies.*

*E. coli* PutA, PutAK9M, and PutA52 were expressed as C-terminal His<sub>6</sub> tag fusion proteins from pET23b in *E. coli* strain BL21(DE3) pLysS using protocols similar to those used previously for PutA proteins (16, 18), except that expression of PutA52 was induced with IPTG at 25° C for 6 h. After purification, PutAK9M was dialyzed into 70 mM Tris (pH 8.1) containing 2 mM EDTA, whereas PutA52 was dialyzed into 50 mM potassium phosphate buffer (pH 7.5) containing 0.2 M NaCl. The C-terminal His<sub>6</sub> tags were retained after purification. The purity of each protein was greater than 95% as judged by SDS-PAGE analysis. Protein concentration was determined using the BCA method (Pierce) with bovine serum albumin as the standard and spectrophotometrically using a molar extinction coefficient of 12,700 M<sup>-1</sup> cm<sup>-1</sup> ( $\lambda$  = 451 nm) for PutA and PutAK9M, and an extinction coefficient of 6970 M<sup>-1</sup> cm<sup>-1</sup> ( $\lambda$  = 280 nm) for PutA52 (4).

### *DNA Binding Assays.*

Nondenaturing gel electrophoretic mobility shift assays were used to test the binding of C-terminally His-tagged PutA proteins (PutA, PutAK9M, PutA52) to *put* control intergenic DNA as described previously (18). Fluorescently labeled *put* intergenic DNA was used in the binding reactions. This reagent was prepared by PCR amplification of the *put* intergenic DNA using a synthetic oligonucleotide labeled at the 5' end with IRdye-700 (LI-COR, Inc.). The PutA-DNA complexes were separated using a polyacrylamide (4%) nondenaturing gel at 4° C as previously described (18). Calf thymus competitor DNA (100 µg/ml) was added to the binding mixtures to prevent nonspecific PutA-DNA interactions. The gels were visualized using a LI-COR Odyssey Imager. Measurement of the dissociation constant for binding of PutA52 to *put* intergenic DNA gave a value of  $10.3 \pm 2.4$  nM (data not shown), which is similar to previously determined values for PutA ( $K_d = 45$  nM) and PutA47 ( $K_d = 15$  nM) (4, 18).

### *Preparation of Selenomethionyl PutA52K9M.*

The selenomethionyl (Se-Met) derivative of PutA52K9M was expressed as C-terminal His<sub>6</sub> tag fusion protein from pET23b in *E. coli* strain BL21(DE3)pLysS using the metabolic inhibition method (20). The cells were grown in 1.0-L of M9 media at 37 °C and 250 rpm until the optical density ( $\lambda = 600$  nm) reached 0.8. An amino acid mixture containing selenomethionine (20) was added to the cultures and incubation continued at 37 °C and 250 rpm for an additional 30 min. The cultures were induced by addition of 0.5 mM IPTG and incubation was continued for 8 h at 22 °C and 200 rpm.

Pelleted *E. coli* cells were resuspended on ice in binding buffer (20 mM Tris, 5 mM imidazole, 0.5 M NaCl, pH 7.9) supplemented with Benzonase nuclease (2.5 U/mL), 1 mM dithiothreitol and five protease inhibitors (0.1 mM TPCK, 0.05 mM AEBSF, 0.1  $\mu$ M Pepstatin, 0.01 mM Leupeptin, 5  $\mu$ M E-64). The resuspended cells were disrupted with two passes through a French pressure cell at 110 MPa and placed on ice. The disrupted cells were centrifuged at 35,000 x g and 4 °C for 45 min. The supernatant fraction was collected, placed on ice, and subsequently added to a Ni-NTA affinity column equilibrated with binding buffer. Contaminating proteins were removed from the column by successive washes with the binding buffer followed by the binding buffer supplemented with 20 mM imidazole. PutA52K9M was eluted with the binding buffer supplemented with 500 mM imidazole. The purified protein sample was dialyzed overnight against 2.0 L of 70 mM Tris pH 8.1 buffer containing 2 mM EDTA, 5 mM dithiothreitol, and 5% glycerol. The protein was concentrated to 18 mg/mL using a centrifugal concentrator. Protein concentration was determined using the Bradford method (Pierce Coomassie Plus). SDS-PAGE analysis using a Coomassie protein stain method demonstrated that the protein preparations were purified to greater than 95 % homogeneity. The C-terminal hexahistidine affinity tag was not removed. Mass spectral analysis of PutA52K9M and Se-Met PutA52K9M indicated molecular masses of 6662.9 Da and 6754.42 Da, respectively, suggesting an average incorporation of 2 Se per polypeptide chain.

*Preparation of PutA52 for Crystallization Trials.*

PutA52 for crystallization trials was expressed as an N-terminal His<sub>8</sub> tag fusion protein in *E. coli* strain BL21(DE3)pLysS. Cells were grown in 4.0-L of LB media at 37 °C with rotation rate of 250 rpm until the optical density reached  $A_{600} = 0.5$ . The temperature and rotation rate were reduced to 25 °C and 220 rpm, respectively, and IPTG was added to a final concentration of 0.5 mM. After 4 hours of induction, the cells were harvested and disrupted as described above for PutA52K9M.

The protein was purified using Ni-affinity chromatography as described above for PutA52K9M. The His<sub>8</sub> tag was removed by incubating the purified protein with His-tagged TEVP for 2 h at 30° C followed by dialysis overnight into 50 mM Tris-HCl pH 8.0 and 300 mM NaCl. The mixture was loaded onto the Ni-affinity column and the flow-through was collected. Tag-free PutA52 was then dialyzed into the pre-crystallization buffer (20 mM Tris pH 8.0, 125 mM NaCl, and 5 mM Imidazole) and concentrated using a centrifugal concentrator to 25 mg/ml. Protein concentration was measured using the BCA method (Pierce).

#### *Crystallization of PutA52 and Se-Met PutA52K9M.*

Crystallization trials for both proteins were performed at 22 °C using the hanging drop method of vapor diffusion with drops formed by mixing equal volumes of the reservoir and protein solutions. Commercially available crystal screen kits (Hampton Research and Decode Genetics) were used to identify initial crystallization conditions.

Rod-shaped crystals of PutA52 grew in Index condition 57 (50 mM ammonium sulfate, 50 mM Bis-Tris pH 6.5, 35% pentaerythritol ethoxylate (15/4 EO/OH)). The mother liquor provided cryoprotection, so the crystals were simply picked up with

Hampton mounting loops and plunged into liquid nitrogen. These crystals belong to the space group  $P4_12_12$  and have unit cell dimensions of  $a = 55.73 \text{ \AA}$ ,  $c = 125.02 \text{ \AA}$ . There are 4 PutA52 chains per asymmetric unit, with 39 % solvent and Matthews coefficient of  $2.0 \text{ \AA}^3/\text{Da}$  (21).

A monoclinic crystal form of Se-Met PutA52K9M was obtained from Index reagent 7. After optimization, crystals were grown over reservoir solutions of 1.7 - 3.0 M NaCl, 0.1 M Na citrate pH 3.0 - 5.5 and 5 mM dithiothreitol. In preparation for low temperature data collection, the crystals were equilibrated in a solution of 3.2 M NaCl, 0.1 M Na citrate pH 3.0, 5 mM dithiothreitol and 30 % glycerol. The crystals have space group C2 with unit cell dimensions of  $a = 72.1$ ,  $b = 91.5$ ,  $c = 69.6$ ,  $\beta = 119.2^\circ$ . There are 6 PutA52K9M chains per asymmetric unit, with 50 % solvent and Matthews coefficient of  $2.5 \text{ \AA}^3/\text{Da}$ .

#### *Structure Determination of PutA52K9M.*

The structure of PutA52K9M was solved using multiwavelength anomalous dispersion (MAD) phasing from data collected at Advanced Light Source beamline 4.2.2. Following acquisition of an absorption spectrum from one crystal, three data sets were collected from a second crystal at wavelengths corresponding to the peak of absorption for Se, ascending inflection point and a remote low energy. Each data set consisted of 360 frames collected with oscillation angle of 1 degree per frame, exposure time of 1 s per frame and detector distance of 125 mm. The data frames were accumulated by iterating through the three collection energies in  $2^\circ$  wedges in order to minimize errors due to radiation damage. Integration and scaling were performed using d\*trek (22).  $I^+$



and  $I^-$  were treated as equivalent reflections during scaling. Data collection and processing statistics are listed in Table 3.1.

SOLVE (23) was used to determine the Se substructure and RESOLVE (24) was used for density modification and automated electron density map interpretation. SOLVE found all 12 of the expected Se sites in the asymmetric unit and the resulting MAD phases had figure of merit = 0.48 for 20 - 2.2 Å data. After density modification, the figure of merit increased to 0.54 for all reflections. The backbone tracing from RESOLVE consisted of 6 chains having a total of 256 residues with the sequence assigned to 229 residues. The model was improved with several rounds of model building in COOT (25) followed by refinement with REFMAC5 (26) against the remote energy data set. Six chloride ions were modeled at the locations of strong spherical electron density features that appeared in  $F_o-F_c$  maps at levels of 11 – 16  $\sigma$ . When water molecules were modeled at these sites,  $F_o-F_c$  density was observed at levels of 6.5 – 8.0  $\sigma$ . Each chloride ion binds between two PutA52K9M protomers and interacts with the guanidinium of Arg27, indole N-H of Trp31 and backbone amine of Met9. Refinement statistics for the final model are listed in Table 3.1. Coordinates and structure factors have been deposited in the Protein Data Bank (PDB, (27)) under accession code 2AY0.

#### *Structure Determination of PutA52.*

A data set consisting of 180 frames was collected at ALS beamline 4.2.2 with oscillation angle of 1° per frame, exposure time of 5 s per frame, and detector distance of 110 mm. Integration and scaling calculations were performed with d\*trek. Data collection and processing statistics are listed in Table 3.1.

The structure of PutA52 was solved with molecular replacement using a PutA52K9M monomer as a search model. Phaser (28) was used for molecular replacement calculation using data in the range 51 - 2.5 Å resolution. The best solution was found for group P4<sub>1</sub>2<sub>1</sub>2 as indicated by an absence of steric clashes, log-likelihood gain value of 1133.63, and Z-score of 25. The model from molecular replacement was improved with rigid-body refinement followed by several alternating rounds of model building with COOT and restrained refinement with REFMAC5 (with TLS). A strong electron density feature consistent with a small planar molecular was present in all four chains near Gln35. The pre-crystallization buffer contained 5 mM imidazole, which lead us to build one imidazole molecule into each location. Each bound imidazole molecule is sandwiched between the aromatic rings of Trp31 and Phe38 and forms a hydrogen bond with the side chain of Gln35. After refinement, the imidazole molecules had an average B-factor of 39 Å<sup>2</sup>. Coordinates and structure factors have been deposited in the PDB and assigned the accession code 2GPE. See Table 3.1 for refinement statistics.

The programs CNS, CCP4i, and Pymol (29) were used for structural analysis. Superimposition of structures and root mean square difference (RMSD) calculations were done using SSM (30) and CE (31).

#### *Modeling of PutA52/DNA Complex.*

Qualitative models of PutA52 complexed with DNA were built using structures of Arc/DNA (PDB code 1BDT,(32)), MetJ/DNA (PDB code MJ2, (33)) and CopG/DNA (PDB code 1B01, (34)) as templates. For each template, a PutA52 dimer, was superimposed onto the template structure by fitting the backbone of PutA52 β1 residues

5-9 (both subunits) to the equivalent residues of the template using CNS (35) (residues 9 - 13 of Arc, residues 23 - 27 of MetJ, residues 4 - 8 of CopG). Missing atoms of Lys9 were added with COOT. The resulting models of PutA52 were combined with the DNA coordinates from the respective template and input to energy minimization to relieve steric conflicts. The minimization protocol consisted of 200 steps of conjugate gradient energy minimization in CNS using polar hydrogen atoms, 13 Å nonbonded cutoff and continuum solvent model (dielectric constant = 80).

**Table 3.1.** Data Collection and Refinement Statistics<sup>a</sup>

Se-Met PutA52K9M				PutA52
	Remote	Inflection	Peak	
Space group	C2	C2	C2	P4 <sub>1</sub> 2 <sub>1</sub> 2
Wavelength (Å)	0.987121	0.979553	0.979144	1.2398
Diffraction resolution (Å)	36.5 - 2.1 (2.17 - 2.10)	36.5 - 2.1 (2.17 - 2.10)	36.5 - 2.1 (2.17 - 2.10)	50.9 - 1.9 (1.97 - 1.90)
No. of observations	171983	172046	172042	211825
No. of unique reflections	23035	23020	23013	16253
Redundancy	7.47 (7.43)	7.47 (7.46)	7.48 (7.49)	13.03 (8.29)
Completeness (%)	99.9 (99.9)	99.9 (99.9)	99.9 (99.9)	100 (98.0)
R <sub>merge</sub>	0.045 (0.335)	0.051 (0.367)	0.075 (0.398)	0.043 (0.317)
Average I/σ	21.4 (5.6)	18.4 (5.0)	13.8 (4.6)	25.5 (5.7)
Wilson B-factor (Å <sup>2</sup> )	41	41	40	32.3
PDB accession code	2AY0			2GPE
No. of protein chains	6			4
No. of protein residues	261			183
No. of protein atoms	2072			1493
No. of water molecules	67			68
No. of chloride ions	6			
No. of imidazole molecules				4
R <sub>cryst</sub>	0.204 (0.234)			0.203 (0.257)
R <sub>free</sub> <sup>b</sup>	0.242 (0.295)			0.249 (0.364)
RMSD <sup>c</sup>				
Bond lengths (Å)	0.011			0.009

Bond angles (deg.)	1.1	1.09
Ramachandran plot <sup>d</sup>		
Favored (%)	97.5	99.4
Allowed (%)	2.5	0.6
Generous (%)	0.0	0.0
Disallowed (%)	0.0	0.0
Average B-factors ( $\text{\AA}^2$ )		
Protein	41	36
Water	38	41
Chloride ions	35	
Imidazole		39

---

<sup>a</sup>Values for the outer resolution shell of data are given in parenthesis.

<sup>b</sup>5%  $R_{\text{free}}$  test set.

<sup>c</sup>Compared to the Engh and Huber force field (36).

<sup>d</sup>The Ramachandran plot was generated with PROCHECK (37).

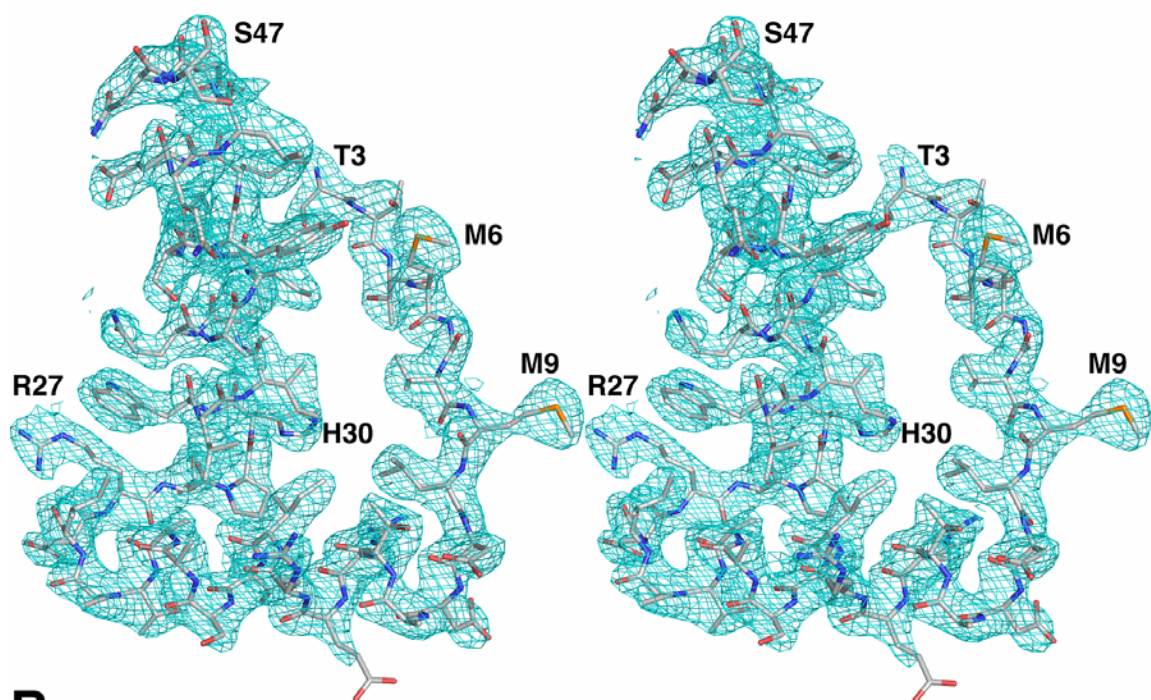
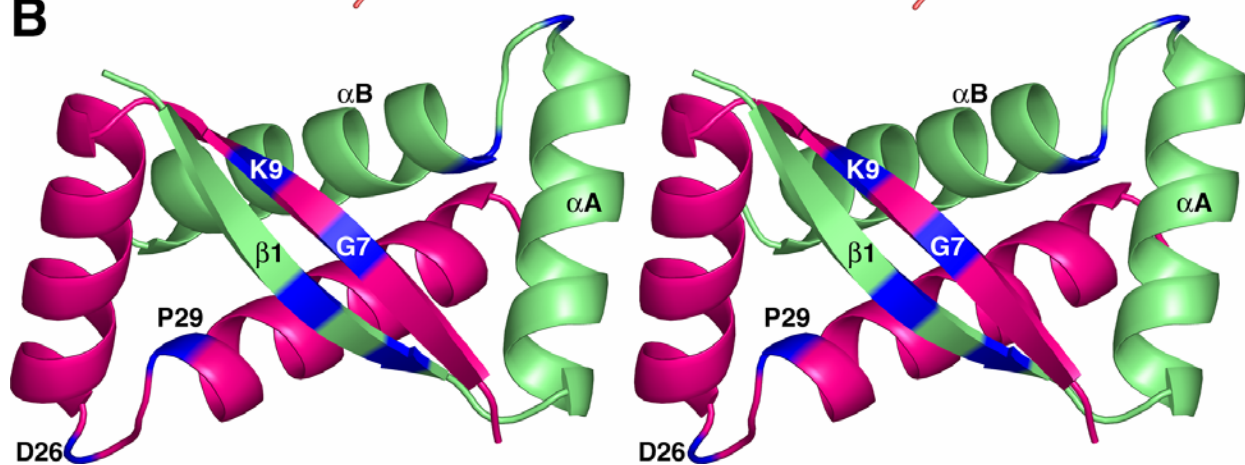
## RESULTS

### *Overall Fold.*

The asymmetric of monoclinic PutA52K9M contains 6 protein chains, which are labeled A-F, 6 chloride ions and 67 water molecules (Table 3.1). The density modified, experimental electron density map was of very high quality (Figure 3.1A) and allowed unambiguous modeling of residues 3-44 of all 6 chains. After a few rounds of model building and refinement,  $2F_o-F_c$  maps revealed the locations of residue 2 of chains C and E, residue 45 of chains A, C, E and F, and residues 45-47 of chain D. The remaining residues of PutA52K9M are presumed to be disordered. The 6 chains have pairwise root mean square differences (RMSDs) of 0.22 - 0.49 Å for  $C_\alpha$  atoms, which indicates that the 6 chains have, essentially, identical overall conformations.

The tetrameric crystal form contains four PutA52 chains (labeled A-D), 68 water molecules and 4 imidazole molecules. Initial electron density maps were of sufficient quality to allow for modeling of residues 2-47 in all chains. After a several rounds of model building and refinement, chain A was extended to include residue 48 and chain D was extended to residue 49. Alternative side chain conformations were modeled for Ser39 and Val8 (C and D chains). The 4 chains have pairwise RMSDs of 0.22 - 0.53 Å for  $C_\alpha$  atoms. RMSDs between PutA52 and PutA52K9M chains are in the range 0.39 - 0.69 Å, which indicates, essentially, identical conformations.

**Figure 3.1.** *Structure of the PutA RHH domain.* (A) Stereo view of one PutA52K9M chain, covered by a figure of merit weighted electron density map ( $1\ \sigma$ ) calculated using amplitudes from the remote energy data set and density-modified MAD phases from RESOLVE. This map was used to build the initial model of PutA52K9M. The model shown is the final, refined structure of the D chain. (B) Ribbon drawing of the PutA RHH dimer. The two subunits are colored red and green. Secondary structure elements are labeled  $\beta 1$ ,  $\alpha A$  and  $\alpha B$  for the green subunit. The blue patches indicate locations of Gly7, Lys9 (Met9), Asp26 and Pro29. This figure, and others, were prepared with PyMol (29).

**A****B**



Each PutA52 chain exhibits the RHH fold (Figure 3.1B), consisting of a  $\beta$ -strand formed by residues 3-11 ( $\beta$ 1) followed by two  $\alpha$ -helices formed by residues 12-25 ( $\alpha$ A) and 29-46 ( $\alpha$ B). The RHH subunits assemble into dimers. Each dimer features an intermolecular antiparallel 2-stranded  $\beta$ -sheet (Figure 3.1B). There are 8 - 9 main chain hydrogen bonds within the sheet. The two subunits of the dimer intertwine such that  $\alpha$ B of one subunit packs against all three secondary structural elements of the other subunit (Figure 3.1B). As a result, both subunits contribute nonpolar side chains to the hydrophobic core of the dimer. The following residues are the largest contributors of intersubunit nonpolar buried surface area: Met6, Val8, Ile37, Tyr40, Leu41. These side chains each bury 66 - 84  $\text{\AA}^2$  of surface area and all 5 residues are highly conserved in PutAs. The intertwined subunits, intermolecular  $\beta$ -sheet and tightly packed hydrophobic intersubunit interface are all commonly found in other RHH proteins, which shows that PutA exhibits the classic RHH dimer structure.

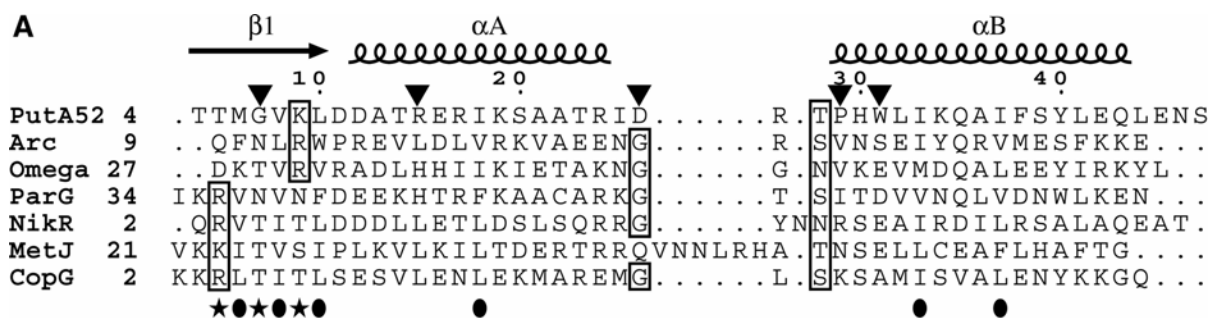
### *Conserved RHH Sequence-Structure Relationships Present in PutA.*

Although the pairwise amino acid sequence identities between members of different RHH subfamilies are only 10-15 % (38), there are a few sequence motifs that are almost universally shared by RHH proteins.

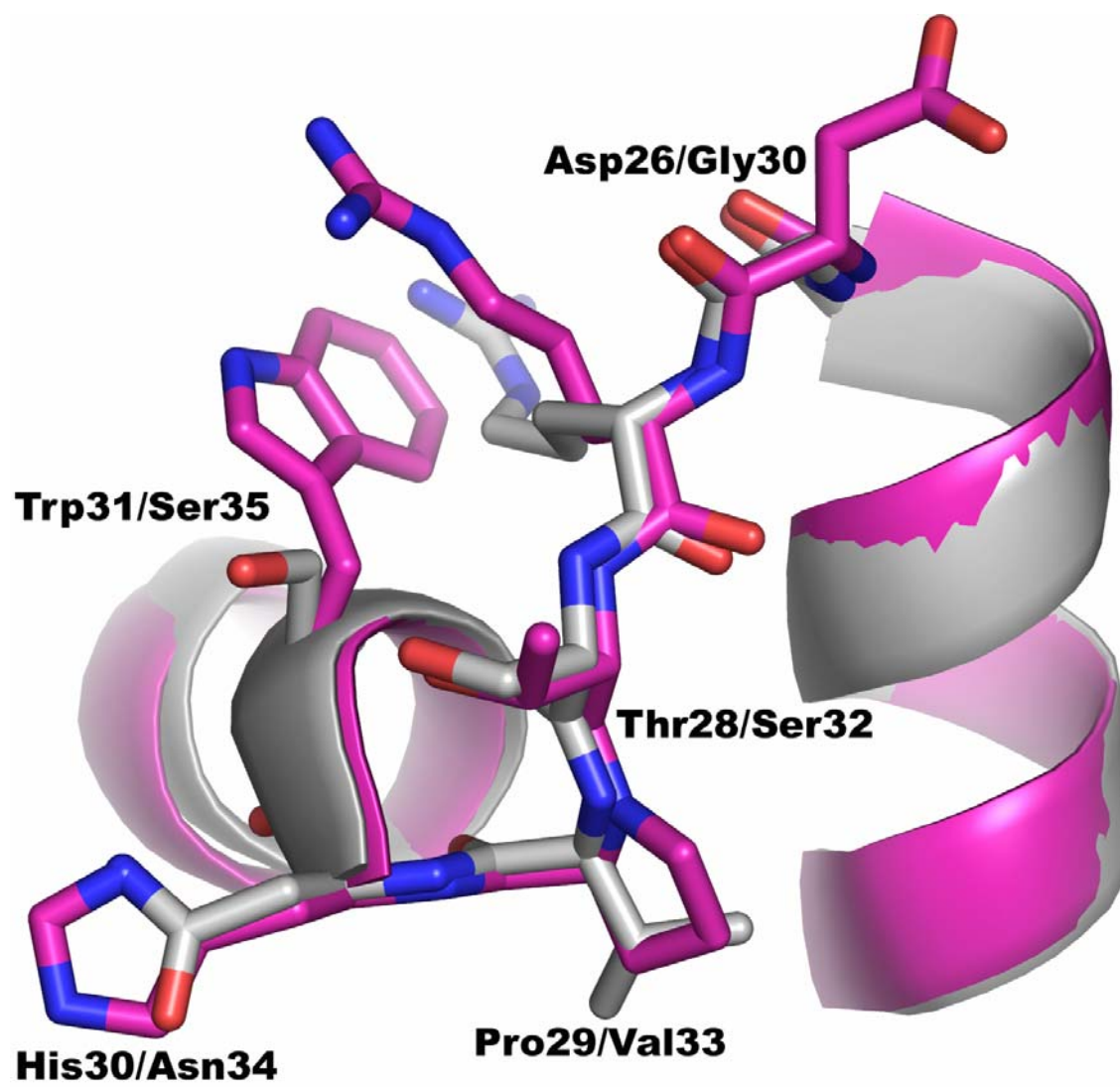
For example, there is usually a basic residue (Arg/Lys) in  $\beta 1$  that interacts directly with DNA bases through the major groove (Figure 3.2A). RHH proteins have been classified according to whether this positively charged residue occurs at the beginning (type I) or at the end (type II) of  $\beta 1$  (39). PutA52 has only one basic residue in  $\beta 1$  (Lys9). Lys9 is located at the end of  $\beta 1$  (Figure 3.1B), which indicates that PutA has a type II RHH fold. For reference, the Arc repressor is another example of a type II RHH protein (32, 40). Lys9 is highly solvent exposed, which is consistent with it playing a direct role in binding DNA. Interestingly, electron density for Lys9 in PutA52 was rather weak and this side chain is disordered in 3 of the 4 chains. Thus, DNA-binding induces structural ordering of Lys9.

RHH proteins also have hydrophobic residues conserved at specific positions in  $\beta 1$ ,  $\alpha A$  and  $\alpha B$  (34), as indicated in Figure 3.2A. All six positions are occupied by hydrophobic residues in PutA (Met6, Val8, Leu10, Ile18, Ile33, Ile37). The side chains of these residues are directed toward the interior of the protein where they contact other hydrophobic side chains, thus forming a tightly packed core. Note that these residues are highly conserved among PutAs (Figure 3.2B), which suggests all PutA RHH domains have the tightly packed hydrophobic core that is widely conserved in the RHH family.

**Figure 3.2.** *Amino acid sequence alignments of RHH and PutA proteins.* (A) Structure-based sequence alignment of the *E. coli* PutA DNA-binding domain (PutA52) with other RHH domains. The numbers above the alignment correspond to *E. coli* PutA. The numbers on the left indicate the starting residue for each protein sequence. Boxes indicate conserved basic residues of  $\beta 1$ , conserved glycine in the turn between  $\alpha A$  and  $\alpha B$  and the  $N_{cap}$  of  $\alpha B$ . Stars below the alignment denote residues of  $\beta 1$  that typically interact with DNA. Ovals below the alignment denote conserved hydrophobic residues of the RHH family. Triangles above the alignment indicate residues that we hypothesize are unique to the PutA RHH subfamily. PDB accession codes for the sequences shown in this alignment are as follows: Arc (1BDT), Omega (1IRQ), ParG (1P94), NikR (1Q5V), MetJ (1CMC) and CopG (2CPG). (B) Sequence alignment of 9 PutA DNA-binding domains. Residues identically conserved in all 9 sequences are indicated in black. Stars below the alignment denote residues of  $\beta 1$  that typically interact with DNA. Ovals below the alignment denote conserved hydrophobic residues of the RHH family.



**Figure 3.3.** *Comparison of the  $\alpha A$ - $\alpha B$  loop regions of PutA52 and Arc.* PutA52 is colored magenta and Arc is colored white. Residues are labeled as PutA52/Arc.



RHH proteins typically have a small, usually polar, residue at the N<sub>cap</sub> position of  $\alpha$ B (Figure 3.2A). The role of this residue in the Arc repressor (Ser32) has been extensively studied using mutagenesis (41). Only small, uncharged residues at this position (Ser, Cys, Ala, Thr, Pro, Asn) preserve both stability and strong operator binding, which suggests that such side chains strike a balance between the thermodynamic requirements of protein stability and DNA recognition (41). As in Arc, the N<sub>cap</sub> of  $\alpha$ B in PutA is a small polar residue (Thr28, Figure 3.2B). The hydroxyl of Thr28 forms a hydrogen bond with the backbone amine of Trp31 (Figure 3.3). This interaction is analogous to the hydrogen bond between Ser32 and the amine of Ser35 in Arc (Figure 3.3). In fact, the hydroxyl of PutA52 Thr28 perfectly overlaps the hydroxyl of Arc Ser32 after superposition of the two loops (Figure 3.3). . This type of N<sub>cap</sub> interaction is also observed in NikR (Asn27) (42), *Streptococcus pyogenes* omega protein (Asn50) (43) and CopG (Ser27) (34), among others. Note that Thr28 is identically conserved in PutAs, thus the corresponding N<sub>cap</sub> hydrogen bond is likely to be conserved as well.

#### *Sequence-Structure Relationships Unique to the PutA RHH Subfamily.*

Amino acid sequence alignments suggest that PutA RHH domain sequences have several conserved residues that are not typically found in other RHH proteins: Gly7, Arg15, Asp26, Pro29, and Trp31 (see Figure 3.2A, residues indicated by triangles). These PutA-specific residues could have important roles that enable the RHH domain to function within the unique context of the PutA polypeptide chain. Accordingly, these

residues may be considered as defining features of the PutA RHH subfamily. We analyzed the PutA52 structures to gain insights into the roles that these residues might play in PutA structure and function.

PutA RHH domains have Gly as the middle residue (Gly7) in the triad of residues in  $\beta 1$  that typically form hydrogen bonds to DNA bases (Figures 3.1B, 3.2A). Gly is an unusual choice for this position since it does not have a side chain capable of forming hydrogen bonds. Gly9 is highly conserved in PutAs (Figure 3.2B), but it is not restricted to PutAs since some NikR sequences also contain Gly at the analogous location (44). More typically, NikR contains Ser or Thr at this position.

Arg15 is conserved in PutAs, whereas hydrophobic residues typically occupy this position in other RHHs. In Arc, for example, the analogous residue is Leu19. The side chain of Arc Leu19 packs against the  $\beta$ -strand of the opposite subunit and contributes to the hydrophobic core. Arg15 of PutA adopts different conformations in the two crystal forms. In native PutA52, the density for Arg15 is rather weak but indicates that the side chain points out into solvent. In Se-Met PutA52K9M, Arg15 is more ordered and forms an intersubunit hydrogen bond with the side chain of Thr4. We note that intersubunit hydrogen bonding to this position on the  $\beta$ -sheet is not unprecedented in RHH domains. For example, NikR Gln2 forms analogous intersubunit hydrogen bonds to Asp 10 and Arg28. As discussed in the next section, Arg15 may play a role in binding the DNA backbone.

RHH proteins typically have a Gly in the turn that connects  $\alpha A$  and  $\alpha B$  (Figure 3.2A). This residue occupies the left-handed  $\alpha$ -helix region of the Ramachandran plot ( $\phi = \psi = 60^\circ$ ) and its unique backbone conformation is presumably important for



establishing the orientation of  $\alpha$ B relative to that of  $\alpha$ A. In PutA, Asp26 replaces the conserved glycine (Figure 3.2A). The carboxyl of Asp26 is directed toward solvent and does not interact with the protein (Figure 3.3). As can be seen from the superimposition of PutA52 and Arc (Figure 3.3), the  $\alpha$ A- $\alpha$ B loop conformation of PutA52 is nearly identical to that of Arc, despite substitution of Asp for Gly.

Accommodation of Asp26 in the  $\alpha$ A- $\alpha$ B is likely due to the fact the left-handed  $\alpha$ -helix region is allowed for Asp (45). In fact, only Gly and Asn are found more frequently than Asp in this region of the Ramachandran plot (45). The complexity of the Ramachandran plot for Asp is due to its role as a helix terminator (45), which is a role that Asp26 plays in PutA (Figure 3.3). We find that Asp26 in our structures occupies the left-handed  $\alpha$ -helix region of the Ramachandran plot, thus Asp26 fulfills the structural role that is more commonly played by the conserved Gly in RHH proteins.

Another unique aspect of PutA RHH domain sequences is that they have a conserved Pro at position 29 (Figure 3.2B). This residue occupies the N-terminal position of  $\alpha$ B (Figures 3.1B, 3.3). Although Pro is sometimes known as a helix breaker, it has, in fact, a high propensity for occurring at the  $\alpha$ -helix N-terminus ( $N_{\text{cap}} + 1$  position) as first discovered by Richardson and Richardson in 1988 (46) and later described in more detail by Kim and Kang (47). Pro29 does not disrupt the  $i$  to  $i+4$  hydrogen bonding of the  $\alpha$ B; the carbonyl of Pro29 forms a hydrogen bond to the amine of Ile33 (3.0 Å). Note that the main chain of Pro29 overlaps perfectly with that of Arc33 (Figure 3.3), showing that proline is accommodated at this position of the RHH fold. Thus, whereas the occurrence of Pro at the N-terminus of  $\alpha$ B is unusual among RHH proteins, Pro is well suited for this location in the RHH fold.

Finally, PutA RHH sequences are unique in that Trp appears at position 31. This residue is typically polar or charged in other RHH proteins (Figure 3.2A). For example, Ser35 of Arc forms a water-mediated hydrogen bond to the DNA backbone (reference 1BDT paper). It is possible that the indole nitrogen of PutA Trp31 plays a similar role, although this N atom does not superimpose well with the hydroxyl of Arc Ser35 (Figure 3.3). Alternatively, Trp31 could play a PutA-specific role. For example, since Trp31 is solvent exposed in PutA52, is possible that the aromatic ring of Trp31 mediates inter-domain interactions by packing against a hydrophobic patch of another PutA domain.

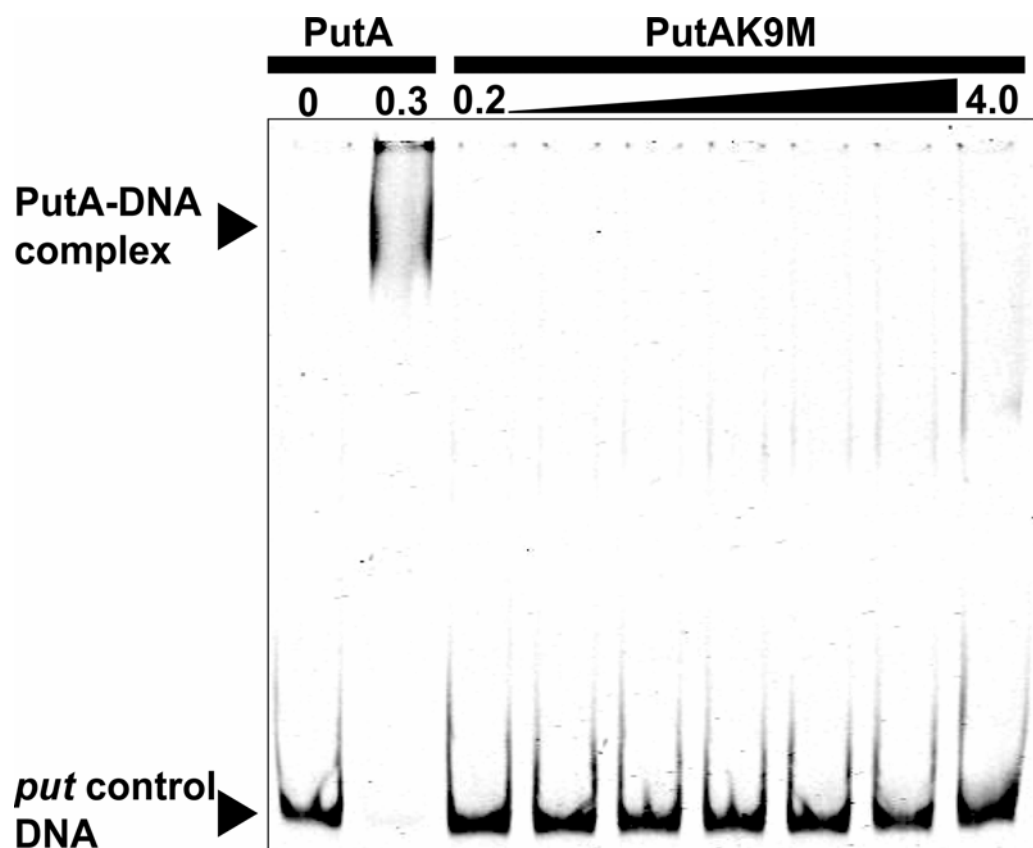
#### *Residues Involved In Binding DNA.*

Crystal structures of Arc (32), MetJ (33) and CopG (34) complexed with DNA show that the  $\beta$ -sheet mediates the majority of protein/DNA interactions. The RHH  $\beta$ -sheet inserts into the major groove of DNA and typically 2-3 residues from each strand form hydrogen bonds with DNA bases and in some cases, the phosphate backbone. These important residues are Gln9, Asn11, Arg13 in Arc, Lys23, Thr25, Ser27 in MetJ and Arg4, Thr6, Thr8 in CopG (Figure 3.2A). The analogous PutA residues are Thr5, Gly7 and Lys9. The side chains of Thr5 and Lys9 are directed toward the solvent in our structure (Figure 3.5), which is consistent with their putative role in binding DNA. The role of Gly7 is unclear at this time.

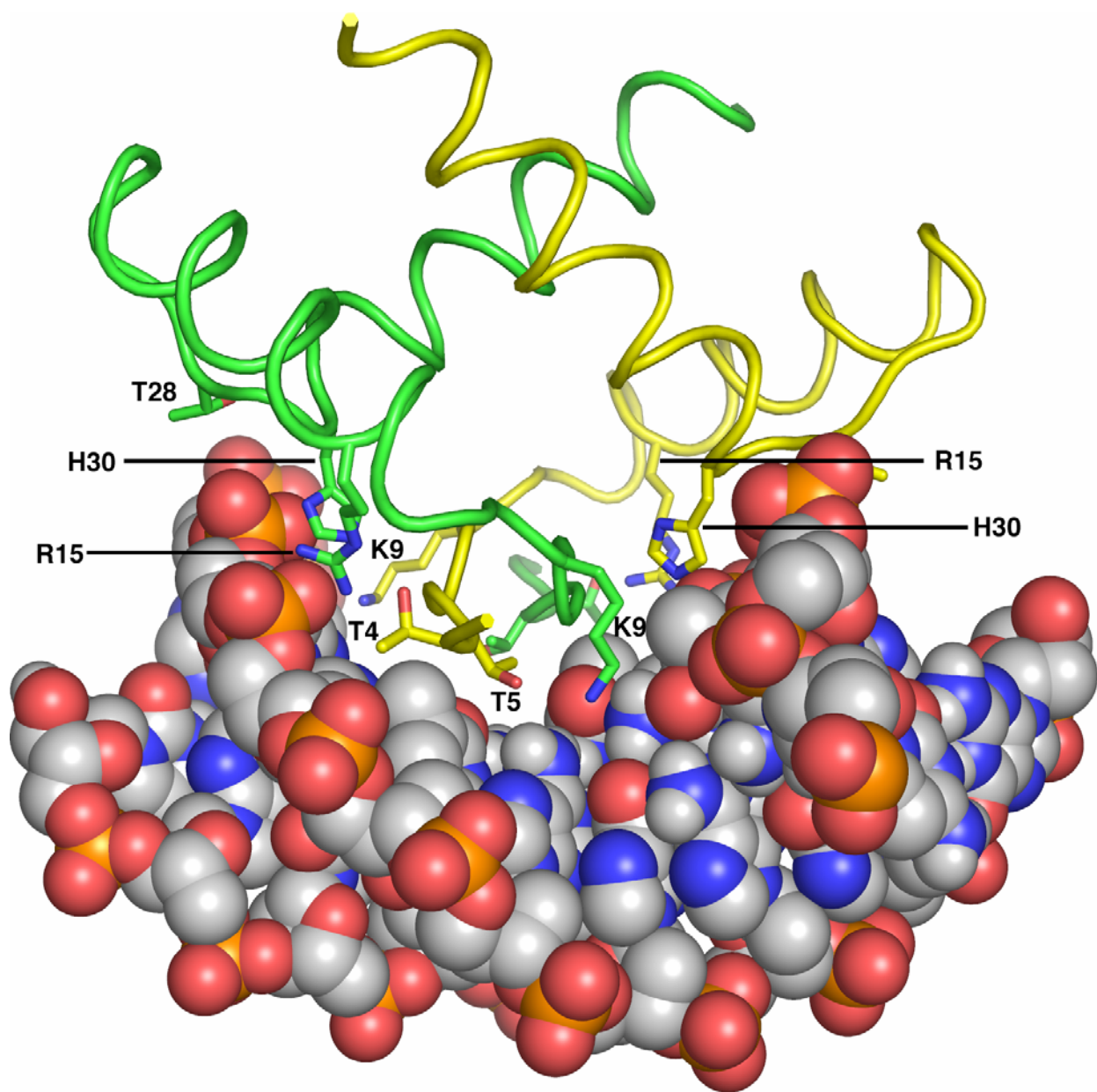
To test the importance of Lys9 for DNA recognition, we engineered and characterized the Lys9Met mutant of full-length *E. coli* PutA (PutAK9M). The biochemical properties of PutAK9M, such as enzyme activity, FAD absorbance spectrum and dimeric structure, were consistent with wild type PutA. However, the DNA binding

activity of PutAK9M was significantly disrupted. Gel-mobility shift assays failed to show binding of PutAK9M to *put* control DNA (5 nM) even at 800-fold excess of PutAK9M to DNA (Figure 3.4). These results clearly show that Lys9 is essential for DNA-binding in PutA.

**Figure 3.4.** *Gel mobility shift assay of wild type PutA and PutA mutant K9M.* Wild type PutA (0.3  $\mu$ M dimer) and PutAK9M (0.2-4  $\mu$ M dimer) were incubated with IRdye-700 labeled *put* control DNA (5 nM) for 20 min at 20° C in 50 mM Tris (pH 8.0) containing 10 % glycerol and 100  $\mu$ g/ml of nonspecific calf thymus DNA. Protein-DNA complexes were separated using a nondenaturing polyacrylamide gel (4 %).



**Figure 3.5.** *Model of the PutA RHH domain interacting with DNA.* Subunits of the PutA52 dimer are shown in green and yellow. Side chains are drawn for Thr4, Thr5, Lys9, Arg15, Thr28 and His30. DNA is shown in CPK mode. This model was built based on homology to the Arc repressor as described in Experimental Procedures.



Our homology models of PutA52/DNA show the ammonium group of Lys9 within 3.8 Å of polar atoms of DNA bases (Figure 3.5). In addition, the model based on the CopG structure has the hydroxyl of Thr5 within 3.0 Å of polar atoms of DNA bases. Thus, we predict that Thr5 and Lys9, which are identically conserved in PutAs, likely form direct hydrogen bonds with DNA bases.

Homology models of PutA52 complexed DNA also suggest that Arg15, Thr28 and His30 could interact with the phosphate backbone (Figure 3.5). All of our models have the side chain of His30 within 3.5 Å of a DNA phosphate group, while models based on Arc and MetJ show the Arg15 guanidinium within 3.4 Å of a phosphate group and the models based on Arc have the side chain of Thr28 3.3 Å from a phosphate group (Figure 3.5).

Arg15 was discussed in the previous section as being one of the PutA-specific residues. Thr28 is the N<sub>cap</sub> of αB, as discussed above. This residue is highly conserved among PutAs (Figure 3.2B) and it is structurally analogous to both the Arc Ser32 and MetJ Thr52. These residues form both direct (3.4 Å) and water-mediated electrostatic interactions with the phosphate backbone in the Arc repressor/operator complex (PDB code 1BDT) and the MetJ repressor/DNA complex (PDB code 1CMA).

His30 is an interesting candidate for making a direct interaction with DNA because it forms intersubunit hydrogen bonding interactions with β1. In the tetragonal crystal form (native PutA52), an imidazole N atom of His30 is within hydrogen bonding distance (2.9 Å) of the backbone N-H group of Lys9, implying that His30 is uncharged. On the other hand, in the monoclinic crystal form (Se-Met PutA52K9M), His30 forms a hydrogen bond with the carbonyl of Gly7. In either case, the hydrogen bond links αB to



$\beta$ 1 of the opposite subunit. This interaction is structurally analogous to the intersubunit hydrogen bond in Arc between Asn34 and the carbonyl of Arg13. We note that interaction with the DNA backbone is observed for both conformations of His30 according to our PutA52/DNA models.

## DISCUSSION

### *The PutA RHH Subfamily.*

Other members of the RHH family that have been structurally characterized include MetJ (33), Arc (32), CopG (34), NikR (42), ParG (39), and most recently the HP0222 protein from *Helicobacter pylori* (48). These proteins are much smaller than PutA, ranging from the 45-residue CopG repressor to the 133-residue NikR protein from *E. coli* with variations due to additional N-terminal or C-terminal domains. The C-terminal domains of different RHH members have been shown to have regulatory functions that enhance DNA-binding. In MetJ, binding of the co-repressor S-adenosylmethionine to the C-terminal domain regulates DNA-binding and repression of the *met* operon by MetJ. NikR is responsive to nickel availability via a C-terminal nickel-binding domain. In response to increasing nickel concentrations, NikR binds to promoter regions of the *nik* operon and represses expression of the nickel ABC transporter. Activation of NikR DNA binding has been proposed to involve nickel-dependent conformational changes that are propagated from the nickel regulatory domain to the RHH domain (42). PutA adds a unique sensory function to the RHH family with an appendage of a flavin redox regulatory domain to the RHH fold.

Comparison of the PutA52 structure to other RHH proteins shows that PutA shares many of the sequence-structure relationships that are commonly found in RHH proteins from different subfamilies. For example, PutA has a basic residue (Lys9) in  $\beta 1$  that is predicted to interact with DNA bases, PutA displays the conserved pattern of hydrophobic residues and there is a small polar residue at the  $\alpha B$  N<sub>cap</sub> position.

PutA also has unique features such as Gly7 serving as one of the residues of  $\beta 1$  that typically contact DNA bases, Arg15, replacement of the conserved glycine in the  $\alpha A$  -  $\alpha B$  turn by Asp26, a proline (Pro29) at the N-terminus of  $\alpha B$ , and Trp31. These unique features help define the PutA RHH subfamily and we suggest that they may play important roles in PutA function.

Gly7 is particularly interesting because polar side chains at this position in other RHH proteins typically interact with DNA bases. In the absence of a PutA/DNA crystal structure it is difficult to predict a role for Gly7 in DNA recognition. It is possible that Gly7 allows increased flexibility, which might be involved in the mechanism by which PutA switches between enzymatic and regulatory roles in response to cofactor reduction. The work presented here provides a foundation for mutagenesis studies that could be used to gain new insights into how these PutA-specific residues might be involved in proline-linked transcriptional regulation of proline utilization genes.

#### *Structural Basis of PutA Dimerization.*

PutAs studied to date purify as apparent homodimers, whether they are trifunctional repressor/enzymes like *E. coli* PutA (4) or bifunctional enzymes such as *Bradyrhizobium japonicum* PutA (49). A truncated form of *E. coli* PutA that starts at residue 86 (PutA $\Delta$ 85) purifies as an apparent monomer, which suggests that the N-terminal RHH domain is essential for dimerization of full-length *E. coli* PutA (18). We find that the PutA DNA-binding domain forms a classic RHH dimer and we assume that a similar dimeric substructure is formed in full-length *E. coli* PutA. The structure

reported here provides a basis for probing the importance of individual RHH residues in stabilizing the dimeric form of *E. coli* PutA.

PutA contains PRODH and P5CDH domains in addition to the RHH domain studied here. The structure of *E. coli* residues 86-669 is a  $\beta_8\alpha_8$  barrel and it forms an apparent monomer in solution. PutA constructs containing only the P5CDH domain have not yet been successfully studied due to solubility problems (J.J. Tanner and D.F. Becker, unpublished results). However, *Thermus thermophilus* P5CDH as well as other aldehyde dehydrogenases such as sheep liver aldehyde dehydrogenase (PDB entry 1BXS, (50)) and retinal dehydrogenase (PDB entry 1BI9, (51)) form dimers mediated by intermolecular  $\beta$ -sheets. Thus, it is possible that *E. coli* PutA dimerization involves both the RHH and P5CDH domains. In bifunctional PutA proteins that lack the RHH domain the dimeric structure may be stabilized entirely by the P5CDH domain. The questions of how the RHH domain is integrated into the architecture of trifunctional PutAs and whether *E. coli* PutA contains a second dimerization domain await structure determination of a full-length trifunctional PutA.

#### *Structural Basis of DNA Recognition.*

Our models of the PutA RHH domain complexed to DNA suggest that the RHH domain spans 5-7 base pairs via major groove binding. This result is qualitatively consistent with the discovery of several PutA binding sites containing the 6 base-pair sequence element GTTGCA in the *put* control DNA (15). Synthetic oligonucleotides of 21 base-pairs that contain GTTGCA bind specifically to PutA with an estimated  $K_d$  of  $\leq$

200 nM (15), which is only 4-fold higher than the  $K_d$  for PutA binding to the entire *put* control DNA ( $K_d = 45$  nM) (4).

We showed that mutation of Lys9 to Met eliminates binding of PutA to *put* control DNA, and our model predicts that Lys9 and Thr5 contact DNA bases directly. The model also predicts roles for Arg15, Thr28 and His30 in binding the DNA backbone (Figure 3.5). The importance of these residues will be tested using mutagenesis. Ultimately, the details of PutA/DNA recognition await structure determination of PutA proteins complexed with DNA.

## **ACKNOWLEDGMENT**

We thank Jay Nix and Darren Sherrell of ALS beamline 4.2.2 for help with data collection and processing. The ALS is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098 at Lawrence Berkeley National Laboratory.

## REFERENCES

1. Menzel, R., and Roth, J. (1981) *J Biol Chem* 256, 9762-6.
2. Brown, E. D., and Wood, J. M. (1993) *J Biol Chem* 268, 8972-9.
3. Surber, M. W., and Maloy, S. (1998) *Arch Biochem Biophys* 354, 281-7.
4. Becker, D. F., and Thomas, E. A. (2001) *Biochemistry* 40, 4714-21.
5. Vinod, M. P., Bellur, P., and Becker, D. F. (2002) *Biochemistry* 41, 6525-32.
6. Zhu, W., and Becker, D. F. (2003) *Biochemistry* 42, 5469-77.
7. White, T. A., and Tanner, J. J. (2005) *Acta Crystallograph Sect F Struct Biol Cryst Commun* 61, 737-9.
8. Phang, J. M. (1985) *Curr Top Cell Regul* 25, 91-132.
9. Menzel, R., and Roth, J. (1981) *J Mol Biol* 148, 21-44.
10. Brown, E. D., and Wood, J. M. (1992) *J Biol Chem* 267, 13086-92.
11. Wood, J. M. (1981) *J Bacteriol* 146, 895-901.
12. Maloy, S. R., and Roth, J. R. (1983) *J Bacteriol* 154, 561-8.
13. Ostrovsky de Spicer, P., O'Brien, K., and Maloy, S. (1991) *J Bacteriol* 173, 211-9.
14. Ostrovsky de Spicer, P., and Maloy, S. (1993) *Proc Natl Acad Sci U S A* 90, 4295-8.
15. Zhang, W., Zhou, Y., and Becker, D. F. (2004) *Biochemistry* 43, 13165-74.
16. Zhang, M., White, T. A., Schuermann, J. P., Baban, B. A., Becker, D. F., and Tanner, J. J. (2004) *Biochemistry* 43, 12539-48.
17. Lee, Y. H., Nadaraia, S., Gu, D., Becker, D. F., and Tanner, J. J. (2003) *Nat Struct Biol* 10, 109-14.
18. Gu, D., Zhou, Y., Kallhoff, V., Baban, B., Tanner, J. J., and Becker, D. F. (2004) *J Biol Chem* 279, 31171-6.
19. Zhu, W., Gincher, Y., Docherty, P., Spilling, C. D., and Becker, D. F. (2002) *Arch Biochem Biophys* 408, 131-6.
20. Doublie, S. (1997) *Methods Enzymol* 276, 523-30.
21. Matthews, B. W. (1968) *J Mol Biol* 33, 491-7.
22. Pflugrath, J. W. (1999) *Acta Crystallogr D Biol Crystallogr* 55, 1718-25.
23. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr D Biol Crystallogr* 55, 849-61.
24. Terwilliger, T. C. (2003) *Methods Enzymol* 374, 22-37.
25. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr D Biol Crystallogr* 60, 2126-32.
26. Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) *Acta Crystallogr D Biol Crystallogr* 57, 122-33.
27. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) *Nucleic Acids Res* 28, 235-42.
28. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C., and Read, R. J. (2005) *Acta Crystallogr D Biol Crystallogr* 61, 458-64.
29. DeLano, W. L. (2002), DeLano Scientific, San Carlos, CA.
30. Krissinel, E., and Henrick, K. (2004) *Acta Crystallogr D Biol Crystallogr* 60, 2256-68.
31. Shindyalov, I. N., and Bourne, P. E. (1998) *Protein Eng* 11, 739-47.

32. Raumann, B. E., Rould, M. A., Pabo, C. O., and Sauer, R. T. (1994) *Nature* 367, 754-7.
33. Somers, W. S., and Phillips, S. E. (1992) *Nature* 359, 387-93.
34. Gomis-Ruth, F. X., Sola, M., Acebo, P., Parraga, A., Guasch, A., Eritja, R., Gonzalez, A., Espinosa, M., del Solar, G., and Coll, M. (1998) *Embo J* 17, 7404-15.
35. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr D Biol Crystallogr* 54, 905-21.
36. Engh, R. A., and Huber, R. (1991) in *Acta Crystallogr A* pp 392-400.
37. Laskowski, R. A., Moss, D. S., and Thornton, J. M. (1993) *J Mol Biol* 231, 1049-67.
38. Chivers, P. T., and Sauer, R. T. (1999) *Protein Sci* 8, 2494-500.
39. Golovanov, A. P., Barilla, D., Golovanova, M., Hayes, F., and Lian, L. Y. (2003) *Mol Microbiol* 50, 1141-53.
40. Schildbach, J. F., Karzai, A. W., Raumann, B. E., and Sauer, R. T. (1999) *Proc Natl Acad Sci U S A* 96, 811-7.
41. Anderson, T. A., and Sauer, R. T. (2003) *Biophys Chem* 100, 341-50.
42. Schreiter, E. R., Sintchak, M. D., Guo, Y., Chivers, P. T., Sauer, R. T., and Drennan, C. L. (2003) *Nat Struct Biol* 10, 794-9.
43. Murayama, K., Orth, P., de la Hoz, A. B., Alonso, J. C., and Saenger, W. (2001) *J Mol Biol* 314, 789-96.
44. Chivers, P. T., and Tahirov, T. H. (2005) *J Mol Biol* 348, 597-607.
45. Hovmoller, S., Zhou, T., and Ohlson, T. (2002) *Acta Crystallogr D Biol Crystallogr* 58, 768-76.
46. Richardson, J. S., and Richardson, D. C. (1988) *Science* 240, 1648-52.
47. Kim, M. K., and Kang, Y. K. (1999) *Protein Sci* 8, 1492-9.
48. Popescu, A., Karpay, A., Israel, D. A., Peek, R. M., Jr., and Krezel, A. M. (2005) *Proteins* 59, 303-11.
49. Krishnan, N., and Becker, D. F. (2005) *Biochemistry* 44, 9130-9.
50. Moore, S. A., Baker, H. M., Blythe, T. J., Kitson, K. E., Kitson, T. M., and Baker, E. N. (1998) *Structure* 6, 1541-51.
51. Lamb, A. L., and Newcomer, M. E. (1999) *Biochemistry* 38, 6003-11.



## VITA

Jermaine Lemont Jenkins was born June 9, 1970 in Easton, Maryland. He graduated from Queen Anne's High School in Centreville, Maryland. He enlisted in the United States Army where he was stationed at Fort Carson, Colorado. He received an Honorable Discharge along with several service medals of accommodation after completing his 3-year military duty obligation.

He earned a Bachelor's of Science in Biology from Salisbury University in Salisbury, Maryland. He obtained a Ph.D. in Biochemistry in 2006 from the Department of Biochemistry. He accepted a post-doctoral fellowship in Cancer Biology from the National Cancer Institute and will work in the lab of Dr. Clara Kielkopf at Johns Hopkins University in Baltimore, Maryland.