

MOLECULAR MECHANISMS OF ENZYME DYSFUNCTION IN HUMAN PHOSPHOGLUCOMUTASE-1 DEFICIENCY

A Dissertation

Presented to

the Faculty of the Graduate School

at the University of Missouri – Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Kyle M. Stiers

and Dr. Lesa J. Beamer

Dissertation Supervisor

May 2019

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

**MOLECULAR MECHANISMS OF ENZYME DYSFUNCTION
IN HUMAN PHOSPHOGLUCOMUTASE-1 DEFICIENCY**

presented by Kyle M. Stiers,

a candidate for the degree of doctor of philosophy,

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

Dr. Lesa J Beamer

Dr. John J. Tanner

Dr. Steven R. Van Doren

Dr. Dong Xu

To Lori and my son...

ACKNOWLEDGEMENTS

I have conducted my graduate research under the excellence guidance of Dr. Lesa J. Beamer. I would like to sincerely thank her for providing a wonderful environment, research opportunities, and mentorship over the last five years. Dr. Beamer has been truly helpful in helping me grow as a person and scientist. I would also like to especially thank her for tolerating my many random computational side projects that rarely led to useful outcomes (but sometimes did).

I would like to give special thanks to Dr. Jack Tanner and Dr. Steve Van Doren who have both provided interesting research collaborations outside of my own project. Furthermore, they have both mentored me a great deal and always provided insight into various areas of research.

I would like to thank Dr. David Korasick for being a uniquely great resource during my graduate work. He taught me a lot, was always happy to lament over bad results with me, and generally to discuss anything and everything.

Many of my cohorts, senior students, and friends also deserve credit for helping me along the way and providing encouragement including, Stuart Roderick, Kwaku Tawiah, Austin Prater, Ashley Campbell, Tyler Kottmann, Max Schaefer, Alex Brechbuhler, Reed Hansen, Mark Schroeder and many more.

The vast majority of my success has been in large part due to the hard work of undergraduates that I was fortunate enough to work with and mentor. A

huge thanks to Bailee Kain, Abby Graham, Andrew Muenks, Zach Addison, and Bana Daghlas.

I would also like to formally thank Dr. Chris Lee for his friendship and mentoring in the various roles we knew each other. He was always a source of reassurance as well as a great scientist and teacher.

Finally, I would like to thank my family and my wife, Lori. They were always willing to listen about experiments they did not understand in an attempt to learn and were frequently my test audience for writing lay descriptions and presentations.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	ii
LIST OF FIGURES.....	ix
LIST OF TABLES.....	xi
LIST OF ABBREVIATIONS.....	xii
ABSTRACT.....	xiii
Chapter	
1. INTRODUCTION.....	1
1.1 Phosphohexomutase superfamily and human PGM1 – 1	
1.2 Human phosphoglucomutase-1 deficiency – 2	
1.3 The molecular basis of enzyme dysfunction in PGM1 deficiency – 3	
1.4 References – 6	
2. INDUCED STRUCTURAL DISORDER AS A MOLECULAR MECHANISM FOR ENZYME DYSFUNCTION IN PHOSPHOGLUCOMUTASE 1 DEFICIENCY.....	9
II.1 Abstract – 9	
II.2 Introduction – 10	
II.3 Results – 13	
II.3.1 Overview of the wild-type structure – 13	
II.3.2 Context of glycines 121 and 291 in the WT structure – 18	
II.3.3 Structure of the G121R missense variant – 18	
II.3.4 Structure of the G291R missense variant – 21	
II.3.5 Correlation between structural and biochemical data – 24	
II.3.6 Missense variants of PGM1 cluster in domain-domain interfaces – 26	

II.4 Discussion – 29	
II.5 Materials and Methods – 34	
II.5.1 Protein expression, purification, and crystallization – 34	
II.5.2 X-ray Diffraction Data Collection and Refinement – 35	
II.6 Acknowledgements – 37	
II.7 References – 38	
II.8 Supporting Material – 43	
II.8.1 Supporting Figures - 43	
3. ASP263 MISSENSE VARIANTS PERTURB THE ACTIVE SITE OF HUMAN PHOSPHOGLUCOMUTASE 1 (PGM1)	44
III.1 Abstract – 44	
III.2 Introduction – 45	
III.3 Results – 47	
III.3.1 Asp263 participates in a conserved salt bridge in WT PGM1 – 47	
III.3.2 Structural impact of the Asp263 mutants – 51	
III.3.3 Arg293 co-opts key interactions needed for binding substrate – 53	
III.3.4 Correlation with previous biochemical data – 54	
III.3.5 Reduction in spontaneous phosphoryl transfer in missense variants – 56	
III.3.6 Arg293 is required for correct protein folding – 57	
III.4 Discussion – 59	
III.5 Experimental Procedures – 61	
III.5.1 Mutagenesis and protein expression – 61	
III.5.2 Protein purification and crystallization – 62	
III.5.3 X-ray diffraction data collection and refinement – 63	

III.5.4	Dynamic light scattering	– 64
III.5.5	Preparation and assessment of phospho-enzyme	– 64
III.6	Author Contributions	– 65
III.7	Acknowledgements	– 65
III.8	References	– 66
4.	A HOTSPOT FOR DISEASE-ASSOCIATED VARIANTS OF HUMAN PGM1 IS ASSOCIATED WITH IMPAIRED LIGAND BINDING AND LOOP DYNAMICS.....	70
IV.1	Summary	– 70
IV.2	Introduction	– 70
IV.3	Results	– 74
IV.3.1	Biochemical studies of variants in the D4 loop	– 74
IV.3.2	Arg503 and Arg515 variants cause unexpected ordering of the D4 loop	– 77
IV.3.3	The D4 loop changes conformation to bind ligand	– 79
IV.3.4	Evaluation of loop mobility by molecular dynamics	– 82
IV.4	Discussion	– 84
IV.5	Acknowledgements	– 87
IV.6	Author Contributions	– 88
IV.7	Declaration of Interests	– 88
IV.8	References	– 89
IV.9	Star Methods	– 95
IV.9.1	Key Resources Table	– 95
IV.9.2	Contact for Reagent and Resource Sharing	– 96
IV.9.3	Experimental Model and Subject Details	– 96
IV.10	Methods Details	– 96

IV.10.1 Mutagenesis, protein expression and purification – 96	
IV.10.2 Biochemical studies – 97	
IV.10.3 Crystallization – 98	
IV.10.4 X-ray diffraction data collection and refinement – 100	
IV.10.5 Molecular dynamics simulations and trajectory analyses – 101	
IV.11 Quantification and Statistical Analysis – 102	
IV.12 Data and Software Availability – 102	
IV.12.1 Accession Numbers – 102	
IV.12.2 Additional Resources – 103	
IV.13 Supplemental Information – 104	
IV.13.1 Supporting Figures – 104	
5. PRINCIPAL COMPONENT ANALYSIS AS A METHOD FOR UNDERSTANDING ENSEMBLE STRUCTURAL CHANGES IN PROTEINS	108
V.1 Abstract – 108	
V.2 Overview of PCA – 108	
V.3 Applying PCA to structural ensembles – 109	
V.3.1 PCA to cluster a structural ensemble into related groups – 112	
V.3.2 PCA to examine residue-specific correlations – 115	
V.3.3 Potential applications of coordinate analysis by PCA – 117	
V.4 Conclusion – 118	
V.5 References – 119	
V.6 Supplemental Material – 120	
V.6.1 PCA script using Bio3D in R – 120	

APPENDIX

1. COMPROMISED CATALYSIS AND POTENTIAL FOLDING DEFECTS IN IN VITRO STUDIES OF MISSENSE MUTANTS ASSOCIATED WITH HEREDITARY PHOSPHOGLUCOMUTASE 1 DEFICIENCY	123
2. SYNCHROTRON-BASED MACROMOLECULAR CRYSTALLOGRAPHY MODULE FOR AN UNDERGRADUATE BIOCHEMISTRY LABORATORY COURSE.....	124
3. BIOLOGY, MECHANISM, AND STRUCTURE OF ENZYMES IN THE A-D- PHOSPHOHEXOMUTASE SUPERFAMILY	125
4. STRUCTURE AND CHARACTERIZATION OF A CLASS 3B PROLINE UTILIZATION A: LIGAND-INDUCED DIMERIZATION AND IMPORTANCE OF THE C-TERMINAL DOMAIN FOR CATALYSIS..	126
5. SEQUENCE-STRUCTURE RELATIONSHIPS, EXPRESSION PROFILES, AND DISEASE-ASSOCIATED MUTATIONS IN THE PARALOGS OF PHOSPHOGLUCOMUTASE 1	127
6. PHOSPHORYLATION-DEPENDENT EFFECTS ON THE STRUCTURAL FLEXIBILITY OF PHOSPHOGLUCOSAMINE MUTASE FROM BACILLUS ANTHRACIS.....	128
7. ASSESSMENT AND IMPACTS OF PHOSPHORYLATION ON PROTEIN FLEXIBILITY OF THE A-D-PHOSPHOHEXOMUTASES	129
8. SYNTHESIS, DERIVATIZATION AND STRUCTURAL ANALYSIS OF PHOSPHORYLATED MONO-, DI- AND TRI-FLUORINATED D- GLUCO-HEPTULOSES BY GLUCOKINASE: TUNABLE PHOSPHOGLUCOMUTASE INHIBITION	130
9. INHIBITORY EVALUATION OF APMM/PGM FROM PSEUDOMONAS AERUGINOSA: CHEMICAL SYNTHESIS, ENZYME KINETIC AND PROTEIN CRYSTALLOGRAPHY STUDY	131
10. GALACTOSE RESTORES GLYCOSYLATION THROUGH REPLENISHING ACTIVATED SUGARS IN PGM1-CDG	132
VITA	133

LIST OF FIGURES

Chapter I

FIGURE I-1 - Generic reaction to the α -D-phosphohexomutase superfamily 2

Chapter II

FIGURE II-1 - A schematic of the catalytic reaction of PGM1 11

FIGURE II-2 - The crystal structure of wild-type human PGM1 16

FIGURE II-3 - Overall structure and detailed views of the G121R crystal structure 19

FIGURE II-4 - Overall and detailed views of the G291R crystal structure 23

FIGURE II-5 - Relationship of the known PGM1 missense variants to domain interfaces of the enzyme 26

FIGURE II-S1 - A listing of PGM1 residues that participate in interdomain interfaces in the WT enzyme 43

Chapter III

FIGURE III-1 - Overview of the mechanism and structure of human PGM1 46

FIGURE III-2 - The context of residue 263 in the structure of WT PGM1 48

FIGURE III-3 - Structural context of residue 263 in the missense variants of PGM1 50

FIGURE III-4 - Biochemical data on phosphorylation and protein expression/solubility 55

Chapter IV

FIGURE IV-1 - Overview of the mechanism and structure of human PGM1 72

FIGURE IV-2 - Crystal structures of missense variants of Arg503 and Arg515.. 78

FIGURE IV-3 - Crystal structure of the PGM1 complex with glucose 6-phosphate 80

FIGURE IV-4 - Analyses of D4 loop flexibility based on the MD simulations of human PGM1 83

FIGURE IV-5 - Sequences of the D4 loop of phosphoglucomutase in diverse eukaryotic organisms.....	85
FIGURE IV-S1 - Representative electrospray ionization mass spectra for the R503Q missense variant, related to Methods.....	104
FIGURE IV-S2 - Electron density maps in the vicinity of the D4 loop of PGM1, related to Figure IV-2 and IV-3	105
FIGURE IV-S3 - Electron density maps of the D4 loop and active site vicinity of rabbit PGM, related to Figure IV-3.....	106
FIGURE IV-S4 Analysis from 10 ns MD simulations of WT PGM1 and the R503Q missense variant, related to Figure IV-4	107
Chapter V	
FIGURE V-1 - Outline of the catalytic cycle of XcPGM, highlighting the various enzyme states captured in this study.....	112
FIGURE V-2 - A scatter plot displaying scores of the various XcPGM structures for the first three principal components of the data set.....	114
FIGURE V-3 PCA of an apo-enzyme ensemble of XcPGM comprised of two cryo and two RT structures.....	116

LIST OF TABLES

Chapter II

TABLE II-1 Data collection and refinement statistics 15

TABLE II-2 Disordered residues and kinetic parameters for the G121R and G291R missense variants relative to WT PGM1 21

Chapter III

TABLE III-1 Data collection and refinement statistics for the Asp263 missense variants 52

Chapter IV

TABLE IV-1 Summary of biochemical data on PGM1 loop mutants 74

TABLE IV-2 Crystallographic data collection and refinement statistics 76

TABLE IV-9.1 Key Resources Table 95

Chapter V

TABLE V-1 Overview of X-ray data sets collected 113

LIST OF ABBREVIATIONS

CAPS: N-cyclohexyl-3-aminopropanesulfonic acid

CHES: N-Cyclohexyl-2-aminoethanesulfonic acid

DTT: dithiothreitol

EDTA: ethylenediaminetetraacetic acid

MOPS: 3-morpholinopropane-1-sulfonic acid

MES: 2-(N-morpholino)ethanesulfonic acid

TRIS: 2-Amino-2-hydroxymethyl-propane-1,3-diol

PDB: Protein Data Bank

PGM1: phosphoglucomutase 1

WT: wild-type

P-Ser: phosphoserine

ESI-MS: electrospray ionization mass spectrometry

DLS: dynamic light scattering

MW: molecular weight

R_h: Hydrodynamic radius

ABSTRACT

Human phosphoglucomutase-1 (PGM1) belongs to the α -D-phosphohexomutase superfamily, an ancient lineage of enzymes critical for carbohydrate metabolism. PGM1 catalyzes the interconversion of glucose-1-phosphate and glucose-6-phosphate, acting as the pivot between energy storage and utilization. Recently, PGM1 has been implicated as the monogenic cause of an inherited metabolic disease in humans, called PGM1 deficiency. The disease presents with highly variable phenotype in patients and is difficult to diagnose. Furthermore, genotype-phenotype relationships remain unclear -even in siblings with the same missense variants, no obvious correlation exists. PGM1 deficiency is a unique research opportunity due to the lack of clear rationale for varying effects of missense variants, availability of patient data, favorable *in vitro* behavior of recombinantly expressed PGM1, and the limited number of structural studies characterizing individual missense variants. In this work we have characterized multiple molecular mechanisms of disease through X-ray crystallography and biochemistry. Thus, this work provides a foundation for physicians to make much more accurate prognostic decisions when advising patients, identifies variants with possible therapeutic interventions, and informs us of key dynamics and structural features required for proper functioning of human PGM1.

Chapter I

Introduction

I.1 Phosphohexomutase superfamily and human PGM1

The α -D-phosphohexomutase (PHM) superfamily is a ubiquitous group of enzymes found in every kingdom of life. These enzymes function in many cellular processes including carbohydrate metabolism, peptidoglycan synthesis, sugar nucleotide biosynthesis, as well as glycolipid and exopolysaccharide synthesis. Despite a long evolutionary history and highly variable sequences, all enzymes in the superfamily appear to share a conserved four domain architecture and similar active sites, located at the confluence of these four domains [1,2].

Enzymes in the PHM superfamily catalyze the reversible conversion of phosphosugar substrates, typically in the cytoplasm of the cell [2]. The reaction is an intramolecular phosphoryl transfer that proceeds through a bisphosphorylated intermediate (see detailed mechanism in subsequent Chapters). The conversion is normally between the 1- and 6-phospho forms of a hexose, such as glucose-1-phosphate and glucose-6-phosphate (Figure I-1). Catalytically active enzyme requires a phosphorylated serine, and a divalent metal ion such as Mg^{2+} . To complete the catalytic cycle, the bisphosphorylated intermediate undergoes an unusual 180° rotation within the active site in between phosphoryl transfer steps [2].

The human genome contains five distinct PHM genes, that differ in their substrate specificity and distribution in tissues. Of these, phosphoglucomutase-1 (PGM1) is the most widely distributed and is responsible for the vast majority of phosphoglucomutase activity in human cells [2,3]. PGM1 is well known due to its function at the critical pivot point between energy storage and mobilization in the form of glycogen. Its role is particularly critical in tissues with high metabolic requirements, such as heart and skeletal muscle.



Figure I-1. Generic reaction of the α -D-phosphohexomutase superfamily. Conversion of a hexose-1-phosphate to hexose-6-phosphate proceeds through a bisphosphorylated intermediate.

I.2 Human phosphoglucomutase-1 deficiency

In 2014, mutations in human PGM1 were identified as the cause of an inherited metabolic disease, PGM1 deficiency [4]. PGM1 deficiency is autosomal recessive in inheritance, and associated with various types of mutations, including frame shifts, aberrant splicing, and missense variants. In the latter category, 13 different mutations were originally identified, affecting 12 residue positions in the protein [4,5]. Since the original study, 9 additional missense variants have been associated with this inherited condition [6–8]. PGM1 deficiency is categorized as both a glycogen storage disease (GSD) and a congenital disorder of glycosylation, due to downstream effects on protein N-glycosylation [4].

Because of the function of PGM1 in glucose metabolism, and its ubiquitous expression in tissues throughout the body, widespread effects from PGM1 mutations are expected [3]. Indeed, patients with PGM1 deficiency have symptoms including dilated cardiomyopathy, exercise intolerance, hepatopathy, hypoglycemia, muscle weakness, and developmental abnormalities such as delayed puberty and cleft palate. However, patient phenotypes are highly variable, even in siblings with the same underlying mutations. The highly variable nature of patient phenotypes makes diagnosing PGM1 deficiency difficult without full exome sequencing as patients may show substantial overlap in symptoms with McArdle disease and other GSDs [4,6,9,10].

I.3 The molecular basis of enzyme dysfunction in PGM1 deficiency

Characterizing the biochemical and structural effects of missense variants of PGM1 has potential implications for individuals affected by this inherited disease. First, such studies can provide a baseline for correlating patient genotype and clinical phenotype, as clinical studies have been done using samples from patients with heterozygous genotypes. Second, knowledge of the biochemical phenotypes of the missense mutants may help target appropriate therapies to patients with PGM1 deficiency. For example, as shown for some other inherited enzyme deficiencies, therapeutic approaches that increase protein folding/stability may be of benefit for PGM1 variants with folding defects [11–18]. Finally, categorization of the underlying biochemical defects may improve predictions of patient prognosis, which range from early childhood death

to adults affected mostly with exercise intolerance. Knowledge of the most detrimental mutations could help target patients for aggressive supportive treatments at an early age [19,20].

In addition to possible benefit for patients, PGM1 has the potential to serve as model system for studying the impacts of missense variant on protein structure and function. Studies of individual disease-related mutants are often complicated by detrimental effects of the mutation on protein expression or stability [16–18]. Also, characterizing crystal structures of missense variants can be time-consuming and frustrating, if suitable crystals cannot be obtained. Because of such difficulties, it is typical for the effects of missense variants to be predicted using computational methods that rely on sequence conservation of the mutated position, the nature of the physicochemical change, calculated steric conflicts from homology modeling, the use of very short molecular dynamics simulations, or some combination of these approaches [21–24]. However, the predictions of these methods are typically not verified experimentally, making their utility questionable.

In this thesis, I present detailed studies of multiple missense variants of PGM1 deficiency, which have proven amenable to detailed biochemical and structural studies. The protein is well expressed from a recombinant system, and many missense variants are soluble and well behaved in solution [25]. Importantly, many of them also crystallize allowing detailed analysis of their structural impacts. Initial biochemical characterization classified missense variants as either catalytic or folding mutants [25]. However, structural

information has deprecated this classification because even soluble, well-behaved PGM1 mutants vary in their impacts, which include induced structural disorder, co-opting of key catalytic residues in the active site, and disruption of loop dynamics necessary for ligand binding and catalysis [26–28]. Thus, PGM1 is a rare, tractable system where many mutants at positions both proximal and distal from the active site can be investigated on a biochemical, structural, and dynamics scale.

In addition to the biochemical and structural studies of PGM1, I include a chapter on the use of principal component analysis of protein coordinates. This one of several computational approaches I developed during the course of my graduate studies.

I.4 References

- [1] D.B. Whitehouse, J. Tomkins, J.U. Lovegrove, D.A. Hopkinson, W.O. McMillan, A phylogenetic approach to the identification of phosphoglucomutase genes, *Mol. Biol. Evol.* 15 (1998) 456–462. doi:10.1093/oxfordjournals.molbev.a025942.
- [2] K.M. Stiers, A.G. Muenks, L.J. Beamer, *Biology, Mechanism, and Structure of Enzymes in the α -D-Phosphohexomutase Superfamily*, 2017. doi:10.1016/bs.apcsb.2017.04.005.
- [3] A.G. Muenks, K.M. Stiers, L.J. Beamer, Sequence-structure relationships , expression profiles , and disease-associated mutations in the paralogs of phosphoglucomutase 1, *1* (2017) 1–21. doi:https://doi.org/10.1371/journal.pone.0183563.
- [4] L.C. Tegtmeyer, S. Rust, M. van Scherpenzeel, B.G. Ng, M.-E. Losfeld, S. Timal, K. Raymond, P. He, M. Ichikawa, J. Veltman, K. Huijben, Y.S. Shin, V. Sharma, M. Adamowicz, M. Lammens, J. Reunert, A. Witten, E. Schrapers, G. Matthijs, J. Jaeken, D. Rymen, T. Stojkovic, P. Laforêt, F. Petit, O. Aumaître, E. Czarnowska, M. Piraud, T. Podskarbi, C.A. Stanley, R. Matalon, P. Burda, S. Seyyedi, V. Debus, P. Socha, J. Sykut-Cegielska, F. van Spronsen, L. de Meirleir, P. Vajro, T. DeClue, C. Ficicioglu, Y. Wada, R.A. Wevers, D. Vanderschaeghe, N. Callewaert, R. Fingerhut, E. van Schaffingen, H.H. Freeze, E. Morava, D.J. Lefeber, T. Marquardt, Multiple phenotypes in phosphoglucomutase 1 deficiency., *N. Engl. J. Med.* 370 (2014) 533–42. doi:10.1056/NEJMoa1206605.
- [5] L.J. Beamer, Mutations in hereditary phosphoglucomutase 1 deficiency map to key regions of enzyme structure and function, *J. Inherit. Metab. Dis.* 38 (2015) 243–256. doi:10.1007/s10545-014-9757-9.
- [6] S.Y.-W. Wong, L.J. Beamer, T. Gadomski, T. Honzik, M. Mohamed, S.B. Wortmann, K.S. Brocke Holmefjord, M. Mork, F. Bowling, J. Sykut-Cegielska, D. Koch, A. Ackermann, C.A. Stanley, D. Rymen, A. Zeharia, M. Al-Sayed, T. Marquardt, J. Jaeken, D. Lefeber, D.F. Conrad, T. Kozicz, E. Morava, Defining the Phenotype and Assessing Severity in Phosphoglucomutase-1 Deficiency, *J. Pediatr.* (2016). doi:10.1016/j.jpeds.2016.04.021.
- [7] N. Loewenthal, A. Haim, R. Parvari, E. Hershkovitz, Phosphoglucomutase-1 deficiency: Intrafamilial clinical variability and common secondary adrenal insufficiency, *Am. J. Med. Genet. Part A.* 167 (2015) 3139–3143. doi:10.1002/ajmg.a.37294.
- [8] Y. Ding, N. Li, G. Chang, J. Li, R. Yao, Y. Shen, J. Wang, X. Huang, X. Wang, Clinical and molecular genetic characterization of two patients with mutations in the phosphoglucomutase 1 (PGM1) gene, *J. Pediatr. Endocrinol. Metab.* 31 (2018) 781–788. doi:10.1515/jpem-2017-0551.

- [9] E. Morava, S. Wong, D. Lefeber, Disease severity and clinical outcome in phosphoglucomutase deficiency, *J. Inherit. Metab. Dis.* 38 (2015) 207–209. doi:10.1007/s10545-014-9769-5.
- [10] N. Preisler, P. Laforêt, A. Echaniz-Laguna, M.C. Ørngreen, E. Lonsdorfer-Wolf, S. Doutreleau, B. Geny, T. Stojkovic, M. Piraud, F.M. Petit, J. Vissing, Fat and Carbohydrate Metabolism During Exercise in Phosphoglucomutase Type 1 Deficiency, *J. Clin. Endocrinol. Metab.* 98 (2013) E1235–E1240. doi:10.1210/jc.2013-1651.
- [11] J.-M. Blouin, Y. Duchartre, P. Costet, M. Lalanne, C. Ged, A. Lain, O. Millet, H. de Verneuil, E. Richard, Therapeutic potential of proteasome inhibitors in congenital erythropoietic porphyria., *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 18238–43. doi:10.1073/pnas.1314177110.
- [12] L.C. Hernández-Ramírez, F. Martucci, R.M.L. Morgan, G. Trivellin, D. Tilley, N. Ramos-Guajardo, D. Iacovazzo, F. D’Acquisto, C. Prodromou, M. Korbonits, Rapid Proteasomal Degradation of Mutant Proteins Is the Primary Mechanism Leading to Tumorigenesis in Patients With Missense *AIP* Mutations, *J. Clin. Endocrinol. Metab.* 101 (2016) 3144–3154. doi:10.1210/jc.2016-1307.
- [13] S. Zampieri, B. Bembi, N. Rosso, M. Filocamo, A. Dardis, Treatment of Human Fibroblasts Carrying NPC1 Missense Mutations with MG132 Leads to an Improvement of Intracellular Cholesterol Trafficking, in: Springer, Berlin, Heidelberg, 2011: pp. 59–69. doi:10.1007/8904_2011_49.
- [14] R.L. Lieberman, B. a Wustman, P. Huertas, A.C. Powe, C.W. Pine, R. Khanna, M.G. Schlossmacher, D. Ringe, G. a Petsko, Structure of acid beta-glucosidase with pharmacological chaperone provides insight into Gaucher disease., *Nat. Chem. Biol.* 3 (2007) 101–107. doi:10.1038/nchembio850.
- [15] P. Compain, O.R. Martin, C. Boucheron, G. Godin, L. Yu, K. Ikeda, N. Asano, Design and synthesis of highly potent and selective pharmacological chaperones for the treatment of Gaucher’s disease, *ChemBioChem.* 7 (2006) 1356–1359. doi:10.1002/cbic.200600217.
- [16] S.W. Gersting, K.F. Kemter, M. Staudigl, D.D. Messing, M.K. Danecka, F.B. Lagler, C.P. Sommerhoff, A.A. Roscher, A.C. Muntau, Loss of Function in Phenylketonuria Is Caused by Impaired Molecular Motions and Conformational Instability, *Am. J. Hum. Genet.* 83 (2008) 5–17. doi:10.1016/J.AJHG.2008.05.013.
- [17] J.A. McKinney, B. Turel, I. Winge, P.M. Knappskog, J. Haavik, Functional properties of missense variants of human tryptophan hydroxylase 2, *Hum. Mutat.* 30 (2009) 787–794. doi:10.1002/humu.20956.
- [18] M. Cerreto, P. Cavaliere, C. Carluccio, F. Amato, A. Zagari, A. Daniele, F. Salvatore, Natural phenylalanine hydroxylase variants that confer a mild

phenotype affect the enzyme's conformational stability and oligomerization equilibrium, *Biochim. Biophys. Acta - Mol. Basis Dis.* 1812 (2011) 1435–1445. doi:10.1016/J.BBADIS.2011.07.012.

- [19] S.Y. Wong, T. Gadowski, M. Van Scherpenzeel, T. Honzik, H. Hansikova, K.S.B. Holmefjord, M. Mork, F. Bowling, J. Sykut-cegielska, D. Koch, J. Hertecant, G. Preston, J. Jaeken, N. Peeters, S. Perez, D. Do Nguyen, K. Crivelly, T. Emmerzaal, K.M. Gibson, K. Raymond, N.A. Bakar, F. Foulquier, G. Poschet, A.M. Ackermann, M. He, Oral D -galactose supplementation in PGM1-CDG, *Nat. Publ. Gr.* 00 (2017) 1–10. doi:10.1038/gim.2017.41.
- [20] E. Morava, Galactose supplementation in phosphoglucomutase-1 deficiency; review and outlook for a novel treatable CDG, *Mol. Genet. Metab.* 112 (2014). doi:10.1016/j.ymgme.2014.06.002.
- [21] A.L. Pey, F. Stricher, L. Serrano, A. Martinez, Predicted Effects of Missense Mutations on Native-State Stability Account for Phenotypic Outcome in Phenylketonuria, a Paradigm of Misfolding Diseases, *Am. J. Hum. Genet.* 81 (2007) 1006–1024. doi:10.1086/521879.
- [22] J. Thusberg, M. Vihinen, Pathogenic or not? and if so, then how? Studying the effects of missense mutations using bioinformatics methods, *Hum. Mutat.* 30 (2009) 703–714. doi:10.1002/humu.20938.
- [23] Y. Yang, B. Chen, G. Tan, M. Vihinen, B. Shen, Structure-based prediction of the effects of a missense variant on protein stability, *Amino Acids.* 44 (2013). doi:10.1007/s00726-012-1407-7.
- [24] I. Adzhubei, D.M. Jordan, S.R. Sunyaev, Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2, *Protoc. Hum. Genet.* 76 (n.d.). doi:10.1002/0471142905.hg0720s76.
- [25] Y. Lee, K.M. Stiers, B.N. Kain, L.J. Beamer, Compromised catalysis and potential folding defects in in vitro studies of missense mutants associated with hereditary phosphoglucomutase 1 deficiency, *J. Biol. Chem.* 289 (2014) 32010–32019. doi:10.1074/jbc.M114.597914.
- [26] K.M. Stiers, B.N. Kain, A.C. Graham, L.J. Beamer, Induced structural disorder as a molecular mechanism for enzyme dysfunction in phosphoglucomutase 1 deficiency, *J. Mol. Biol.* 428 (2016) 1493–1505. doi:10.1016/j.jmb.2016.02.032.
- [27] K.M. Stiers, A.C. Graham, B.N. Kain, L.J. Beamer, Asp263 missense variants perturb the active site of human phosphoglucomutase 1, *FEBS J.* 284 (2017) 937–947. doi:10.1111/febs.14025.
- [28] K.M. Stiers, L.J. Beamer, A Hotspot for Disease-Associated Variants of Human PGM1 Is Associated with Impaired Ligand Binding and Loop Dynamics., *Structure.* 26 (2018) 1337–1345.e3. doi:10.1016/j.str.2018.07.005.

Chapter II

Induced structural disorder as a molecular mechanism for enzyme dysfunction in phosphoglucomutase 1 deficiency

Adapted with permission from Kyle M. Stiers, Bailee N. Kain, Abigail C. Graham, Lesa J. Beamer, *Journal of molecular biology* 428 (8), 1493-1505. Copyright 2016 Elsevier

II.1 Abstract

Human phosphoglucomutase 1 (PGM1) plays a central role in cellular glucose homeostasis, mediating the switch between glycolysis and gluconeogenesis through the conversion of glucose 1-phosphate and glucose 6-phosphate. Recent clinical studies have identified mutations in this enzyme as the cause of PGM1 deficiency, an inborn error of metabolism classified as both a glycogen storage disease and a congenital disorder of glycosylation. Reported here are the first crystal structures of two disease-related missense variants of PGM1, along with the structure of the wild-type enzyme. Two independent glycine to arginine substitutions (G121R and G291R), both affecting key active site loops of PGM1, are found to induce regions of structural disorder, as evidenced by a nearly complete loss of electron density for as many as 23 amino acids. The disordered regions are not contiguous in sequence to the site of mutation, and even cross domain boundaries. Other structural rearrangements include changes in the conformations of loops and side chains, some of which occur nearly 20 Å away from the site of mutation. The induced structural disorder is correlated with increased sensitivity to proteolysis and lower resolution diffraction, particularly for the G291R variant. Examination of the

multi-domain effects of these G→R mutations establishes a correlation between interdomain interfaces of the enzyme and missense variants of PGM1 associated with disease. These crystal structures provide the first insights into the structural basis of enzyme dysfunction in PGM1 deficiency, and highlight a growing role for biophysical characterization of proteins in the field of precision medicine.

II.2 Introduction

The enzyme PGM1 is well known for its central role in human metabolism, where it mediates the switch between glycolysis and gluconeogenesis. It catalyzes the interconversion of glucose 1-phosphate and glucose 6-phosphate, which also serve as precursors for the nucleotide sugar, galactose, and pentose phosphate pathways. In 2014, a landmark study identified PGM1 deficiency as a hereditary genetic disorder, with characteristics of both a glycogen storage disease (GSDXIV, MIM 612934) and a congenital disorder of glycosylation (CDG) of types I and II [1]. Affected individuals show varying clinical phenotypes, including hepatopathy, dilated cardiomyopathy, hypoglycemia, muscle weakness, exercise intolerance, growth retardation, and congenital malformations of the head such as cleft palate. PGM1 deficiency is autosomal recessive in inheritance, and associated with various types of mutations, including frame shifts, aberrant splicing, and missense variants. To date, the number of affected patients identified is small (~35) but growing [1–7]. Due to the recent characterization of this inherited condition, a biochemical understanding of the disease is not complete, and many questions remain regarding patient prognosis and possible treatments [8].

PGM1 is a cytoplasmic protein with 562 amino acids and is expressed in many human tissues. It belongs to the ubiquitous α -D-phosphohexomutase superfamily, members of which are found in organisms from all kingdoms of life [9]. The catalytic mechanism of phosphoglucomutase from rabbit (97% sequence identity to human) was extensively characterized over several decades by Ray and coworkers ([10] and references therein), who proposed the scheme shown in Figure II-1. The reaction entails two consecutive phosphoryl transfers and proceeds via a bisphosphorylated intermediate. The first phosphoryl transfer takes place from a conserved phosphoserine residue (Ser117) to substrate, creating glucose 1,6-bisphosphate. The intermediate reorients in the active site,

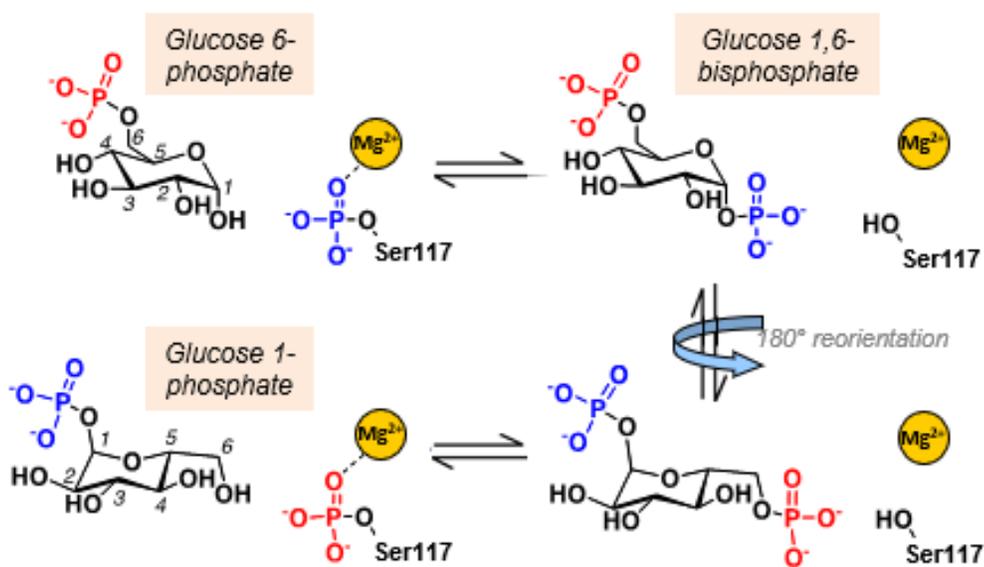


Figure II-1. A schematic of the catalytic reaction of PGM1, showing the reversible conversion of glucose 6-phosphate to glucose 1-phosphate. The bisphosphorylated intermediate undergoes a 180° reorientation (indicated by blue arrow) in the active site prior to the second phosphoryl transfer.

and the second phosphoryl transfer occurs from the intermediate back to the protein, creating product and regenerating the active (phosphorylated) version of the enzyme. The reaction is highly reversible, facilitating the role of the enzyme as a metabolic switch.

The identification of PGM1 deficiency as an inherited disease provides new opportunities for understanding this critical human enzyme and its role in metabolism. A number of disease-related missense variants have been biochemically characterized, showing effects on key regions/residues with proposed roles in enzyme function [11]. To date, however, no structural information is available for these mutants, leaving open questions regarding the nature and extent of their impact on protein structure. Here, we address this for the first time by solving crystal structures of two missense variants of PGM1, both of which contain a single Gly→Arg substitution. Previous work had shown that recombinant versions of these proteins, G121R and G291R, are well behaved in solution, but have significant catalytic defects [11]. Through comparison with the high-resolution crystal structure of wild-type (WT) human PGM1, also reported here, we find that both mutant enzymes suffer from regions of induced structural disorder and other effects resulting from the dramatic change in physicochemical characteristics at the site of mutation. The structural impacts of the mutation include both local and longer-range (i.e., >10 Å away) effects, and in some cases extend across domain boundaries. We also establish a correlation between the location of other known missense variants and interdomain interfaces of PGM1. These studies provide the first insights into the molecular origin of enzyme

dysfunction responsible for PGM1 deficiency, and may have relevance for future in vivo and clinical studies of this inborn error of metabolism.

II.3 Results

II.3.1 Overview of the wild-type structure

The high-resolution structure of human WT PGM1 is reported here for the first time. Previous models of human PGM1 relied on the known structure of the rabbit enzyme, which has been refined to a nominal resolution of 2.4 Å (PDB code: 3PMG) [12]. Our 1.85 Å structure (Table II-1 and Materials and Methods) confirms many of the features of this model, but provides additional detail on inter-residue interactions, as well as solvent and sulfate binding sites. We first describe the overall structure of the enzyme, and then the context of the glycines affected by the disease-related mutations G121R and G291R. Although the two copies of the polypeptide chain found in the asymmetric unit are chemically identical, chain A is more highly ordered in the crystals, and is used as the reference in structural descriptions below, unless otherwise noted.

Human PGM1 is a monomeric protein that shares the four-domain architecture common to all enzymes in the α -D-phosphohexomutase superfamily [9]. Its four domains are of roughly equal size, and arranged in an overall heart-shape (Figure II-2A). As noted in the structure of rabbit PGM [13], domains 1-3 share a common mixed α/β core, with a central 4-stranded β -sheet flanked on each side by an α -helix. The active site of PGM1 is located in a large central cleft, formed at the confluence of its four structural domains. This cleft has a

molecular surface area of 1670 Å² as calculated by CASTp [14], and involves 64 residues. It is quite hydrophilic, with nearly 70% of residues being charged or polar (including 11 arginines and lysines), as would be expected for binding of the phosphosugar substrates.

Proposed functional roles of four critical loops within the active site of PGM1 have been previously described (for detailed review see [15]). These loops are from domains 1-4 (D1-D4), respectively, and contain: i) phosphoserine 117 that participates in phosphoryl transfer; ii) the metal-binding loop including the three coordinating aspartates (residues 290, 292 and 294); iii) a sugar-binding loop; and iv) the phosphate-binding site that interacts with the phosphate group of the substrates. Figure II-2B shows a close-up view of these regions in the active site of PGM1; their context in the structure is described below.

Loops i and ii. The phosphoserine and metal binding loops in D1 and D2 are found in close proximity to one another, as required for their joint function in catalysis: Mg²⁺ serves as an electron-withdrawing group to facilitate phosphoryl transfer from phosphoserine 117 (Figure II-1). Phosphoserine 117 resides within a highly conserved TASHNP sequence motif (residues 115-120), and forms an Ω loop that is partially exposed to solvent. Ser117 is found in its dephosphorylated state in the crystal structure, and its side chain O_γ provides one ligand to the bound Mg²⁺ ion. The metal ion is coordinated in octahedral geometry: in addition to Ser117, ligands include three aspartates (residues 288, 290, and 292) and two water molecules. Other nearby residues with proposed roles in phosphoryl transfer include Arg23, which is strictly conserved in the enzyme

superfamily, and a potential candidate for the general acid in the reaction [16].

Lys389 is a candidate for the general base: it is analogous to Lys409 of S.

typhimurium PGM, which when mutated to alanine shows a 3000-fold reduction in kcat [16].

Table II-1. Data collection and refinement statistics

	WT	G121R	G291R
space group	<i>P4₁2₁2</i>	<i>P4₁2₁2</i>	<i>P4₁2₁2</i>
unit cell parameters (Å)	<i>a</i> = <i>b</i> = 172.6, <i>c</i> = 99.8	<i>a</i> = <i>b</i> = 171.4, <i>c</i> = 99.79	<i>a</i> = <i>b</i> = 170.8, <i>c</i> = 99.3
resolution (Å)	61.1–1.75 (1.88–1.85)	54.2 – 2.50 (2.58–2.50)	60.5–2.75 (2.87–2.75)
observations	4273578	740313	544355
unique reflections	162886	51905	38736
<i>R</i> _{merge} (<i>I</i>)	0.150 (1.959)	10.5 (1.618)	0.098 (1.566)
<i>R</i> _{pim} (<i>I</i>)	0.035 (0.558)	0.041 (0.663)	0.039 (0.640)
Mean <i>I</i> /σ(<i>I</i>)	18.5 (1.6)	19.6 (1.5)	20.0 (1.6)
Mean <i>CC</i> _{1/2}	0.999 (0.885)	0.999 (0.859)	0.999 (0.751)
completeness (%)	100.0 (99.9)	100.0 (99.9)	100.0 (100.0)
multiplicity	33.4 (25.9)	14.3 (13.3)	14.1 (13.4)
no. of protein residues	1118	1066	1063
no. of atoms	9413	8196	7463
no. of SO ₄ /glycerol molecules	8/9	8/2	5/0
no. of water molecules	819	80	38
<i>R</i> _{cryst}	0.170 (0.309)	0.192 (0.337)	0.210 (0.313)
<i>R</i> _{free}	0.198 (0.351)	0.243 (0.328)	0.271 (0.407)
rmsd bond lengths (Å)	0.007	0.008	0.010
rmsd bond angles (°)	0.998	1.125	1.167
Ramachandran plot ^a			
favored (%)	98.0	95.8	95.2
outliers (residues)	0	6	8
MolProbity score (%-tile)	100	95	99
Average B (Å ²)			
protein	40.0	77.0	80.4
water	44.7	53.0	69.8
coordinate error (Å) ^b	0.19	0.34	0.45
PDB code	5EPC	5F9C	5HSH

Values for the outer resolution shell of data are given in parentheses.

^aRamachandran plots generated with Molprobity via the PDB validation server. ^bMaximum likelihood-based coordinate error estimate reported by PHENIX.

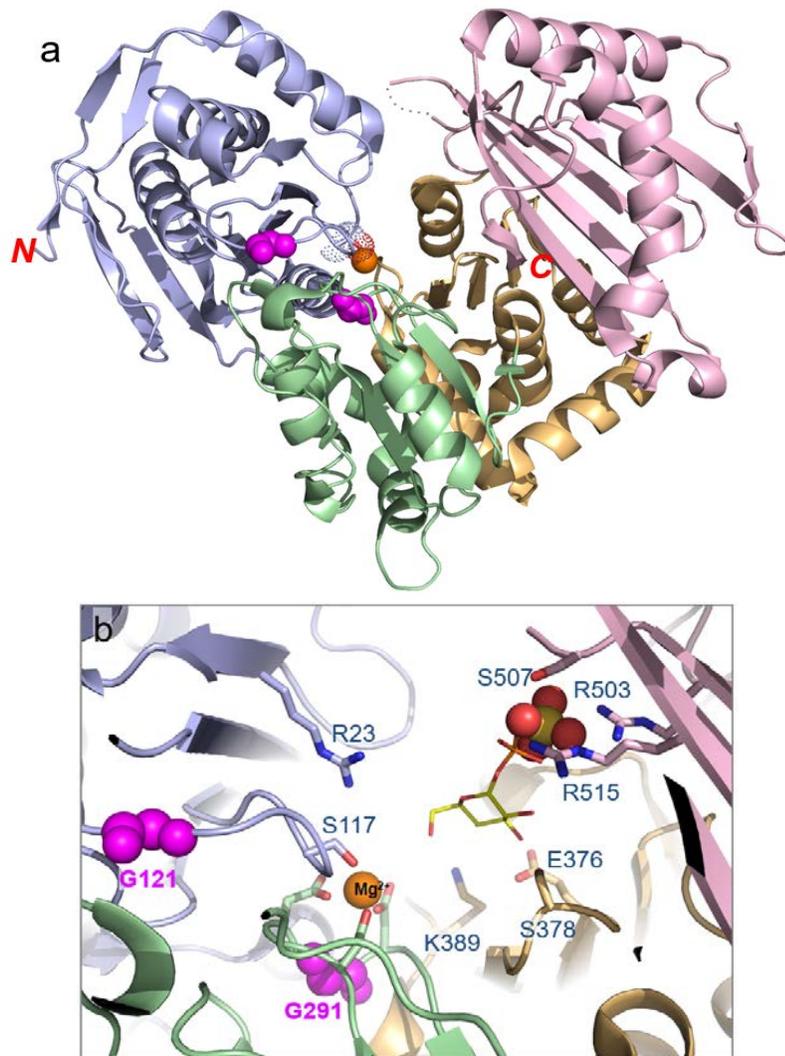


Figure II-2. The crystal structure of wild-type human PGM1. (a) A ribbon diagram of PGM1 colored by domain. Domain 1 (residues 1-191) is shown in blue, domain 2 (residues 192-304) in green, domain 3 (residues 305-421) in gold, and domain 4 (residues 422-562) in pink. Phosphoserine 117 is highlighted with dots, and Mg^{2+} is shown as an orange sphere. Glycines 121 and 291 are highlighted in magenta. For clarity, the three metal-binding aspartates (residues 288, 290 and 292) are not labeled. Missing residues in the D4 active site loop are shown with a dotted line. N- and C-termini are shown with red letters. (b) A close-up view of the active site WT PGM1 showing the four regions (i-iv) discussed in text; side chains of key active site residues are shown as sticks. Colors as in panel (a). The sulfate ion from the crystallization buffer that binds in the phosphate-binding site is shown with spheres. A model for the proposed binding of substrate glucose 1-phosphate is shown in thin lines (yellow), based on a structural superposition with a related enzyme-ligand complex (PDB code: 1P5D).

Loop iii. Residues in this loop of D3 are proposed to make key contacts with the O3 and O4 hydroxyl groups of the phosphosugar substrates. The GEESF motif (residues 375-379) is highly conserved in phosphoglucomutases [17], with both Glu376 and Ser378 of human PGM1 proposed to make direct interactions with the O3 and O4 sugar hydroxyls (see model of ligand in Figure II-2B). These residues are precisely oriented such that they can contact the hydroxyl groups of either glucose 1-phosphate or glucose 6-phosphate, which bind in two different orientations in the active site [18]. This dual-binding mode is essential to the reversibility of the PGM1 reaction, allowing it to use either glucose 1-phosphate or glucose 6-phosphate as substrate.

Loop iv. This region contains residues from two β -strands and a loop in D4 that are proposed to interact with the phosphate group of the substrate/product when bound in the active site. In the crystal structures of WT enzyme and both mutants, a sulfate ion derived from the crystallization buffer is found in this location (Figure II-2B). This sulfate acts as structural mimic for the phosphate group of the substrate, and confirms the identity of key residues involved in direct phosphate contacts, including the side chains of Arg503, Ser505, and Arg515, as well as the backbone amide of Gly506. Several residues in this loop (507-509) are missing from the electron density maps, suggesting they are mobile. This would be consistent with motion of the loop during the catalytic cycle, whereby it closes upon substrate binding and opens to release product.

II.3.2 Context of glycines 121 and 291 in the WT structure

Gly121 is located immediately following loop (i), just after the highly conserved TASHNP motif. In the WT structure, Gly121 is in a partially buried region between two turns, and has ϕ/ψ angles of $\sim 130/175^\circ$, which are uncommon except for glycine. Gly121 makes no direct contacts to other residues, but is located at the interface between D1 and D2 of the protein (Figure II-2B), adjacent to a short helix (residues 257-259) and extended region (residues 260-269) in D2.

Gly291 is found within loop (ii), the metal binding loop of PGM1. This loop is deeply buried within the active site cleft, and adjacent to residues from both D1 and D3 (Figure II-2B). Gly291 is the intervening residue between Asp290 and Asp292, both of which coordinate the metal ion. Its ϕ/ψ angles are $\sim 80/10^\circ$, values preferred for glycine. Gly291 makes several interactions with other residues, including a hydrogen bond between its backbone amide proton and the side chain of Asp288, and two hydrogen bonds from its backbone carbonyl to the amide of Gly391 and a water-mediated interaction to the backbone amide of Leu393.

II.3.3 Structure of the G121R missense variant

The crystal structure of the G121R missense variant was determined to 2.35 Å (Table II-1; Materials and Methods). Figure II-3A,B show an overview of the structure of the mutant, superimposed with that of WT enzyme. While the two structures are quite similar with an overall root-mean-square-deviation (rmsd)

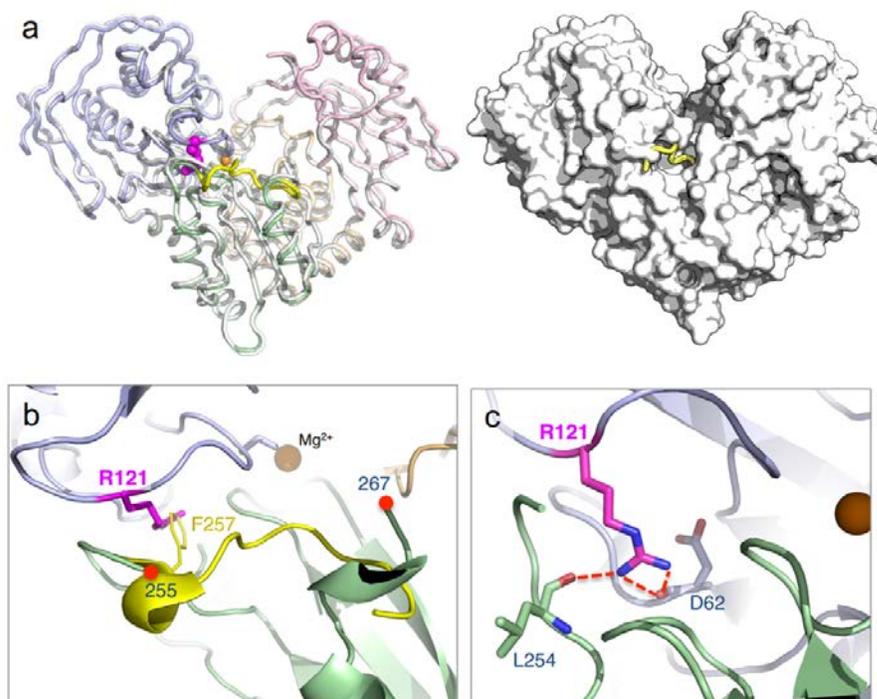


Figure II-3. Overall structure and detailed views of the G121R crystal structure. (a) A superposition of the G121R missense variant (colored by domain as in Fig. 2) and WT PGM1 (white) showing their overall structural similarity. Arg121 is shown in magenta; the region of the WT enzyme corresponding to the disordered residues in the G121R structure is in yellow. (b) The G121R mutant shown with a surface, and superimposed with WT enzyme (yellow), to highlight the patch of disordered residues in the mutant. (c) A close-up view comparing the vicinity of residue 121 in the WT and the G121R structures. The side chain of Phe257 (yellow), which lies in the interface between D1 and D2 in the WT structure, is displaced by Arg121 (magenta), causing disorder from residues 258-264 (between red circles). (d) Adventitious hydrogen bonds made between Arg121 and the backbone carbonyl groups of several residues in the D1-D2 interface.

of 0.51 Å for 540 C α pairs), the mutation produces both local and longer-range structural changes in the enzyme. In the vicinity of residue 121, the introduced arginine displaces the side chain of a nearby residue, Phe257, which packs in the interface between D1 and D2 in the WT structure (Figure II-3C). In its place, Arg121 makes adventitious hydrogen bond interactions with the backbone carbonyls of Leu254 and Asp62 (Figure II-3D).

These fortuitous interactions appear to compensate for the introduction of the polar, positively charged side chain of the arginine, which would otherwise be unfavorable in this buried environment. Despite the proximity of the G121R mutant to active site loops (i) and (ii), few changes to this region are observed, at least in chain A in the crystal structure. In chain B, which is less well ordered, neither the side chain of Arg121 nor loop (i) is clearly observed.

In addition to this local structural rearrangement, longer-range changes also occur, particularly in the adjacent region of D2. Electron density for the displaced side chain of Phe257 cannot be located in the maps, and subsequent to this residue, density for the entire polypeptide chain disappears and does not clearly reappear until approximately residue 265, presumably due to dynamic disorder in the crystal. A comparison of this area in the G121R and WT enzyme structures is shown in Figure II-3C. Although these residues have limited regular secondary structure in the WT enzyme, a number of well-defined inter-residue interactions involving this region are evident, all of which are lost in the structure of the mutant. The induced structural disorder appears to originate from the steric rearrangement of Phe257 forced by the arginine substitution, and subsequent accommodation of its side chain in the D1-D2 interface. Despite the extended region of disorder in the middle of the sequence, the polypeptide chain eventually returns to its proper course, adopting an essentially WT fold following residue 270. In total, the propagated effects of the G121R substitution extend more than 17 Å from the site of the mutation, with induced disorder of seven amino acids (Table II-2).

Table II-2. Disordered residues and kinetic parameters for the G121R and G291R missense variants relative to WT PGM1.

Protein	Domain of mutation	Disordered residues ^a (no.)	Domain	k_{cat} (s^{-1})	K_m (μM)	% k_{cat} / K_m rel. WT
WT	-	-	-	143±2	80±4	100
G121R	1	258-264 (7)	2	0.368±0.004	56±3	0.4
G291R	2	64-65 (2)	1	b.d.	b.d.	-
		117-125 (9)	1			
		256-267 (12)	2			

^aDisordered residues are listed relative to WT enzyme as found in chain A of each crystal structure. Kinetic parameters are from [11]; b.d., below detection.

II.3.4 Structure of the G291R missense variant

The crystal structure of the G291R missense variant was determined to 2.75 Å (Table II-1; Materials and Methods). Despite crystals of similar size, diffraction data collected from these crystals were noticeably lower in resolution compared to WT (1.85 Å) and the G121R mutant (2.5 Å). The structure of the G291R missense variant superimposed with that of WT enzyme is shown in Figure II-4A,B. As is the case for G121R, the overall structure of G291R is quite similar to that of WT (C α rmsd of 0.83 Å for 525 residues). Nevertheless, the single arginine substitution causes widespread structural changes, including induced disorder in three loops near the active site, as well as conformational rearrangements and differences in residue packing (Figure II-4c), as detailed below.

In the G291R structure, loop (ii) shows changes in the conformation of both its polypeptide backbone and side chains. The corresponding areas of the WT and G291R structures are displayed in Figure II-4D,E. The backbone angles of the polypeptide chain of G291R have altered to accommodate the introduced arginine side chain and the less flexible backbone of this residue relative to

glycine, moving the loop to a more solvent-exposed position in the active site cleft. Accompanying the backbone changes are significant differences in the orientations of the three metal-coordinating aspartates, with concomitant destruction of the metal-binding site. Aspartates 290 and 292, in particular, reorient dramatically: Asp290 becomes fully solvent exposed, while Asp292 makes a novel hydrogen bond interaction with Thr93 (Figure II-4E). The introduced arginine side chain is apparent in the electron density maps of chain A, where it points directly toward the center of the positively charged active site cleft. This unfavorable electrostatic environment may be mitigated by interactions with a sulfate ion and the high pH of the crystals (see Materials and Methods), which would tend to deprotonate Arg291 and/or other nearby positive residues.

Another apparent outcome of the conformational rearrangement of loop (ii) is induced disorder in the phosphoserine loop (i) of D1. As can be seen by comparing Figure 4d and e, the native conformation of this loop would be in steric conflict with the Figure II-4D and E, the native conformation of this loop would be in steric conflict with the rearranged position of loop (ii) in the mutant. Consequently, residues 117-125 are disordered in the electron density maps of the G291R mutant. In an apparent series of propagated changes, several other nearby loops are also disordered, including another in D1 (residues 64-65) and a larger region in D2 (residues 256-267). In total, 23 residues in two domains are affected by induced disorder in the structure of G291R mutant (Table II-2), including some that are nearly 20 Å away from the site of the mutation. Other

structural effects include changes in the packing of residues, such as Trp359 (Figure II-4C), which is one of several residues near loop (ii) that appear to have adjusted in response to its rearranged conformation. Thus, while the overall structure of the enzyme remains largely intact, changes due to this single residue substitution propagate across the protein, affecting multiple residues and domains.

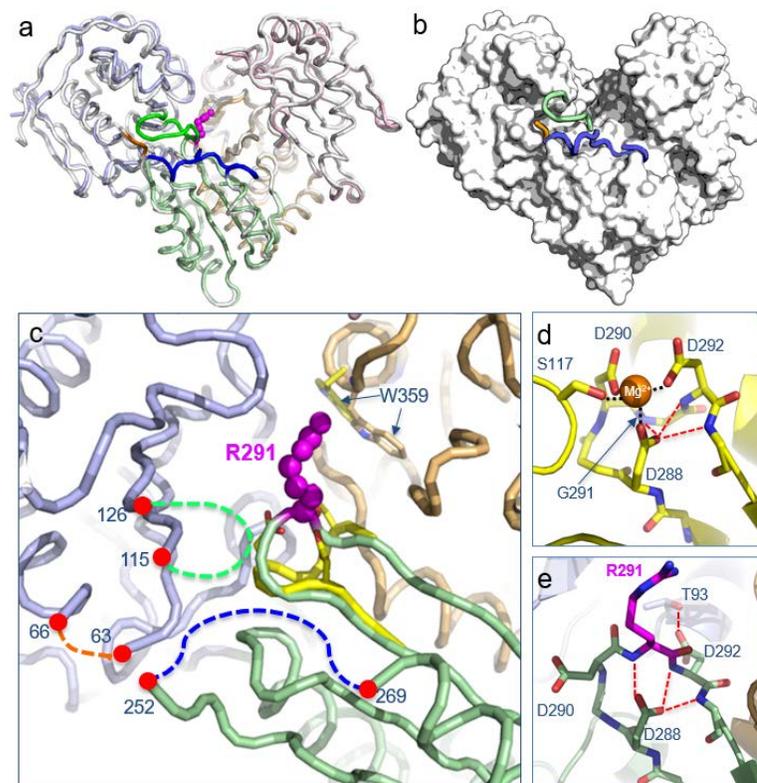


Figure II-4. Overall and detailed views of the G291R crystal structure. (a) A superposition of the G291R missense variant (colored by domain as in Fig. 2) and WT PGM1 (white) showing their overall structural similarity. Arg291 is shown in magenta; regions of the WT enzyme corresponding to the disordered residues in the mutant structure are highlighted in bright colors: residues 64-65 in orange; residues 116-125 in green; and residues 253-268 in blue. (b) The G291R mutant shown with a surface, and superimposed with WT enzyme as in (a). (c) The vicinity of Arg291 showing its location in the 3-way interface of domains 1, 2, and 3, and locations of the three disordered loops in the mutant structure (dashed lines). The metal binding loop from WT PGM1 is shown in yellow; the position of Trp359 in the WT and G291R structures shows its rearrangement in packing. (d) The metal-binding loop of WT PGM1, showing the bound Mg^{2+} ion, its coordinating aspartates, and location of Ser117. Hydrogen bonds made between Asp288 and other residues in the loop are in dashed red lines; coordinating interactions to the Mg^{2+} ion are in black dotted lines. (e) The region corresponding to (d) in the G291R structure, showing the rearrangement of the three aspartates and loss of the bound Mg^{2+} . Hydrogen bond interactions of Asp288 and Asp292 are shown with red dashed lines.

II.3.5 Correlation between structural and biochemical data

The crystal structures of the G121R and G291R missense variants provide a valuable opportunity for understanding biochemical data on these disease-related mutants [11]. Previous work on G121R showed that this protein is well expressed in *E. coli*, soluble, and its apparent molecular weight in solution is similar to WT enzyme. Likewise, its near-UV circular dichroism spectrum and T_m (only 1° lower) are quite similar to those of WT, although its interaction with the fluorescent dye 1-anilinonaphthalene-8-sulfonic acid (ANS) shows a moderate increase. Its catalytic activity is nevertheless significantly compromised, with k_{cat} / K_m only 0.4% of WT (Table II-2), although Ser117 in the mutant protein can still be phosphorylated by the activator glucose 1,6-bisphosphate. The structural studies of this variant, showing various perturbations near the active site cleft, are consistent with its catalytic impairment, which is significant but not complete. At least in chain A of the crystal structure, the position of Ser117 and the metal-binding loop are quite similar to WT, implying the mutant should be competent for catalysis. Disorder of loop (i) in chain B of the crystals, however, suggests that the mutant protein may exist as an inactive conformer more frequently than the WT enzyme. Together with other effects, including many small changes in/near the active site environment due to the disordered regions, it is not difficult to rationalize an overall decrease in the catalytic efficiency of G121R.

In the case of G291R, the structural disruption is more extensive, and this is reflected in the extreme catalytic defect of this mutant, which is the most

impaired of all the PGM1 missense variants characterized to date (activity below detection in in vitro kinetic assays; Table II-2) [11]. The inactivity of G291R is consistent with structural effects of the mutation on key loops within the active site, perturbing both metal binding and phosphoryl transfer, two essential features of the reaction. Defects in these processes are also reflected by the inability of the G291R variant to be phosphorylated under the conditions tested. While not directly linked to catalysis, the disordered loop containing residues 64-65 disrupts a conserved salt-bridge between Asp62 and Arg64, which in turn makes multiple interactions with loop (i) in the WT structure. Absence of these anchoring interactions with the critical phosphoserine loop would be a further impediment to phosphoryl transfer by this mutant protein.

The structural effects of the G291R missense variant are also interesting to consider in light of other biochemical data on this protein [11]. The mutant is generally well expressed in *E. coli* and soluble, although not quite at WT levels. In solution, the purified protein is well behaved with a similar apparent molecular weight to WT enzyme, while its T_m is a bit lower ($\sim 3^\circ$). Notably, however, its binding to ANS is significantly increased, and the G291R mutant is also more susceptible to limited proteolysis. These results make sense based on the observed changes in the crystal structure, particularly the regions of induced disorder, which could be easily proteolyzed. Similarly, the perturbation of the 3-way domain interface in which this residue is involved, and concomitant changes in residue packing (Figure II-4C), can be easily envisioned to allow enhanced binding of ANS, which typically occurs in apolar crevices in protein structures

[19]. In the future, it may be informative to conduct studies such as ANS-binding and limited proteolysis in the presence and absence of ligand, to determine whether ligand binding can mitigate the disordered regions of structure in mutants such as G121R and G291R.

II.3.6 Missense variants of PGM1 cluster in domain-domain interfaces

In light of the multi-domain effects of the G121R and G291R mutants, we analyzed other disease-related missense variants of PGM1 relative to the interdomain interfaces of the protein. The conserved four-domain architecture of PGM1 and the location of the active site cleft at the juncture of these structural domains (Figure II-2A) imply that mutations in these regions could be disproportionately disruptive to enzyme function. Moreover, in addition to the 13 originally characterized [11], several new missense mutants of PGM1 associated

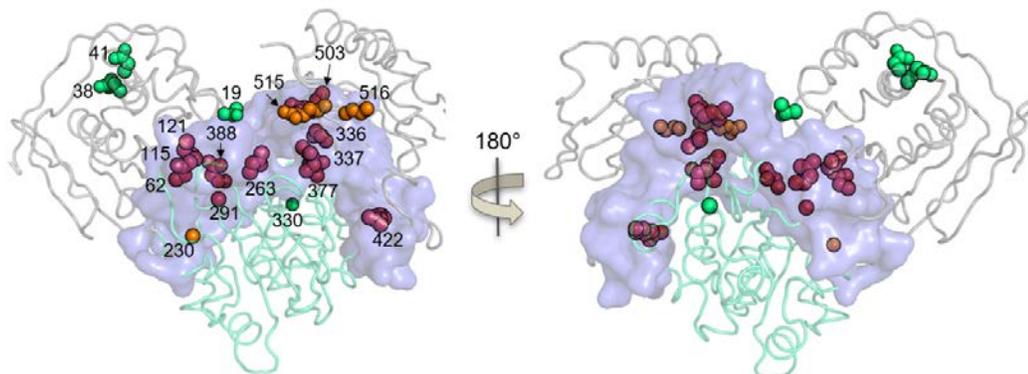


Figure II-5. Relationship of the known PGM1 missense variants to domain interfaces of the enzyme. A backbone trace of PGM1 showing the interface (blue surface) between the D2/3 unit (cyan) and the rest of the enzyme, i.e., domains 1 and 4 (gray). Most of the known disease-related missense variants cluster in (magenta spheres) or very near (orange spheres) to this extensive interface. The four missense variants that fall outside of this region are shown in green.

with disease have recently been identified [6]. These include G230E, T337M, P336R, R422W, R503Q, and R515L (Figure II-5). It can be seen that these new variants expand the structural regions associated with enzyme dysfunction: while many are near the active site cleft, others, such as G230E and R422W, are more distant from known regions of functional importance.

To assess the potential correlation between domain interfaces and the pathogenicity of missense variants, residues in the structure of WT PGM1 involved in each possible domain-domain interface were identified using LigPlot+ [20] and are detailed in Supplementary Figure II-S1. Of the six potential interfaces, four have contacting residues in the crystal structure: D1-D2, D1-D3, D2-D3, and D3-D4. Each interface includes at least one known disease-related mutation, although most are in either the D1-D2 or D3-D4 interface (see following paragraph). Notably, only one of the known missense variants resides in two distinct domain interfaces: G291R, which contacts residues in D1, D2, and D3 (see also Figure II-4C). Although not affected by known mutations, 13 other residues in PGM1 are also involved in 3-way domain interactions (Supplementary Figure II-S1) and might also be expected to be especially sensitive to structural perturbation by mutation.

As many of the known missense mutations localize to the D1-D2 and D3-D4 interfaces, it is convenient to consider domains 2 and 3 (residues 192-421) as a single structural unit (D2/3), and the interface between this unit and the rest of the enzyme as a single, extended surface (blue in Figure II-5). This large interface contains 101 residues out of 562 in the protein (18%), including much of

the active site cleft, but also extends to include regions between D1-D2 and D3-D4 that are relatively far from the catalytic center. Using this convention, it can be seen that the interface with the D2/3 unit encompasses 11 of the 18 residues affected by missense variants of PGM1, while four others are immediately adjacent (a combined total of 83%). Thus, it appears that mutations in this interface account for a striking proportion of the currently recognized disease-related variants of PGM1. It remains to be seen whether this distribution will change as new missense mutants are identified.

In addition to G121R and G291R, these interdomain mutations include many previously categorized as deleterious to catalysis (D62H, T115A, G121R, G230E, D263G/Y, G291R, T337M, and R503Q) [6,11]. However, several also appear to cause folding problems (P336R, E377K, E388K, and R422W) [6,11]. This suggests that a gradation of effects can occur due to substitutions at this interdomain interface, with the functional outcome dependent on factors such as the physicochemical change of the mutation and the structural interactions of the WT residue. It should be noted that several missense variants of PGM1 fall well outside of the D2/3 interface, suggesting the presence of additional “residue clusters” (e.g., residues 38 and 41, Figure II-5) with distinct functional roles, perhaps in protein folding. Mutants affecting residue 38 and 41 also tend to have milder clinical phenotypes than those affecting the interface residues [6], although the heterozygous backgrounds and limited number of patients prohibits a definitive analysis.

II.4 Discussion

Missense mutations are the most common genetic variation associated with monogenic disease, with >70,000 examples currently found in the Human Gene Mutation Database [21]. Decreased stability and folding problems are estimated to account for the pathogenicity of more than two-thirds of these [22,23], resulting in problems such as aggregation, improper trafficking, and rapid degradation. Less frequently, missense mutants are identified that directly impact function, i.e., affecting key catalytic residues or causing discrete structural changes like disruption of hydrogen bonding networks. While biochemical studies have been conducted on hundreds of missense variants responsible for human disease, parallel structural studies are not nearly as common, particularly for mutants with effects on protein folding/stability. Hence, while the importance of missense variants in human disease is well appreciated, a molecular understanding of the results of the mutation is often elusive [24], with many studies relying on modeling or other computational methods to predict likely effects.

Here we provide direct structural insights into the molecular basis of enzyme dysfunction for two Gly→Arg missense mutations associated with PGM1 deficiency. The structural studies of these two variants reveal multiple changes as a result of the single amino acid substitution, including the complete loss of previously ordered regions of the polypeptide chain. In lieu of structural information, biochemical data on the G121R and G291R variants suggested they be categorized as a “catalytic mutants” rather than defective in protein folding.

The crystallographic studies reveal that induced structural disorder affecting key residues in/near the active site is likely involved in this catalytic impairment, along with other structural changes such as conformational rearrangements. In essence, this blends the two previously suggested categories of enzyme dysfunction [11], suggesting that small regions of disordered structure may be one source of catalytic defects in missense variants of PGM1.

The G→R substitution in the G121R and G291R mutants is one of the more dramatic physicochemical changes among the disease-related variants of PGM1 [15], and is also one of the two variants most frequently associated with human disease in general [25,26]. Given the obvious differences between glycine and arginine, structural repercussions on the enzyme are not unexpected. Indeed, it may be more surprising that the overall tertiary fold of PGM1 remains largely intact, despite the structural impacts of the mutation. It appears that the structural scaffold of PGM1 has a high threshold of robustness, i.e., an excess of stability that buffers the destabilizing effects of mutations [23], which may be a testimony to the long evolutionary history of phosphoglucomutases. Also, as established in other systems [27], the multi-domain architecture of the enzyme could offer advantages, allowing individual domains “fold up” around the site(s) of structural disruption, despite their internal locations in the polypeptide chain. For several related proteins ([28,29] and PDB code: 1WJW), isolated versions of D4 fold correctly, supporting the notion of independent domain folding in this enzyme superfamily.

As is true for many enzymes, the active site of PGM1 is formed at the juncture of its structural domains, and this cleft includes many of the known disease-related mutants. However, by considering the D2/3 interface, a larger area of high mutational density is defined, that includes several additional missense variants. This perspective appears to connect disparate structural regions of PGM1 with the observed pathogenic effects of the mutations, and thus may have potential utility for predicting clinical phenotypes as new missense variants of PGM1 are identified. Moreover, we expect this theme could be relevant to many human diseases, given the preponderance of multi-domain proteins in the proteomes of higher organisms [30]. Several other proteins involved in inherited disorders also have missense variants that localize to interdomain interfaces [31–33], although their distribution across the 3D-structure is much wider overall, perhaps due to the larger number of mutations identified.

By combining the insights gained from the crystal structures of PGM1 with previous biochemical data [11], it is possible to assess, in retrospect, the utility of various assays used to characterize these mutant proteins. On one hand, assays monitoring global changes in protein behavior, like circular dichroism or dynamic light scattering, were not informative, at least with regard to the presence of induced structural disorder. On the other hand, methods such as the binding of ANS and limited proteolysis, appear to correlate well with the structural data. Accordingly, we can propose that several other disease-related PGM1 mutants may also suffer from induced structural disorder. This would include D62H and T115A, two mutants with profiles that fall between those of G121R and G291R in

ANS-binding spectra [11]. In contrast, other missense mutants, such as D263Y and T19A, show little change from WT enzyme in this assay, suggesting that their defects discretely affect catalysis/ligand binding, rather than protein structure. As many other proteins have missense variants that show differences in such assays (e.g., [34–36]), it is possible that induced disorder is a common outcome of disease-related mutations, which is underappreciated due to lack of structural information.

Overall, it seems that a combination of structural, kinetic, and other biochemical assays should be considered to fully describe the biochemical phenotype of the PGM1 missense variants. These characterizations may help direct and interpret future in vivo studies. For example, it will be intriguing to see whether “slightly misfolded” variants of PGM1, such as G121R and G291R, are targeted for degradation by the proteasome in human cells. Such studies have not yet been conducted, but if this turns out to be the case, use of these assays on newly identified variants might help suggest patients who would benefit from treatment with proteasome inhibitors. Proteostasis regulators have proven useful in the treatment of other inherited diseases associated with variants that affect protein folding and or stability [37].

Due to the recent recognition of PGM1 deficiency in the clinical literature, and the mild phenotypes of some patients, the number of individuals confirmed with this disorder is still quite small. To date, only one person with the G121R mutation in PGM1 has been identified, who was homozygous for the mutation and deceased at age 8 with end stage dilated cardiomyopathy [1]. In the case of

G291R, two unrelated patients are known, both heterozygous for other mutations in PGM1 (either a nonsense or different missense mutant), and who have mild-moderate phenotypes [1,2,6]. Our structural studies therefore represent “personalized” crystal structures, a new tool in the growing field of precision medicine [38,39]. Such structures are likely to have a growing role in the understanding and treatment of inherited disease. In the case of PGM1 deficiency, structural studies may be especially relevant given difficulties in correlating in vitro enzyme activity with patient phenotype [6]. Additional crystal structures of other missense variants PGM1 will likely provide further insights into the molecular bases of enzyme dysfunction. In combination with future clinical and in vivo studies, this should help unravel the multiple patient phenotypes associated with PGM1 deficiency.

II.5 Materials and Methods

II.5.1 Protein expression, purification, and crystallization

Human WT PGM1 and the G121R and G291R missense variants were expressed recombinantly in *E. coli* and purified to homogeneity via an N-terminal histidine tag, as previously described [11]. Prior to crystallization, the histidine tag was removed by cleavage with tobacco etch virus (TEV) protease as follows. Proteins were dialyzed into Buffer 1 (0.3 M NaCl and 50 mM Tris, pH 8.0), then mixed at 10-20 fold ratio (w/w) with TEV protease (Buffer 1 with 0.5 mM DTT and 0.5 mM EDTA) and glycerol (10% v/v final). The mixture was incubated at room temperature for 3-4 hours, kept overnight at 4° C, and extensively dialyzed into Buffer 1 before loading onto a pre-equilibrated Ni-NTA column. Cleaved protein eluted in the flow through; TEV and any residual His-tagged proteins were retained on the column, and eluted with Buffer 1 plus 300 mM imidazole, pH 8.0. Successful cleavage was confirmed by electrospray ionization mass spectrometry; all proteins were the expected molecular weight with no indication of proteolysis. The purified, cleaved proteins were dialyzed into a solution of 50 mM MOPS, pH 7.4, with 1 mM MgCl₂, and concentrated to ~10 mg / mL or higher. If not used immediately, samples were flash-frozen in liquid nitrogen and stored at -80° C. Prior to crystallization, the protein was diluted as desired and 0.22 µm filtered.

Initial crystallization screens were set up with Crystal Screen kits 1 and 2 (Hampton Research) and Wizard screen kits 1 and 2 (Emerald BioSystems Inc.)

at protein concentrations ranging from 10-18 mg / mL using the hanging drop vapor diffusion method at 20° C. Drops containing 2 µl protein solution and 2 µl crystallization buffer were sealed over a 0.5 mL reservoir. The WT protein crystallized successfully in a range of conditions from ammonium or lithium sulfate (1.35 – 1.85 M) with either 0.1 M buffer of MES, pH 6.0, or Tris HCl, pH 7.5 – 8.5. Crystals grew in approximately one week. The G121R mutant crystallized in ~1.5 M ammonium sulfate with lithium sulfate (0.12-0.18 M) and 0.1 M CAPS, pH 10.5. The G291R mutant crystallized in ~1.0 M sodium citrate with 0.1 M CHES buffer, pH 9.7.

All crystals were cryoprotected using a solution of well buffer supplemented with 30% glycerol (v/v), mounted on Hampton loops, and flash cooled in liquid nitrogen. WT and mutant proteins crystallized in space group P41212 with ~60% solvent ($VM = 3.04 \text{ \AA}^3/\text{Da}$). The asymmetric unit contains two copies of the polypeptide chain.

II.5.2 X-ray Diffraction Data Collection and Refinement

Diffraction data were collected at a wavelength of 1.00003 Å from single crystals on beamline 4.2.2 of the Advanced Light Source using a Taurus-1 CMOS detector in shutterless mode. The data were processed using XDS [40] and AIMLESS [41] via CCP4i [42]. Data processing statistics are listed in Table II-1. Values of $CC1/2 > 0.30$ [43] and R_{pim} [44] were considered when determining the high resolution cutoff, due to the large number of images (1800-3600 per data set) and high redundancy obtained with the shutterless data collection. In addition, all data sets were anisotropic, especially those of G121R

and G291R, which contributes to poor Rmerge statistics at high resolution for these two data sets, although CC1/2 is acceptable in the chosen range.

Crystallographic refinement calculations for the WT enzyme were initiated using coordinates derived from the 2.4 Å resolution structure of rabbit PGM (PDB code: 3PMG), which crystallizes isomorphously with human PGM1 (97% sequence identity). Refinement was performed with REMAC 5.0 [45] and PHENIX [46]; progress was monitored by following R free with 5% of each data set was set aside for cross validation. The B-factor model consisted of an isotropic B-factor for each atom; TLS refinement was used as automated in PHENIX. COOT [47] was used for model building. Additional details of the refinement are on Table II-1. The structures of the G121R and G291R missense variants were determined using the model of WT enzyme, starting with molecular replacement using MOLREP [48] or rigid-body followed by restrained refinement. The Rfree data sets for both mutants were constrained to match those of the WT data. The structures were validated using MolProbity [49]. Refinement statistics are listed in Table II-1. Structural figures were prepared with PYMOL [50]. Coordinates and structure factor amplitudes have been deposited in the PDB under the accession numbers listed in Table II-1.

II.6 Acknowledgements

We thank Jay Nix of ALS beamline 4.2.2 for assistance with data collection and processing, Ritcha Mehra-Chaudhary of the University of Missouri Structural Biology Core for assistance with protein purification, and Brian Mooney of the University of Missouri Charles W. Gehrke Proteomics Center for mass spectrometry. This work was supported by grants to LJB from the Patton Trust of the Kansas City Area Life Sciences Research Foundation and the National Science Foundation (MCB-0918389). Part of this work was performed at the Advanced Light Source. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under contract DE-AC02-05CH11231.

II.7 References

- [1] L.C. Tegtmeyer, S. Rust, M. van Scherpenzeel, B.G. Ng, M.-E. Losfeld, S. Timal, et al., Multiple Phenotypes in Phosphoglucomutase 1 Deficiency, *N. Engl. J. Med.* 370 (2014) 533–542.
- [2] B. Pérez, C. Medrano, M.J. Ecay, P. Ruiz-Sala, M. Martínez-Pardo, M. Ugarte, et al., A novel congenital disorder of glycosylation type without central nervous system involvement caused by mutations in the phosphoglucomutase 1 gene, *J. Inherit. Metab. Dis.* 36 (2012) 535–542.
- [3] S. Timal, A. Hoischen, L. Lehle, M. Adamowicz, K. Huijben, J. Sykut-Cegielska, et al., Gene identification in the congenital disorders of glycosylation type I by whole-exome sequencing., *Hum. Mol. Genet.* 21 (2012) 4151–4161.
- [4] A. Kucukcongar, L. Tumer, F. Suheyl Ezgu, C. Seher Ksapkara, J. Jaeken, G. Matthijs, et al., A case with rare type of congenial-disorder of glycosylation: PGM1-CDG, *Genet. Couns.* 26 (2015) 87–90.
- [5] N. Ondruskova, T. Honzik, A. Vondrackova, M. Tesarova, J. Zeman, H. Hansikova, Glycogen storage disease-like phenotype with central nervous system involvement in a PGM1-CDG patient, *Neuro Endocrinol Lett.* 35 (2014) 137–141.
- [6] S. Wong, L.J. Beamer, T. Gadomski, T. Honzik, M. Mohamed, S.B. Wortmann, et al., Defining the phenotype and assessing severity in Phosphoglucomutase-1 deficiency, *Under Revision*.
- [7] N. Loewenthal, A. Haim, R. Parvari, E. Hershkovitz, Phosphoglucomutase-1 deficiency: Intrafamilial clinical variability and common secondary adrenal insufficiency, *Am. J. Med. Genet. Part A.* 167 (2015) 3139–3143.
- [8] E. Morava, Galactose supplementation in phosphoglucomutase-1 deficiency; review and outlook for a novel treatable CDG., *Mol. Genet. Metab.* 112 (2014) 275–279.
- [9] G.S. Shackelford, C.A. Regni, L.J. Beamer, Evolutionary trace analysis of the alpha-D-phosphohexomutase superfamily., *Protein Sci.* 13 (2004) 2130–2138.
- [10] W.J. Ray, J.W. Burgner, C.B. Post, Characterization of vanadate-based transition-state-analogue complexes of phosphoglucomutase by spectral and NMR techniques., *Biochemistry.* 29 (1990) 2770–8.
- [11] Y. Lee, K.M. Stiers, B.N. Kain, L.J. Beamer, Compromised catalysis and potential folding defects in in vitro studies of missense mutants associated with hereditary phosphoglucomutase 1 deficiency., *J. Biol. Chem.* 289 (2014) 32010–32019.

- [12] Y. Liu, W.J. Ray, S. Baranidharan, Structure of rabbit muscle phosphoglucomutase refined at 2.4 Å resolution, *Acta Crystallogr D.* 53 (1997) 392–405.
- [13] J.-B. Dai, W.J. Ray, M. Konno, The crystal structure of muscle phosphoglucomutase refined at 2.7 Å resolution, *J. Biol. Chem.* (2001) 1–16.
- [14] J. Dundas, Z. Ouyang, J. Tseng, A. Binkowski, Y. Turpaz, J. Liang, CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues, *Nucleic Acids Res.* 34 (2006) 116–118.
- [15] L.J. Beamer, Mutations in hereditary phosphoglucomutase 1 deficiency map to key regions of enzyme structure and function, *J. Inherit. Metab. Dis.* (2014) 1–14.
- [16] Y. Lee, R. Mehra-Chaudhary, C. Furdui, L.J. Beamer, Identification of an essential active-site residue in the alpha-D-phosphohexomutase enzyme superfamily, *FEBS J.* 280 (2013) 2622–2632.
- [17] R. Mehra-Chaudhary, J. Mick, J.J. Tanner, M.T. Henzl, L.J. Beamer, Crystal structure of a bacterial phosphoglucomutase, an enzyme involved in the virulence of multiple human pathogens., *Proteins.* 79 (2011) 1215–29.
- [18] C. Regni, L. Naught, P. a. Tipton, L.J. Beamer, Structural Basis of Diverse Substrate Recognition by the Enzyme PMM/PGM from *P. aeruginosa*, *Structure.* 12 (2004) 55–63.
- [19] A. Hawe, M. Sutter, W. Jiskoot, Extrinsic fluorescent dyes as tools for protein characterization., *Pharm. Res.* 25 (2008) 1487–99.
- [20] R.A. Laskowski, M.B. Swindells, LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery, *J. Chem. Inf. Model.* 51 (2011) 2778–2786.
- [21] D.N. Cooper, P.D. Stenson, N. a Chuzhanova, The Human Gene Mutation Database (HGMD) and its exploitation in the study of mutational mechanisms., *Curr. Protoc. Bioinformatics.* Chapter 1 (2006) Unit 1.13.
- [22] P. Yue, Z. Li, J. Moult, Loss of Protein Structure Stability as a Major Causative Factor in Monogenic Disease, *J. Mol. Biol.* 353 (2005) 459–473. doi:10.1016/j.jmb.2005.08.020.
- [23] N. Tokuriki, D.S. Tawfik, Stability effects of mutations and protein evolvability, *Curr. Opin. Struct. Biol.* 19 (2009) 596–604.
- [24] S. Stefl, H. Nishi, M. Petukh, A.R. Panchenko, E. Alexov, Molecular Mechanisms of Disease-Causing Missense Mutations, *J. Mol. Biol.* 425 (2013) 3919–3936.

- [25] M. Petukh, T.G. Kucukkal, E. Alexov, On human disease-causing amino acid variants: Statistical study of sequence and structural patterns, *Hum. Mutat.* 36 (2015) 524–534.
- [26] T.A. Peterson, E. Doughty, M.G. Kann, Towards precision medicine: Advances in computational approaches for the analysis of human variants, *J. Mol. Biol.* 425 (2013) 4047–4063.
- [27] J.-H. Han, S. Batey, A. a Nickson, S. a Teichmann, J. Clarke, The folding and evolution of multidomain proteins., *Nat. Rev. Mol. Cell Biol.* 8 (2007) 319–30.
- [28] A.M. Schramm, D. Karr, R. Mehra-Chaudhary, S.R. Van Doren, C.M. Furdui, L.J. Beamer, Breaking the covalent connection: Chain connectivity and the catalytic reaction of PMM/PGM., *Protein Sci.* 19 (2010) 1235–42.
- [29] Y. Wei, T.C. Marcink, J. Xu, A.G. Sirianni, A.V.S. Sarma, S.H. Prior, et al., Chemical shift assignments of domain 4 from the phosphohexomutase from *Pseudomonas aeruginosa* suggest that freeing perturbs its coevolved domain interface, *Biomol. NMR Assign.* 8 (2014) 329–333.
- [30] D. Ekman, A.K. Björklund, J. Frey-Skött, A. Elofsson, Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions., *J. Mol. Biol.* 348 (2005) 231–43.
- [31] A.L. Pey, F. Stricher, L. Serrano, A. Martinez, Predicted effects of missense mutations on native-state stability account for phenotypic outcome in phenylketonuria, a paradigm of misfolding diseases., *Am. J. Hum. Genet.* 81 (2007) 1006–1024.
- [32] U. Ohto, K. Usui, T. Ochi, K. Yuki, Y. Satow, T. Shimizu, Crystal structure of human β -galactosidase: Structural basis of G M1 gangliosidosis and morquio B diseases, *J. Biol. Chem.* 287 (2012) 1801–1812.
- [33] K. Fulop, L. Barna, O. Symmons, P. Zavodsky, A. Varadi, Clustering of disease-causing mutations on the domain-domain interfaces of ABCC6, *Biochem. Biophys. Res. Commun.* 379 (2009) 706–709.
- [34] M. Bozzi, A. Cassetta, S. Covaceuszach, M.G. Bigotti, S. Bannister, W. Hübner, et al., The Structure of the T190M Mutant of Murine α -Dystroglycan at High Resolution: Insight into the Molecular Basis of a Primary Dystroglycanopathy., *PLoS One.* 10 (2015) e0124277.
- [35] M.S. Lee, R. Green, S.M. Marsillac, N. Coquelle, R.S. Williams, T. Yeung, et al., Comprehensive analysis of missense variations in the BRCT domain of BRCA1 by structural and functional assays, *Cancer Res.* 70 (2010) 4880–4890.
- [36] E. Bjørgo, P.M. Knappskog, A. Martinez, R.C. Stevens, T. Flatmark, Partial characterization and three-dimensional-structural localization of

- eight mutations in exon 7 of the human phenylalanine hydroxylase gene associated with phenylketonuria., *Eur. J. Biochem.* 257 (1998) 1–10.
- [37] A.C. Muntau, J. Leandro, M. Staudigl, F. Mayer, S.W. Gersting, Innovative strategies to treat protein misfolding in inborn errors of metabolism: pharmacological chaperones and proteostasis regulators., *J. Inherit. Metab. Dis.* 37 (2014) 505–523.
- [38] B.M. Kroncke, C.G. Vanoye, J. Meiler, A.L. George, C.R. Sanders, Personalized Biochemistry and Biophysics., *Biochemistry.* 54 (2015) 2551–2559.
- [39] E. Alexov, *Advances in Human Biology : Combining Genetics and Molecular Biophysics to Pave the Way for Personalized Diagnostics and Medicine*, *Adv. Biol.* 2014 (2014).
- [40] W. Kabsch, Software XDS for image rotation, recognition and crystal symmetry assignment, *Acta Crystallogr., Sect. D Biol. Crystallogr.* 66 (2010) 125–132.
- [41] P.R. Evans, G.N. Murshudov, How good are my data and what is the resolution?, *Acta Crystallogr. D. Biol. Crystallogr.* 69 (2013) 1204–14.
- [42] E. Potterton, P. Briggs, M. Turkenburg, E. Dodson, A graphical user interface to the CCP 4 program suite, *Acta Crystallogr. Sect. D Biol. Crystallogr.* 59 (2003) 1131–1137.
- [43] P.A. Karplus, K. Diederichs, Linking crystallographic model and data quality, *Science (80-.).* 336 (2012) 1030–1033..
- [44] M.S. Weiss, Global indicators of X-ray data quality, *J. Appl. Crystallogr.* 34 (2001) 130–135.
- [45] G.N. Murshudov, P. Skubák, A. a. Lebedev, N.S. Pannu, R. a. Steiner, R. a. Nicholls, et al., REFMAC5 for the refinement of macromolecular crystal structures, *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67 (2011) 355–367.
- [46] P.D. Adams, P. V. Afonine, G. Bunkóczi, V.B. Chen, I.W. Davis, N. Echols, et al., PHENIX: A comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66 (2010) 213–221.
- [47] P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics, *Acta Crystallogr. Sect. D Biol. Crystallogr.* 60 (2004) 2126–2132.
- [48] A. Vagin, A. Teplyakov, Molecular replacement with MOLREP, *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66 (2010) 22–25.
- [49] V.B. Chen, W.B. Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, et al., MolProbity: all-atom structure validation for macromolecular crystallography., *Acta Crystallogr. D. Biol. Crystallogr.* 66 (2010) 12–21.

[50] W.L. DeLano, The PyMOL Molecular Graphics System, Schrödinger LLC
Www.pymol.org. Version 1. (2002)

II.8 Supporting Material

II.8.1 Supporting Figures

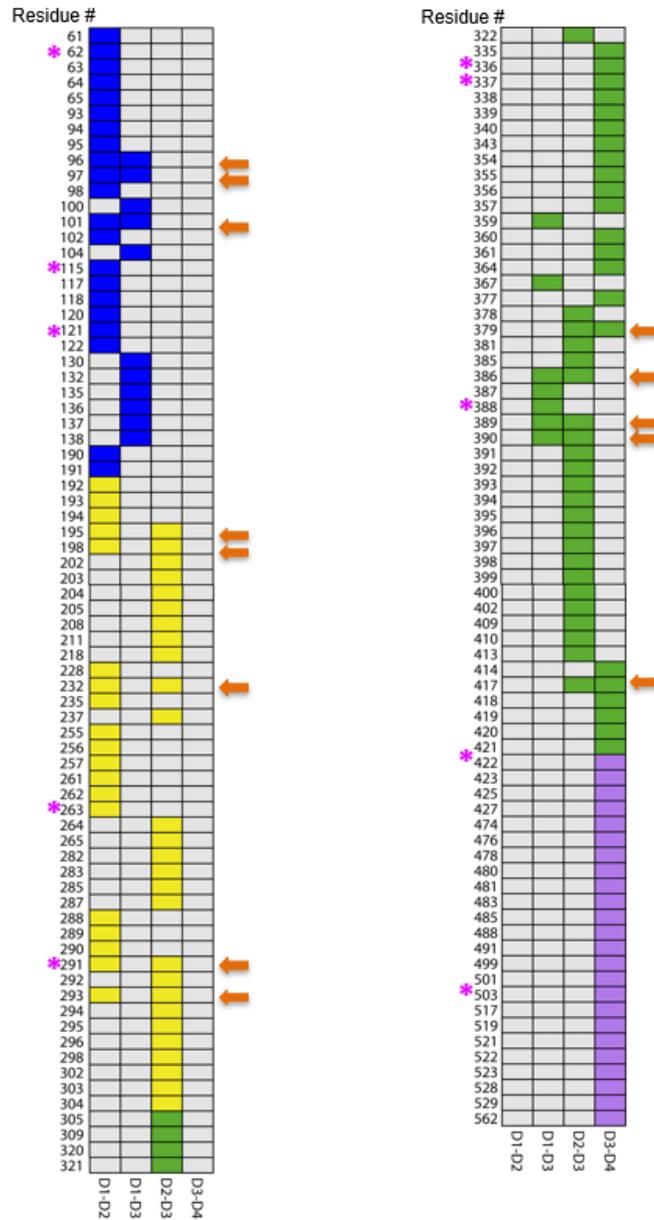


Figure II-S1. A listing of PGM1 residues that participate in interdomain interfaces in the WT enzyme. Interface residues as determined by DIMPLOT are listed vertically; individual domain-domain interfaces are indicated across the bottom. Residues found in D1, D2, D3 and D4 are shown using different colored boxes of blue, yellow, green, and purple, respectively. Residues involved in more than one interface are highlighted with orange arrows; asterisks mark residues that correspond to disease-related missense variants.

Chapter III

Asp263 missense variants perturb the active site of human phosphoglucomutase 1 (PGM1)

Adapted with permission from Kyle M. Stiers, Abigail C. Graham, Bailee N. Kain, Lesa J. Beamer, *The FEBS Journal* 284 (6), 937-947. Copyright 2017 Federation of European Biochemical Societies

III.1 Abstract

The enzyme phosphoglucomutase 1 (PGM1) plays a central role in glucose homeostasis. Clinical studies have identified mutations in human PGM1 as the cause of PGM1 deficiency, an inherited metabolic disease. One residue, Asp263, has two known variants associated with disease: D263G and D263Y. Biochemical studies have shown that these mutants are soluble and well folded, but have significant catalytic impairment. To better understand this catalytic defect, we determined crystal structures of these two missense variants, both of which reveal a similar and indirect structural change due to the loss of a conserved salt bridge between Asp263 and Arg293. The arginine reorients into the active site, making interactions with residues responsible for substrate binding. Biochemical studies also show that the catalytic phosphoserine of the missense variants is more stable to hydrolysis relative to wild-type enzyme. The structural perturbation resulting from mutation of this single amino acid reveals the molecular mechanism underlying PGM1 deficiency in these missense variants.

III.2 Introduction

Human PGM1 (EC 5.4.2.2) has a pivotal role in glucose metabolism, mediating the switch between glycolysis and gluconeogenesis. Recently, mutations in this enzyme have been identified as the cause of an inherited metabolic disorder, PGM1 deficiency, which has characteristics of both a glycogen storage disease (GSDXIV, OMIM 612934) and a congenital disorder of glycosylation of types I and II [1–3]. Affected individuals have variable clinical phenotypes [4–6], which include dilated cardiomyopathy, hepatopathy, hypoglycemia, muscle weakness, growth retardation, delayed puberty, and congenital malformations such as cleft palate. Some affected individuals live relatively normal lives, while others face life-threatening complications, including four patients (ages 8 – 33 years) who suffered cardiac arrest and others who were listed for heart transplants [4].

PGM1 accomplishes its role in glucose metabolism through the isomerization of 1- and 6-phosphosugars via a bisphosphorylated intermediate [7]. Catalysis entails two consecutive phosphoryl transfers: first from a phosphoserine residue of the protein to substrate, and next from the intermediate back to the protein (Figure III-1A). The reaction is highly reversible, and a single Mg^{2+} ion is required for activity. The crystal structure of wild-type (WT) human PGM1 has recently been reported [8]. The protein has 562 amino acids, with four domains arranged in an overall heart shape (Figure III-1B). The active site is found in a large, centrally located cleft and comprises key functional regions, including the catalytic phosphoserine, a loop that binds the Mg^{2+} ion, as well as

residues involved in contacting the sugar hydroxyls and phosphate group of the substrates [8]. The catalytic mechanism and key active site residues are highly conserved across the ubiquitous α -D-phosphohexomutase enzyme superfamily [9], of which PGM1 is a member.

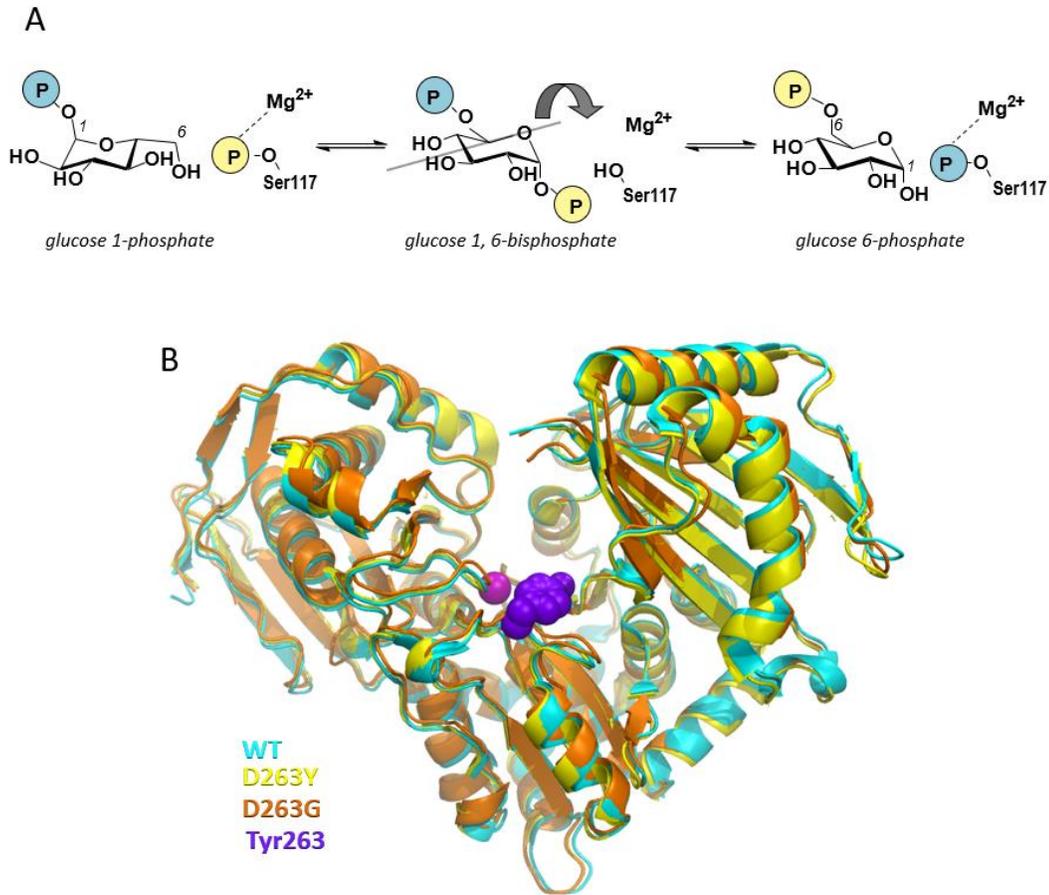


Figure III-1. Overview of the mechanism and structure of human PGM1. (A) A schematic of the catalytic reaction, showing the reversible conversion of glucose 1-phosphate to glucose 6-phosphate. The glucose 1,6-bisphosphate intermediate undergoes a 180° reorientation in between the two phosphoryl transfer steps of the reaction (gray line indicates axis of rotation). (B) A superposition showing the overall similarity between the structures of WT human PGM1 and the D263G and D263Y missense variants. The WT structure is shown in cyan, D263G in orange, and the D263Y in yellow. The bound metal ion as shown as a magenta sphere and the side chain of Tyr263 with purple spheres.

PGM1 deficiency is autosomal recessive in inheritance, and associated with various types of mutations, including frame shifts, aberrant splicing, and missense variants [1–4,10–12]. In the latter category, ~20 different mutations are currently known, affecting 19 residue positions in the protein. Among these, residue 263 of PGM1 stands out as the only position affected by two missense mutants, D263Y and D263G, with confirmed association to disease. Previous biochemical studies indicated that both of these variants were folded and of similar stability to WT enzyme. However, despite the lack of a known role for Asp263 in catalysis or ligand binding, both mutants exhibited significant catalytic impairment, with k_{cat} of 1-2% that of WT enzyme [13].

To better understand the origin of enzyme dysfunction in PGM1 deficiency, we report herein the crystal structures of the D263Y and D263G missense variants. In combination with biochemical analyses, the structural work reveals that an indirect structural change impacting the active site is responsible for their catalytic impairment. This result expands our understanding of the molecular bases of enzyme dysfunction in missense variants of human PGM1.

III.3 Results

III.3.1 Asp263 participates in a conserved salt bridge in WT PGM1

To understand the structural impacts of the D263Y/G mutations, we first examine the context of this residue in WT PGM1. The protein may exist as either an active, phospho-enzyme or inactive, dephospho-enzyme, depending on the phosphorylation state of its catalytic serine (Ser117); both versions are presented

in Figure III-2A. In this figure, the structure of dephospho-enzyme is that of WT human PGM1 from PDB ID 5EPC at 1.9 Å resolution [8], and the phospho-enzyme structure is of rabbit PGM (97% identity with human enzyme) from PDB

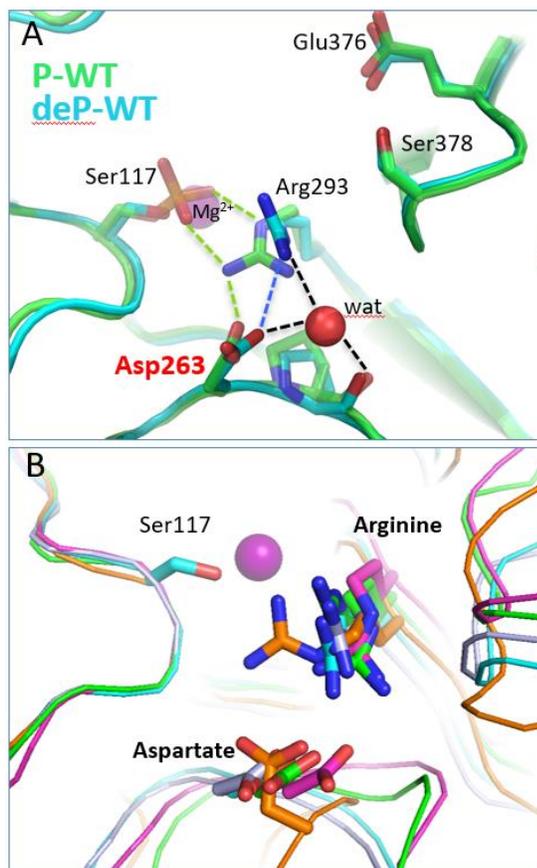


Figure III-2. The context of residue 263 in the structure of WT PGM1. (A) Direct and water-mediated contacts made by the side chain of Asp263, including those with Arg293. Ser117 and the metal ion are shown for reference. WT human PGM1 (PDB ID 5EPC; cyan) is shown in its dephospho-state; the phospho-enzyme version of rabbit PGM (PDB ID 3PMG; 97% identical to human enzyme) is shown in green. Contacts made between Asp263, Arg293, and Ser117 are shown by dashed lines in green for those unique to 3PMG, in blue for 5EPC, and in black when common to both structures. (B) A superposition of related enzymes with PGM1 (cyan) showing the conservation of the analogous aspartate – arginine residue pair in enzymes from different sub-groups and diverse organisms within the superfamily. Structures are from *P. aeruginosa* phosphomannomutase/phosphoglucomutase (PDB ID 1P5D) in green; *Salmonella typhimurium* PGM (PDB ID 3NA5) in magenta; parafusin from *Paramecium tetraaurelia* (PDB ID 1KFI) in gray; and *Candida albicans* N-acetylglucosamine-phosphate mutase (PDB ID 2DKA) in orange. Ser117 of human PGM1 is shown for reference.

ID 3PMG at 2.4 Å resolution [14]. As chain A is better ordered in the PGM1 crystals, it is used for all structural analyses herein; similar interactions are found in chain B, though bond distances vary.

Asp263 is located near the active site cleft, in the vicinity of several important catalytic regions. These include phosphoserine (P-Ser) 117 in domain 1, required for phosphoryl transfer, as well as the Mg²⁺-binding loop in domain 2 (Figure III-2A). In dephospho-enzyme, the side chain of Asp263 makes just one direct interaction with another residue: a salt bridge to Arg293, the residue immediately following the conserved Mg²⁺-binding loop. In the phospho-enzyme, Asp263 also makes a salt bridge with Arg293, which in turn contacts P-Ser117. In both structures, Asp263 makes several water-mediated contacts as well, including a 3-way interaction that includes Arg293 and the carbonyl backbone of Pro264.

Both Asp263 and Arg293 are near the active site cleft of PGM1, which is composed of more than 80 residues in total. Arg293 has been proposed to interact with the reaction intermediate [15], while Asp263, which lies further outside the active site, has no known role in catalysis or substrate binding. Both of these residues are highly conserved in the enzyme superfamily, and similar structural interactions between them can be found in multiple related proteins (Figure III-2B). Proteins shown in this superposition are from different subgroups of the large α -D-phosphohexomutase superfamily, which differ in their substrate preferences (see legend of Figure III-2), and come from diverse organisms including bacteria, yeast, and higher eukaryotes. Pairwise sequence

identities between these proteins and human PGM1 range from 22 to 53 %.

Despite considerable overall sequence diversity, the Asp263 - Arg293 interaction

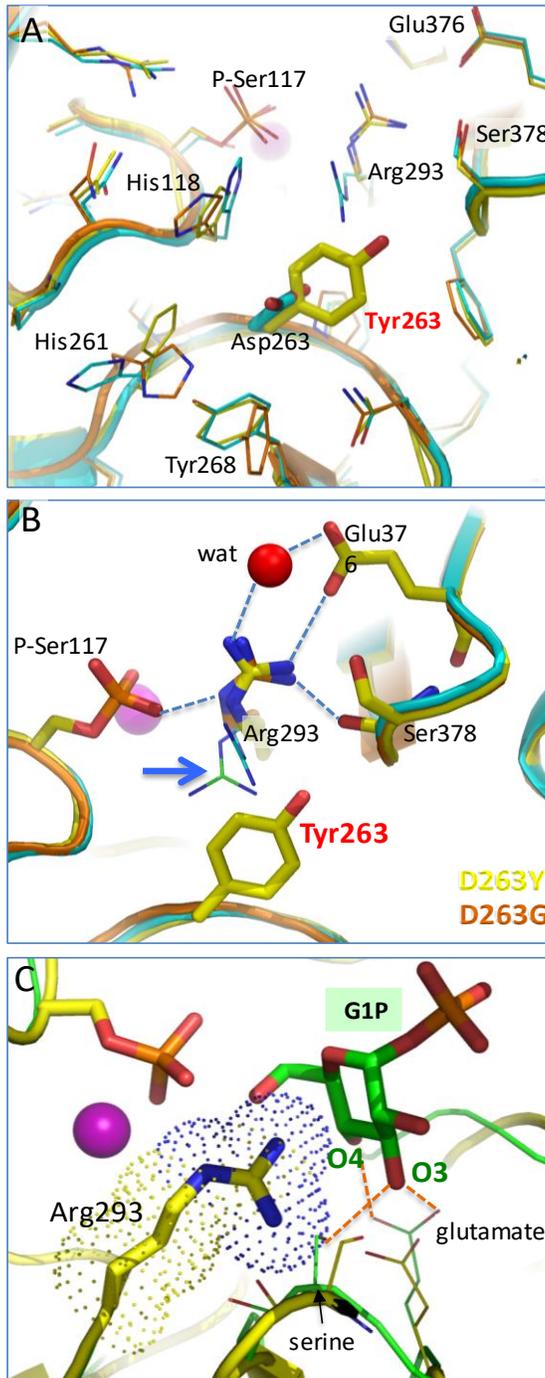


Figure III-3. Structural context of residue 263 in the missense variants of PGM1. (A) The vicinity of residue 263 in the WT PGM1 (cyan), D263G (orange), and D263Y structures (yellow). Conformational changes in the side chains of His118, His261, Tyr268, and Arg293 are observed relative to the WT enzyme. (B) A close-up view of Arg293 in the D263G and D263Y structures (yellow and orange, respectively) showing its distinct conformation relative to WT enzyme. For comparison, the position of Arg293 in the both phospho- (green) and dephospho-versions (cyan) of WT enzyme is shown in thin sticks (see blue arrow). The loss of the conserved interaction between Asp263 and Arg293 causes the arginine side chain to adopt a novel conformation that makes contacts (dashed lines) with two key ligand-binding residues, Glu376 and Ser378. A water that makes a bridging interaction between Arg293 and Glu376 is shown in red. For clarity, side chains are shown only for D263Y, except for Arg293 where both variants are shown. (C) A model of bound glucose 1-phosphate (green) superimposed with the structure of D263Y (yellow). Arg293 is highlighted with a dotted surface, showing potential steric conflicts with the binding of substrate. The glutamate and serine are highly conserved in the enzyme superfamily, suggesting similar ligand-binding roles in PGM1. The model is based on a superposition of an enzyme-ligand complex (green) from a related protein (PDB ID 1P5D); contacts between G1P and the protein are indicated by dashed orange lines.

has been maintained over vast evolutionary distances, as is true for many active site residues in this enzyme superfamily [9].

III.3.2 Structural impact of the Asp263 mutants

The crystal structures of the D263Y and D263G missense variants were determined to 1.85 Å and 2.55 Å, respectively (Table III-1; Experimental Methods). The overall structures of the mutants are nearly identical to that of WT enzyme (Figure III-1B), with a root-mean-square-deviation of ~0.2 Å for both D263Y and D263G for 532-543 C α pairs. Near the site of the mutation, several minor structural changes are seen. The superposition in Figure III-3A highlights these, including small movements of several nearby histidine side chains. Another difference is that Ser117 is found in its phosphorylated state in the mutant structures, as opposed to the dephosphorylated state in the human WT structure (see following section). Overall, these changes from the WT structure appear quite minor, and are not easily correlated with the poor activity of these variants.

The structures of these two mutants, however, do reveal a change with clear significance to enzyme function: reorientation of the side chain of Arg293. As noted above, this is the residue with which Asp263 makes a conserved salt bridge in the structure of WT enzyme. Neither the tyrosine nor glycine of the missense variants is able to make compensating interactions, and, without this, the side chain of Arg293 rotates to a position deep within the active site cleft (Figure III-3B). In the case of D263Y, due to the larger size of the introduced tyrosine relative to aspartate, steric factors could also contribute to the

reorientation of the arginine. However, since a nearly identical conformer of Arg293 is seen in the D263G mutant structure, it appears that loss of the salt bridge with Asp263 is the major determinant of the arginine reorientation.

Table III-1. Data collection and refinement statistics for the Asp263 missense variants

Protein	D263Y	D263G
space group	<i>P</i> 4 ₁ 2 ₁ 2	<i>P</i> 4 ₁ 2 ₁ 2
unit cell parameters (Å)	<i>a</i> = <i>b</i> = 173.0, <i>c</i> = 100.0	<i>a</i> = <i>b</i> = 172.0, <i>c</i> = 99.0
resolution (Å)	57.7–1.75 (1.78–1.75)	57.4–2.5 (2.59–2.50)
observations	2187519	621502
unique reflections	151916	44375
<i>R</i> _{merge} (<i>I</i>)	0.132 (1.587)	0.194 (2.027)
<i>R</i> _{pim} (<i>I</i>)	0.052 (0.657)	0.054 (0.565)
Mean <i>I</i> /σ(<i>I</i>)	15.9 (1.5)	16.3 (1.3)
Mean <i>CC</i> _{1/2}	0.999 (0.648)	0.998 (0.592)
completeness (%)	100.0 (100.0)	85.4 (100.0)
multiplicity	14.4 (13.3)	14.0 (13.7)
no. of protein residues	1112	1113
no. of atoms	10013	8233
no. of sulfate ions	7	4
no. of water molecules	1423	204
<i>R</i> _{cryst}	0.1799 (0.2756)	0.2200 (0.3051)
<i>R</i> _{free}	0.2088 (0.3041)	0.2919 (0.3802)
rmsd bond lengths (Å)	0.008	0.0097
rmsd bond angles (°)	1.089	1.066
Ramachandran plot ^a		
favored (%)	98.0	94.3
outliers (residues)	1	4
MolProbity score (%-tile)	99	86
Average B (Å ²)		
protein	36.8	56.70
water	36.2	42.95
coordinate error (Å) ^b	0.20	0.38
PDB code	5JN5	5TR2

Values for the outer resolution shell of data are given in parentheses. ^aThe Ramachandran plots generated with Molprobity via the PDB validation server. ^bMaximum likelihood-based coordinate error estimate reported by PHENIX.

III.3.3 Arg293 co-opts key interactions needed for binding substrate

Without the salt bridge to Asp263, it is unsurprising that the side chain of Arg293 finds new residue interactions that help neutralize its charge, particularly given the preponderance of positive charge in the active site of PGM1 [8]. The novel, intrasteric (i.e., within the active site) interactions made by the reoriented conformer of Arg293 include contacts to the side chain of Glu376 and the backbone carbonyl of Ser378 (Figure III-3B). These interactions are noteworthy due to the important roles of these residues in substrate binding: Glu376 and Ser378 reside in the highly conserved sugar-binding loop of the enzyme superfamily, and are proposed to make direct contacts to the O3 and O4 hydroxyls of the phosphosugar substrates [7,8]. By comparing the proposed interactions of these residues to substrate (Figure III-3C) with those observed in the structure of the D263Y mutant (Figure III-3B), it can be seen that the reoriented side chain of Arg293 has effectively co-opted interactions with these important ligand binding residues, either directly for Glu376, or indirectly, due to steric hindrance from the nearby backbone contact, in the case of Ser378. The perturbation of the substrate binding site due to the reorientation of Arg293 is consistent with the observed detrimental impact of these variants on kinetic activity [13].

The functional importance of residues in the sugar-binding loop PGM1 is supported by biochemical data, including site-directed mutagenesis and kinetic studies in several members of the superfamily. In phosphomannomutase/phosphoglucomutase from *Pseudomonas aeruginosa*

and PGM from *Acetobacter xylinum*, E→A substitutions at the positions corresponding to Glu376 result in $k_{cat}/K_m < 0.5\%$ that of the WT enzymes [16,17]. This large effect reflects the loss of the structurally conserved bidentate interaction between this glutamate and the O3/O4 hydroxyls of the phosphosugar substrate, observed in several high resolution (2.2 Å or better) crystal structures of enzyme-substrate complexes in related proteins [18,19]. Although crystal structures of enzyme-substrate complexes are not yet available for human PGM1, analogous roles for Glu376 and Ser378 in substrate binding are supported by two unpublished enzyme-ligand complexes for phosphoglucomutase from rabbit PGM PDB ID 1C4G, 1C47 at 2.7 Å resolution.

III.3.4 Correlation with previous biochemical data

Previous studies showed that kinetic impairment of the D263Y and D263G variants is largely due to a reduction in k_{cat} (1.1 – 2.2% that of WT), while K_m is nearly unchanged (~2-fold > WT) [13]. Both of these proteins can also be activated by glucose 1,6-bisphosphate (as measured by phosphorylation of Ser117) at levels similar to that of WT enzyme [13]. In addition, these two variants also have similar melting temperatures to WT enzyme, suggesting their overall ΔG of folding is not significantly changed, and the crystal structures show that the overall structures of the enzyme and active site are largely unperturbed, except for the reorientation of Arg293. Altogether, these data support the notion that these variants would be competent for catalysis upon the occasion that Arg293 moves out of its detrimental position in the active site. The inherent flexibility of the arginine side chain suggests that this could occur, at least

infrequently, explaining their residual activity. Indeed, given that PGM1 is phosphorylated by its reaction intermediate glucose 1,6-bisphosphate, which binds to dephospho-enzyme in a fashion similar to that of substrate, it would be necessary for Arg293 to adopt a WT-like conformer in order for these variants to become phosphorylated in the first place.

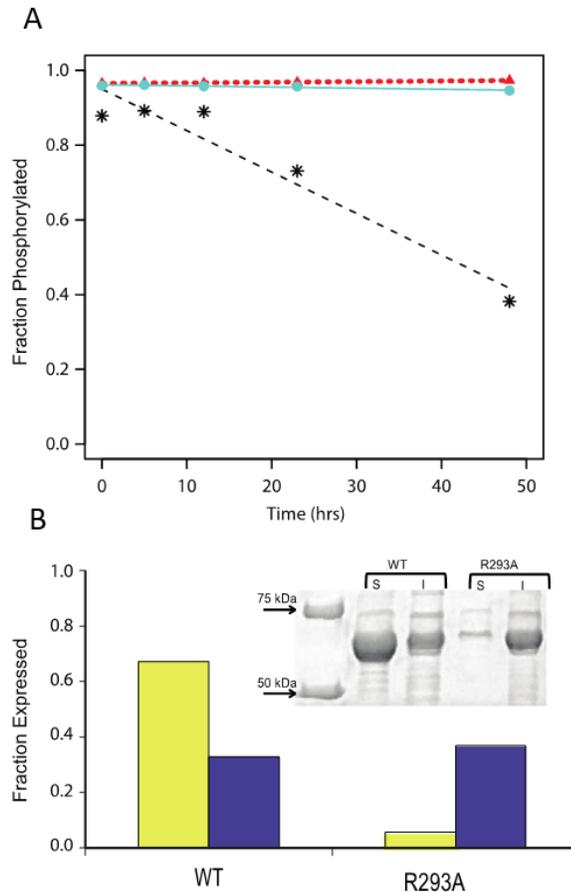


Figure III-4. Biochemical data on phosphorylation and protein expression/solubility. (A) Time course of dephosphorylation from hydrolysis for WT PGM1 (*, dashed line) and the D263Y (circles, solid line) and D263G (triangles, dotted line) missense variants as determined by ESI-MS. (B) Reduced expression and solubility of the R293A mutant relative to WT PGM1. SDS/PAGE (top) and corresponding histogram (bottom) showing the soluble (yellow) and insoluble (purple) fractions of cell extracts from *E. coli* cultures of WT PGM1 and the R293A mutant. Relative protein levels on the histogram are normalized to a value of 1.0 using the combined amounts of soluble/insoluble protein obtained for WT PGM1.

III.3.5 Reduction in spontaneous phosphoryl transfer in missense variants

As noted previously, in the D263Y/G variants, the catalytic phosphoserine (P-Ser117) is observed in its phosphorylated state (occupancy ~0.8-1.0). This is a notable difference from WT human PGM1, where Ser117 completely lacks phosphorylation, when produced using the same purification and crystallization protocols [8]. (The phospho-enzyme version of rabbit PGM in Figure III-2A was produced with different protocols). Based on our anecdotal experience, the ratio of phospho- vs. dephospho-enzyme varies in different samples and batches of recombinantly purified PGM1. However, the large contrast between the D263Y/G mutants and WT enzyme observed in the crystal structures suggested that the missense variants, and the associated novel conformer of Arg293, might affect the phosphorylation level of P-Ser117.

To test this biochemically, hydrolysis of P-Ser117 over time was assessed for WT PGM1 and both the D263Y and D263G variants. [Both time and increased temperature are associated with a reduction in the level of P-Ser of related enzymes [20–24]]. Fully phosphorylated (>90%) versions of WT PGM1 and the D263Y/G variants were prepared using established protocols ([13] and Experimental Procedures). Samples were kept at room temperature for two days, aliquots collected at different time points, and changes in the levels of intact phospho- and dephospho-enzyme estimated using electrospray ionization mass spectrometry (ESI-MS). Figure III-4A shows a shift in relative phosphorylation from ~90% to ~40% over 48 hours for WT enzyme. Strikingly, neither D263Y nor D263G (>95% initial phosphorylation) undergo a significant

change over this time period, consistent with the observation of P-Ser117 in the crystal structures of the D263Y/G variants.

The reduced susceptibility of phosphoserine of the missense variants to hydrolysis is presumably associated with the observed structural perturbation of the active site. This could be due to multiple factors, including a potential increase in the strength of the charge-charge interaction between Arg293 and the P-Ser (due to loss of the Asp-Arg salt bridge), which could serve to stabilize the phospho-enzyme. (While the novel contacts made to Glu376 and Ser378 would appear to somewhat compensate for the missing aspartate, even small changes in bond distances/angles could make these interactions less optimal.) It is also possible that the new conformer observed for Arg293 makes hydrolysis of the phosphoryl group less sterically favorable, either directly or indirectly by precluding a necessary conformational change of the enzyme. A full understanding of the issues involved will require detailed investigations, which might also address whether the rate of phosphoryl transfer to substrate is similarly reduced in these missense variants, a currently unanswered question.

III.3.6 Arg293 is required for correct protein folding

To further explore the role of the Asp263-Arg293 interaction, we characterized an R293A mutant of PGM1. This mutant was constructed and expressed in *Escherichia coli* (Experimental Procedures), as done for the WT enzyme and disease-related missense variants [13]. The R293A mutant shows a moderate reduction in overall protein expression relative to WT enzyme (Figure III-4B). However, a striking decrease in the amount of soluble protein obtained

(only 8% that of WT) is apparent. The soluble fraction of the R293A protein was purified and further characterized by dynamic light scattering (DLS) (Experimental Procedures). The sample indicated a molecular weight (MW) in solution of 102 kD, based on a hydrodynamic radius (R_h) of 4.3 nm. This is larger than the calculated MW for PGM1 of 65 kD, as well as the observed MW from DLS for WT enzyme, which is 72 kD (R_h 3.8 nm) [13]. This result indicates aggregation of purified R293A in solution. As this complicates interpretation of other biochemical experiments, no further studies were conducted.

The impact of the R293A mutation is strikingly different from that of the Asp263 variants, which are both well folded and soluble, suggesting that Arg293 has importance beyond this inter-residue interaction. Indeed, a review of domain-domain interactions in WT PGM1 showed that Arg293 is one of only 13 residues (of 562 total) that participate in a 3-way domain-domain interface [8]. This suggests relevance of Arg293 to interdomain interactions within PGM1, such as maintaining appropriate charge balance between domains during the folding process. A number of missense variants in PGM1 are located in interdomain interfaces of PGM1, and are also known to affect protein folding/solubility [8]. Although no disease-related variants of Arg293 have yet been identified from patients, the observed disruption of protein folding *in vitro* suggests that individuals with mutations at residue 293 would suffer from PGM1 deficiency.

III.4 Discussion

The disease-related missense variants of human PGM1 have varying biochemical phenotypes, which include both those defective in catalysis and those with apparent folding problems [13]. Until recently, no structural information on the molecular basis of enzyme dysfunction for any missense variants of PGM1 was available. However, newly reported crystal structures of two Gly → Arg variants showed significant impacts on enzyme structure, including multiple regions of induced structural disorder [8]. This study showed that the two-category perspective of enzyme dysfunction (defects in catalysis vs. defects in folding) was too simple, with some variants causing problems in both. In contrast to these recent structures, the D263Y/G variants show no new regions of disorder or other long-range impacts on structure. Rather, these two disparate substitutions (tyrosine vs. glycine) have a common, albeit indirect, impact on the enzyme: loss of a conserved interaction with Arg293, with a concomitant rotameric rearrangement of this side chain into the active site. Considering that biochemical studies also show no difference in stability of these variants relative to WT PGM1 [13], dysfunction of the D263Y/G variants appears to be truly catalytic in nature.

In inherited disease, mutants that impair catalysis are less common than those that affect protein folding/stability, but have been found in many systems [for reviews see [25–27]]. Given previous functional knowledge, such as residues comprising the active site, candidates for catalytic mutants are easily proposed based on sequence information. However, catalytic mutants may also

arise from residues beyond those with known direct roles in catalysis (bond making/breaking) or ligand binding, through indirect effects like changing the size of a ligand binding pocket [28] or blocking access to the active site [29]. The missense variants of Asp263 in PGM1 are similar to these, producing an indirect, propagated structural change due to disruption of a finely tuned network of residues surrounding the catalytic cleft of the enzyme. Residues involved in such indirect effects may be difficult to identify, as they may not affect stability, and could even be located outside of key functional regions or not be sequence conserved. In such cases, crystallographic studies of individual disease-related variants (i.e., “personalized biophysics” [27]) may provide many insights.

The Asp263 missense variants of PGM1 cause an intrasteric rearrangement of Arg293 that not only blocks the substrate-binding site, but also co-opts interactions with key ligand-binding residues (e.g., Glu376). Thus, while the arginine side chain is chemically dissimilar from substrate, it nevertheless is able to make interactions with the residues normally responsible for contacting the O3 and O4 hydroxyls of the sugar ring (Figure III-3C). This intrasteric interaction is reminiscent of proteins regulated by autoinhibition through “self interactions” [30]. Although most cases of intrasteric regulation involve domains or groups of residues, some involve more subtle changes, such as the tyrosine residue in the active sites of the insulin receptor tyrosine kinase and MAP kinase ERK2 that moves from its autoinhibitory position upon phosphorylation [30–32]. Although PGM1 is not known to be regulated by autoinhibition, the missense variants at residue 263 mimic this intrasteric effect in principle.

Asp263 is a highly conserved residue in the α -D-phosphohexomutase superfamily, and therefore mutations at this position would typically be considered high impact in terms of disease potential [33]. Moreover, the loss of salt-bridges, as seen in the D263Y/G variants, is also associated with disease-causing mutations [25]. However, an appreciation of the underlying mechanism of enzymatic dysfunction, namely the reorientation of Arg293 with its resulting intrasteric interactions and effects on phosphoryl transfer, would be not be forthcoming from computational or modeling studies. Similarly, the common structural impacts of these two missense variants would tend to give different predicted effects due to the significant physicochemical differences between tyrosine and glycine [25]. Our recent studies of other PGM1 missense variants that cause long-range, induced structural disorder [8] also provided unique and otherwise unpredictable insights into the molecular basis of disease. Together, these studies further emphasize the value of direct structural characterization of missense variants to the field of inherited disease.

III.5 Experimental Procedures

III.5.1 Mutagenesis and protein expression

Construction and expression of the D263Y and D263G missense variants of PGM1 were previously described [13]. The R293A mutant was prepared similarly, transformed into *E. coli* BL21(DE3), and protein expression and solubility assessed in parallel with WT enzyme, following the protocol in [13]. After induction of cell cultures, 1.5 ml of each was centrifuged, and pellets were

resuspended in 200 μ l B-PER Protein Extraction Reagent (ThermoFisher Scientific, Waltham MA, USA) and processed according to manufacturer's instructions. Equal volumes of samples obtained from the soluble and insoluble fractions were mixed with SDS sample buffer and loaded on a 10% SDS/PAGE gel for analysis.

III.5.2 Protein purification and crystallization

PGM1 proteins were purified as previously described [13]. Prior to crystallization, the N-terminal His₆-affinity tags of the missense variant were cleaved using tobacco etch virus protease as in [8]. Purified proteins were dialyzed into a solution of 50 mM MOPS, pH 7.4, with 1 mM MgCl₂, and concentrated to 10-12 mg / mL for D263Y and D263G, and ~ 1 mg / mL for R293A. If not used immediately, samples were flash-frozen in liquid nitrogen and stored at -80° C.

Crystals screens were conducted at a protein concentration of 10-12 mg / mL using the hanging drop vapor diffusion method at 20° C. Drops containing 2 μ l protein solution and 2 μ l crystallization buffer were sealed over a 0.5 mL reservoir. The mutant proteins crystallized in conditions similar to those of WT enzyme: D263Y crystals grew in solutions including lithium or ammonium sulfate (1.4 – 1.55 M) with 0.1 M buffer of either MES, pH 6.0, or Tris HCl, pH 7.5.; crystals of D263G grew from 1.6 M ammonium sulfate, 0.1 M NaCl, and 0.1 M HEPES, pH 7.5. Crystals grew in approximately one week and were cryoprotected using a solution of well buffer with 30% (vol/vol) glycerol, mounted on Hampton loops, and flash cooled in liquid nitrogen. Both variants crystallized

isomorphously with WT enzyme in space group $P4_12_12$ with 60% solvent ($V_M = 3.0 \text{ \AA}^3/\text{Da}$) and two copies of the polypeptide chain in the asymmetric unit.

III.5.3 X-ray diffraction data collection and refinement

Diffraction data were collected at a wavelength of 1.00003 \AA from a single crystal on beamline 4.2.2 of the Advanced Light Source using a Taurus-1 CMOS detector in shutterless mode. Data were processed using XDS [34] and AIMLESS [35] via CCP4i [36]. Data processing statistics are listed in Table III-1. Values of $CC_{1/2} > 0.30$ [37] and R_{pim} [38] were used to determine the high resolution cutoff due to the large number of images (~ 1800 per data set) and high redundancy obtained with the shutterless data collection. The lower resolution diffraction obtained for the D263G variant was due to a smaller crystal size.

Crystallographic refinement calculations were initiated using coordinates of the WT enzyme (PDB ID 5EPC). Refinement was performed with PHENIX [39]; progress was monitored by following R_{free} with 5% of the data set aside for cross validation. The R_{free} data set for the mutants was constrained to match that of the WT diffraction data. The B-factor model consisted of an isotropic B-factor for each atom; TLS refinement was used as automated in PHENIX. COOT [40] was used for model building. The structures were validated using MolProbity [41]. Refinement statistics are listed in Table III-1. Structural figures were prepared with PYMOL [42].

Both mutant proteins are found in as phospho-enzymes (P-Ser117) in the crystal structures. Each monomer contains one bound ion in its metal binding

site, which was modeled as Ca^{2+} in these structures, due to positive electron density remaining after refinement as Mg^{2+} ; it is likely that a mixture of metals derived from the crystallization buffer occupy this site. (The metal ion in the crystal structure is not relevant to kinetic activity of the protein, as assays are always done with Mg^{2+} -containing enzyme.) Seven sulfate ions are included in the final model of D263Y and four in D263G, including two bound in the active site of each chain.

III.5.4 Dynamic light scattering

The R293A mutant was prepared at 1 mg/ml in 50 mM MOPS, pH 7.4, and 1 mM MgCl_2 and centrifuged prior to data collection. Data were collected on a Protein Solutions DynaPro 99 instrument at a wavelength of 8363 Å for at least 200 s (10 s each for 20 acquisitions) at 25° C. Polydispersity of the sample was 23%. R_h was used to estimate MW in solution using software provided with the instrument.

III.5.5 Preparation and assessment of phospho-enzyme

Phospho-enzyme versions of WT PGM1 and the D263Y/G variants were prepared as previously described [13]. Briefly, protein samples at concentrations between 200- 300 μM in 50 mM MOPS, pH 7.4, 1 mM MgCl_2 were incubated with a 6-fold molar excess of the activator glucose 1,6-bisphosphate for 18 h at 4 °C. Glucose 1,6-bisphosphate was subsequently removed via extensive dialysis. This protocol produces ~90% phospho-enzyme [13]. The stability of the phosphoserine to hydrolysis was monitored at 21° C over 48 hours. Aliquots

were removed at various time points for mass spectrometric analyses. For ESI-MS, 10 μL protein samples at 1 pmol/ μL in 1% formic acid were analyzed by nano-LC/nanospray QTOF (Agilent 6520, Agilent Technologies, Santa Clara, CA, USA) in positive ion mode with a Zorbax C8 trap column. Data were examined using the QUAL software provided with the instrument. The mass error between samples is 0.11 Da (2.1 p.p.m.) and the quantification error is 2%. The percentage phospho-enzyme was calculated by normalizing the sum of the dephosphorylated and phosphorylated peak heights to 1.0.

III.6 Author Contributions

K.M.S. and L.J.B designed the experiments; K.M.S., A.C.G., and B.N.K. performed the experiments; L.J.B. wrote the manuscript.

III.7 Acknowledgments

We thank Jay Nix of the Advanced Light Source beamline 4.2.2 for assistance with data collection and processing, Carol Deakyne for helpful discussions, and Brian Mooney of the University of Missouri Charles W. Gehrke Proteomics Center for mass spectrometry. KMS was supported by National Institutes of Health training grant T32 GM008396-26 from NIGMS. This work was supported by grants to LJB from the Patton Trust of the Kansas City Area Life Sciences Research Foundation and the National Science Foundation (MCB-091838[9]). Part of this work was performed at the Advanced Light Source, which is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under contract DE-AC02-05CH11231.

III.8 References

- [1] Timal S, Hoischen A, Lehle L, Adamowicz M, Huijben K, Sykut-Cegielska J, Paprocka J, Jamroz E, van Spronsen FJ, Körner C, Gilissen C, Rodenburg RJ, Eidhof I, Van den Heuvel L, Thiel C, Wevers RA, Morava E, Veltman J & Lefeber DJ (2012) Gene identification in the congenital disorders of glycosylation type I by whole-exome sequencing. *Hum. Mol. Genet.* 21, 4151–4161.
- [2] Stojkovic T, Vissing J, Petit F, Piraud M, Orngreen MC, Andersen G, Claeys KG, Wary C, Hogrel J-Y & Laforêt P (2009) Muscle Glycogenosis Due to Phosphoglucomutase 1 Deficiency. *N. Engl. J. Med.* 361, 425–427.
- [3] Pérez B, Medrano C, Ecay MJ, Ruiz-Sala P, Martínez-Pardo M, Ugarte M & Pérez-Cerdá C (2012) A novel congenital disorder of glycosylation type without central nervous system involvement caused by mutations in the phosphoglucomutase 1 gene. *J. Inherit. Metab. Dis.* 36, 535–542.
- [4] Tegtmeyer LC, Rust S, van Scherpenzeel M, Ng BG, Losfeld M-E, Timal S, Raymond K, He P, Ichikawa M, Veltman J, Huijben K, Shin YS, Sharma V, Adamowicz M, Lammens M, Reunert J, Witten A, Schrapers E, Matthijs G, Jaeken J, Rymen D, Stojkovic T, Laforêt P, Petit F, Aumaitre O, Czarnowska E, Piraud M, Podskarbi T, Stanley CA, Matalon R, Burda P, Seyyedi S, Debus V, Socha P, Sykut-Cegielska J, van Spronsen F, de Meirleir L, Vajro P, DeClue T, Ficicioglu C, Wada Y, Wevers RA, Vanderschaeghe D, Callewaert N, Fingerhut R, van Schaftingen E, Freeze HH, Morava E, Lefeber DJ & Marquardt T (2014) Multiple Phenotypes in Phosphoglucomutase 1 Deficiency. *N. Engl. J. Med.* 370, 533–542.
- [5] Morava E, Wong S & Lefeber D (2014) Disease severity and clinical outcome in phosphoglucomutase deficiency. *J. Inherit. Metab. Dis.* 38, 207–209.
- [6] Wong SYW, Beamer LJ, Gadomski T, Honzik T, Mohamed M, Wortmann SB, Brocke Holmefjord KS, Mork M, Bowling F, Sykut-Cegielska J, Koch D, Ackermann A, Stanley CA, Rymen D, Zeharia A, Al-Sayed M, Marquardt T, Jaeken J, Lefeber D, Conrad DF, Kozicz T & Morava E (2015) Defining the Phenotype and Assessing Severity in Phosphoglucomutase-1 Deficiency. *J. Pediatr.* 175, 130–136.
- [7] Beamer LJ (2015) Mutations in hereditary phosphoglucomutase 1 deficiency map to key regions of enzyme structure and function. *J. Inherit. Metab. Dis.* 38, 243–56.
- [8] Stiers KM, Kain BN, Graham AC & Beamer LJ (2016) Induced structural disorder as a molecular mechanism for enzyme dysfunction in phosphoglucomutase 1 deficiency. *J. Mol. Biol.* 428, 1493–1505.
- [9] Shackelford GS, Regni CA & Beamer LJ (2004) Evolutionary trace analysis of the alpha-D-phosphohexomutase superfamily. *Protein Sci.* 13, 2130–2138.

- [10] Ondruskova N, Honzik T, Vondrackova A, Tesarova M, Zeman J & Hansikova H (2014) Glycogen storage disease-like phenotype with central nervous system involvement in a PGM1-CDG patient. *Neuro Endocrinol Lett* 35, 137–141.
- [11] Loewenthal N, Haim A, Parvari R & HersHKovitz E (2015) Phosphoglucomutase-1 deficiency: Intrafamilial clinical variability and common secondary adrenal insufficiency. *Am. J. Med. Genet. Part A* 167, 3139–3143.
- [12] Wong SY-W, Beamer LJ, Gadomski T, Honzik T, Mohamed M, Wortmann SB, Brocke Holmefjord KS, Mork M, Bowling F, Sykut-Cegielska J, Koch D, Ackermann A, Stanley CA, Rymen D, Zeharia A, Al-Sayed M, Marquardt T, Jaeken J, Lefeber D, Conrad DF, Kozicz T & Morava E (2016) Defining the Phenotype and Assessing Severity in Phosphoglucomutase-1 Deficiency. *J. Pediatr.*
- [13] Lee Y, Stiers KM, Kain BN & Beamer LJ (2014) Compromised catalysis and potential folding defects in in vitro studies of missense mutants associated with hereditary phosphoglucomutase 1 deficiency. *J. Biol. Chem.* 289, 32010–32019.
- [14] Liu Y, Ray WJ & Baranidharan S (1997) Structure of rabbit muscle phosphoglucomutase refined at 2.4 Å resolution. *Acta Crystallogr D* 53, 392–405.
- [15] Dai J-B, Liu Y, Ray WJ & Konno M (1992) The crystal structure of muscle phosphoglucomutase refined at 2.7 Å resolution. *J. Biol. Chem.* 267, 6322–6337.
- [16] Schramm AM, Mehra-Chaudhary R, Furdui CM & Beamer LJ (2008) Backbone flexibility, conformational change, and catalysis in a phosphohexomutase from *Pseudomonas aeruginosa*. *Biochemistry* 47, 9154–9162.
- [17] Brautaset T, Petersen SB & Valla S (2000) In vitro determined kinetic properties of mutant phosphoglucomutases and their effects on sugar catabolism in *Escherichia coli*. *Metab. Eng.* 2, 104–114.
- [18] Regni C, Naught L, Tipton PA & Beamer LJ (2004) Structural Basis of Diverse Substrate Recognition by the Enzyme PMM/PGM from *P. aeruginosa*. *Structure* 12, 55–63.
- [19] Nishitani Y, Maruyama D, Nonaka T, Kita A, Fukami TA, Mio T, Yamada-Okabe H, Yamada-Okabe T & Miki K (2006) Crystal structures of N-acetylglucosamine-phosphate mutase, a member of the ??-D-phosphohexomutase superfamily, and its substrate and product complexes. *J. Biol. Chem.* 281, 19740–19747.
- [20] Sarma AVS, Anbanandam A, Kelm A, Mehra-Chaudhary R, Wei Y, Qin P, Lee Y, Berjanskii M V, Mick JA, Beamer LJ & Van Doren SR (2012) Solution

- NMR of a 463-residue phosphohexomutase: domain 4 mobility, substates, and phosphoryl transfer defect. *Biochemistry* 51, 807–819.
- [21] Xu J, Lee Y, Beamer LJ & Van Doren SR (2015) Phosphorylation in the catalytic cleft stabilizes and attracts domains of a phosphohexomutase. *Biophys. J.* 108, 325–37.
- [22] Ray, William J J, Long JW & Owens JD (1976) An Analysis of the Substrate-Induced Rate Effect in the Phosphoglucomutase System. *Biochemistry* 15, 4006–4017.
- [23] Ray WJ & Long JW (1976) Thermodynamics and mechanism of the PO₃ transfer process in the phosphoglucomutase reaction. *Biochemistry* 15, 3993–4006.
- [24] Lee Y, Furdui CM & Beamer LJ (2017) Data on the phosphorylation state of the catalytic serine of enzymes in the α-D-phosphohexomutase superfamily. *Data Br.* 10, 398–405.
- [25] Kucukkal TG, Petukh M, Li L & Alexov E (2015) Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. *Curr. Opin. Struct. Biol.* 32, 18–24.
- [26] Stefl S, Nishi H, Petukh M, Panchenko AR & Alexov E (2013) Molecular Mechanisms of Disease-Causing Missense Mutations. *J. Mol. Biol.* 425, 3919–3936.
- [27] Kroncke BM, Vanoye CG, Meiler J, George AL & Sanders CR (2015) Personalized Biochemistry and Biophysics. *Biochemistry* 54, 2551–2559.
- [28] Karakas E, Wilson HL, Graf TN, Xiang S, Jaramillo-Busquets S, Rajagopalan K V. & Kisker C (2005) Structural insights into sulfite oxidase deficiency. *J. Biol. Chem.* 280, 33506–33515.
- [29] Decroos C, Bowman CM, Moser JS, Christianson KE, Deardor M a & Christianson DW (2015) Compromised Structure and Function of HDAC8 Mutants Identified in Cornelia de Lange Syndrome Spectrum Disorders. *ACS Chem. Biol.* 9, 2157–64.
- [30] Kobe B & Kemp BE (1999) Active site-directed protein regulation. *Nature* 402, 373–376.
- [31] Hubbard SR, Wei L, Ellis L & Hendrickson WA (1994) Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature* 372, 746–54.
- [32] Zhang F, Strand a, Robbins D, Cobb MH & Goldsmith EJ (1994) Atomic structure of the MAP kinase ERK2 at 2.3 Å resolution. *Nature* 367, 704–711.
- [33] Petukh M, Kucukkal TG & Alexov E (2015) On human disease-causing amino acid variants: Statistical study of sequence and structural patterns. *Hum. Mutat.* 36, 524–534.

- [34] Kabsch W (2010) Software XDS for image rotation, recognition and crystal symmetry assignment. *Acta Crystallogr., Sect. D Biol. Crystallogr.* 66, 125–132.
- [35] Evans PR & Murshudov GN (2013) How good are my data and what is the resolution? *Acta Crystallogr. Sect. D, Biol. Crystallogr.* 69, 1204–14.
- [36] Potterton E, Briggs P, Turkenburg M & Dodson E (2003) A graphical user interface to the CCP 4 program suite. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 59, 1131–1137.
- [37] Karplus PA & Diederichs K (2012) Linking crystallographic model and data quality. *Science* 336, 1030–1033.
- [38] Weiss MS (2001) Global indicators of X-ray data quality. *J. Appl. Crystallogr.* 34, 130–135.
- [39] Adams PD, Afonine P V., Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC & Zwart PH (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66, 213–221.
- [40] Emsley P & Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 60, 2126–2132.
- [41] Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS & Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D, Biol. Crystallogr.* 66, 12–21.
- [42] DeLano WL (2002) The PyMOL Molecular Graphics System. *Schrödinger LLC wwwpymol.org* Version 1., <http://www.pymol.org>.

Chapter IV

A hotspot for disease-associated variants of human PGM1 is associated with impaired ligand binding and loop dynamics

Adapted with permission from Kyle M. Stiers, Lesa J. Beamer, *Structure* 26 (10), 1337-1345. e3. Copyright 2018 Elsevier.

IV.1 Summary

Human phosphoglucomutase 1 (PGM1) plays a central role in cellular glucose homeostasis, catalyzing the conversion of glucose 1-phosphate and glucose 6-phosphate. Recently, missense variants of this enzyme were identified as causing an inborn error of metabolism, PGM1 deficiency, with features of a glycogen storage disease and a congenital disorder of glycosylation. Previous studies of selected PGM1 variants have revealed various mechanisms for enzyme dysfunction, including regions of structural disorder and side chain rearrangements within the active site. Here, we examine variants within a substrate-binding loop in domain 4 (D4) of PGM1 that cause extreme impairment of activity. Biochemical, structural, and computational studies demonstrate multiple detrimental impacts resulting from these variants, including loss of conserved ligand binding interactions and reduced mobility of the D4 loop, due to perturbation of its conformational ensemble. These potentially synergistic effects make this conserved ligand-binding loop a hotspot for disease-related variants in PGM1 and related enzymes.

IV.2 Introduction

The effects of disease-related variants on the structure and function of

human proteins are often not fully understood [1]. Because of the large number of such variants, comprehensive experimental investigation is prohibitive, leading to the widespread use of computational algorithms for predicting impact [2,3]. Even in systems where experimental studies are initiated, many variants prove intractable due to problems with protein folding, stability, or cellular trafficking [4,5]. In a few cases, however, biochemical and atomic resolution structural studies of disease-related variants have proven feasible, shedding light on the molecular basis of disease. One such example is a key metabolic enzyme in humans known as phosphoglucomutase 1 (PGM1), which is responsible for maintaining cellular glucose homeostasis. Approximately 20 missense variants of this enzyme are associated with inherited PGM1 deficiency, an autosomal recessive disease with characteristics of both a glycogen storage disorder (GSD XIV, OMIM 614921) and a congenital disorder of glycosylation of types I and II [6–13]. Recent X-ray crystallographic studies of several PGM1 missense variants revealed distinct mechanisms for enzyme dysfunction, providing insights well beyond and sometimes differing from their predicted effects [14,15]. This work highlights the importance of direct structural characterization for understanding disease-related variants and the value of human PGM1 as a model system for such studies.

The identification of PGM1 deficiency as an inherited disease also provides opportunities for understanding this critical enzyme and its role in human metabolism. PGM1 is a cytosolic enzyme found in most tissues [16]. Its catalytic reaction entails two consecutive phosphoryl transfers and proceeds via

a bisphosphorylated intermediate (Figure IV-1A). The first phosphoryl transfer takes place from a conserved phosphoserine residue (Ser117) to substrate, creating glucose 1,6-bisphosphate. The intermediate must then reorient in the active site, allowing the second phosphoryl transfer to occur from the intermediate to the protein, creating product and regenerating the active (phosphorylated) version of the enzyme. PGM1 has 562 residues, comprising

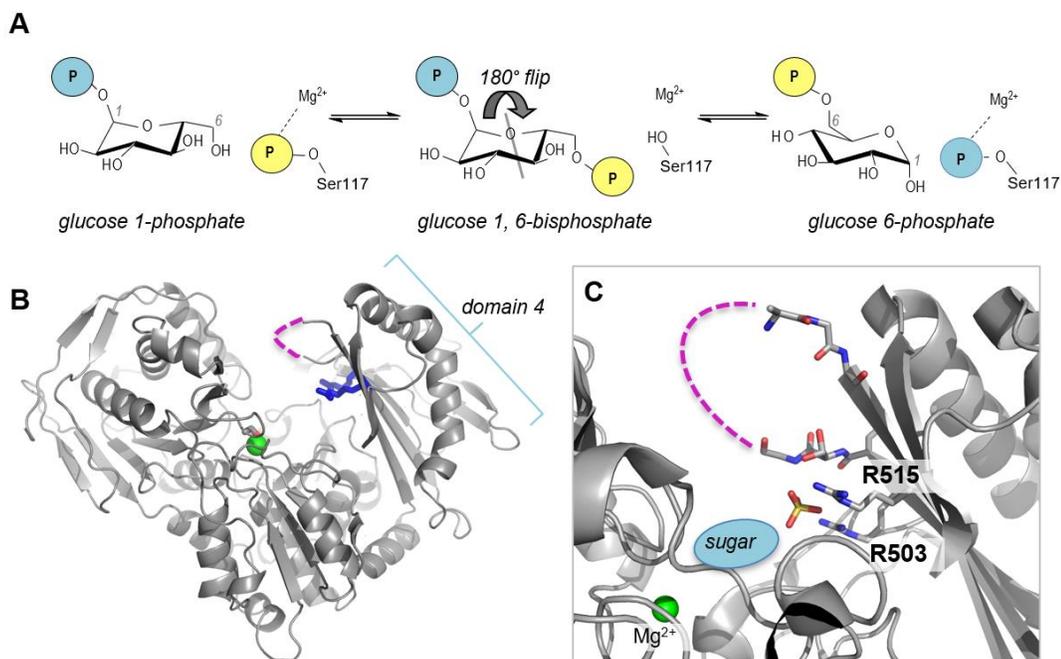


Figure IV-1. Overview of the mechanism and structure of human PGM1. (A) A schematic of the catalytic reaction, showing the reversible conversion of glucose 1-phosphate to glucose 6-phosphate. Glucose 1,6-bisphosphate undergoes a 180° reorientation in between the two phosphoryl transfer steps of the reaction (gray line indicates axis of rotation). (B) The crystal structure of WT human PGM1. Ser117, Arg503 and Arg515 are highlighted as sticks; bound metal ion is shown as black sphere. The missing residues in the D4 loop are shown with dashed line. (C) A close-up of the active site of PGM1. The bound sulfate ion in WT enzyme that acts as a structural mimic for the phosphate group of the substrate is shown with spheres; bound glucose 6-phosphate (in yellow, this report) is superimposed. The Mg^{2+} ion near the site of phosphoryl transfer is shown for reference. Missing residues (507-509) in the D4 loop are shown with dashed line.

four domains of approximately equal size (Figure IV-1B). Its active site is found within the large central cleft and has four conserved loops with critical functional roles: i) the phosphoserine loop that participates in phosphoryl transfer; ii) a loop that binds a Mg^{2+} ion needed for activity; iii) a sugar-binding loop; and iv) the phosphate-binding loop that interacts with the phosphate group of the substrates (see [17] for a detailed review).

In the present study we focus on PGM1 variants affecting residues within loop (iv) in the C-terminal domain (domain 4 or D4) of the enzyme: the 13-residue D4 loop. This region is composed of two β -strands and a connecting loop, and includes a number of conserved residues with proposed roles in binding the phosphate group of substrate, a key determinant of ligand recognition for the entire α -D-phosphohexomutase enzyme superfamily [18]. In the 1.85 Å crystal structure of wild-type (WT) PGM1, three residues within this loop (507-509) were missing in the electron density maps, consistent with high mobility [15]. This suggested a potential functional role during catalysis, whereby the mobile flap would close upon substrate binding and open to release product. Movement of the flap could also be necessary to permit the 180° reorientation of the intermediate in the midst of the catalytic cycle (Figure IV-1A).

The functional relevance of the mobile flap and the impacts of missense variants affecting conserved ligand-binding residues in the D4 loop are further investigated herein, using biochemical, structural and computational approaches. Variants affecting this region, identified either from patients with PGM1 deficiency or from exome sequencing databases, are shown to have multiple detrimental

effects on the enzyme, culminating in a profound decrease in catalytic activity. These studies provide insights into key functional regions of PGM1 and further expand its utility for probing the molecular effects of disease-related mutations.

Table IV-1. Summary of biochemical data on PGM1 loop mutants

Name	Source mutation	Sol. protein (%)	MW (kD)	Polydispersity (%)	Rh (nm)	T _{0.5} (°)	k _{cat} (1/s)	K _m (μM)	k _{cat} / K _m (μM/s)	Phosphorylation before/after G16P (%)
WT ^a	-	90	67	20	3.7	46.3	143 +/- 2	80 +/- 4	1.8 +/- 0.1	35 / 90
R503A	in-house	90	63	21	3.8	n.d.	n.c.	n.c.	n.c.	40 / 85
R503Q ^b	patient	90	72	0	3.8	48.7	n.c.	n.c.	n.c.	25 / 92
G508R	ExAC	80	10	18	4.3	-	-	-	-	34 / 79
G511R	ExAC, Cosmic	80	22	0	6.0	-	-	-	-	13 / 82
R515A	in-house	90	51	19	3.3	n.d.	n.c.	n.c.	n.c.	31 / 87
R515L ^b	patient	75	68	15	3.7	46.8	n.c.	n.c.	n.c.	33 / 49
R515Q	Cosmic	90	53	19	3.3	47.6	n.c.	n.c.	n.c.	34 / 82
R515W	ExAC, Cosmic	70	66	16	3.6	44.9	n.c.	n.c.	n.c.	17 / 36

Abbreviations: n.c. – not characterizable; n.d. – not done; (-) unable to assess due to aggregation; ExAC: Exome Aggregation Consortium [33] ; Cosmic: Catalog of Somatic Mutations in Cancer [34]. ^a Data for WT PGM1 previous published [19] except for T_{0.5} (this study). ^b Kinetic data previously published [11]. Phosphorylation assessed via electrospray ionization mass spectrometry as described in Methods.

IV.3 Results

IV.3.1 Biochemical studies of variants in the D4 loop

Eight mutants affecting the D4 loop of PGM1 were studied (Table IV-1). Two of these were identified from patients affected by PGM1 deficiency: R503Q and R515L [11]. Four others (G508R, G511R, and R515Q/W) were identified as likely detrimental to enzyme function, from a survey of human PGM1 variants found in various genome databases (Methods and [16]). Two additional mutants, R503A and R515A, were engineered to further examine the effects of the amino

acid side chain at these residue positions. Mutant proteins were expressed recombinantly in *E. coli* and purified as described in Methods. Results of the biochemical characterization are summarized in Table IV-1. Soluble protein was obtained for all of the variants, most at levels similar to that of WT enzyme. Potential aggregation was assessed with dynamic light scattering. Six of the eight variants showed hydrodynamic radii (R_h) consistent with the expected molecular weight of the protein. However, G508R and G511R had larger R_h and correspondingly increased estimated molecular weights in solution. Due to apparent aggregation, neither of these variants was studied further.

The six remaining variants were characterized for thermal stability, enzyme activity, and the competence of their catalytic serine for phosphorylation by the known activator and reaction intermediate glucose 1,6-bisphosphate (Table IV-1) [19]. As assessed by thermal shift assays, none of the variants showed significant changes in stability relative to WT enzyme: $T_{0.5}$ values range from 44.9 to 48.7° C, generally within 1-2° C of WT enzyme ($T_{0.5} = 46.3^\circ$ C). In contrast, all six variants had profoundly impaired enzyme activity. While low phosphoglucomutase activity could be observed over extended time (hours) at very high enzyme concentration (100 times WT), it was too low to permit characterization of the steady state kinetic parameters. For comparison, other missense variants with k_{cat} as low as 0.1% of the WT enzyme are still quantifiable [19]. Despite their impaired activity, all of the variants were nevertheless competent for phosphorylation of the catalytic serine (P-Ser117) as assessed by electrospray ionization mass spectrometry (Figure IV-S1), although

for several this was less efficient than for WT enzyme (Table IV-1) under the conditions tested. Thus, it appears that impaired catalysis in these variants is not

Table IV-2. Crystallographic data collection and refinement statistics

Protein	R503Q	R515L	R515Q	R515W	G6P complex
Space Group	P4 ₁ 2 ₁ 2				
Unit cell parameters (Å)	172.1 172.1 99.8	171.3 171.3 99.4	174.0 174.0 99.5	172.3 172.3 99.4	171.9 171.9 99.4
Resolution (Å)	60.95 (1.95)	60.68 (2.20)	61.5 (2.60)	60.93 (1.75)	60.82 (2.30)
Observations	1557274 (74492)	893507 (42745)	660415 (62010)	2043537 (59107)	878402 (37536)
Unique reflections	108994 (5314)	75286 (4379)	47453 (4566)	150294 (6946)	64976 (3790)
Rmerge (I)	0.14 (1.902)	0.279 (2.141)	0.185 (1.846)	0.089 (1.58)	0.160 (1.774)
Rpim (I)	0.038 (0.526)	0.084 (0.72)	0.052 (0.518)	0.025 (0.561)	0.045 (0.573)
Mean I/σ(I)	15.1 (1.5)	9.8 (1.2)	15.2 (1.4)	20.1 (1.1)	14.8 (1.3)
CC _{1/2}	0.999 (0.709)	0.997 (0.693)	0.998 (0.796)	0.999 (0.595)	0.998 (0.656)
Completeness (%)	100.0 (99.9)	100.0 (99.9)	100.0 (100.0)	99.7 (93.6)	97.8 (82.6)
Multiplicity	14.3 (14)	11.9 (9.8)	13.9 (13.6)	13.6 (8.5)	13.5 (9.9)
No. of protein residues	1126	1128	1124	1125	1128
No. of atoms	9725	9316	8281	9898	9259
No. of sulfate ions	18	6	13	5	0
No. of water molecules	949	635	91	1180	571
Rcryst	0.1693	0.1897	0.2386	0.1652	0.1674
Rfree	0.2129	0.2507	0.2954	0.1909	0.2263
RMSD bond lengths (Å)	0.007	0.007	0.010	0.006	0.007
RMSD bond angles (°)	0.812	0.853	1.384	0.783	0.894
Ramachandran Plot					
Favored (%)	98.03	96.44	96.78	98.21	96.80
Outliers (%)	0.18	0.18	0.63	0.09	0.18
MOLPROBITY score (%-tile)	98	98	98	99	99
Average B (Å ²)	41.82	52.17	71.27	32.07	41.11
Protein	40.57	52.03	70.94	30.31	40.89
Water	45.53	50.04	52.38	42.13	42.48
Sulfate ions (/no.)	112.54/18	108.43/6	129.77/13	67.15/5	-
Glucose 6-P (/no.)	-	-	-	-	79.21/2
Coordinate Error (Å)	0.197	0.295	0.454	0.211	0.270
PDB code	5VG7	5VEC	5VIN	5VBI	6BJ0

Values for the outer resolution shell of data are given in parentheses. ^aRamachandran plots generated with Molprobit via the PDB validation server. ^bMaximum likelihood-based coordinate error estimate reported by PHENIX.

attributable to lack of phosphorylation of Ser117, which is required but not sufficient for enzyme activity.

IV.3.2 Arg503 and Arg515 variants cause unexpected ordering of the D4 loop

The crystal structures of four variants of human PGM1, with substitutions at either residue 503 or 515, were determined (Methods). The proteins crystallized under similar conditions as WT enzyme and with isomorphous crystal lattices (Table IV-2). Two identical copies of the polypeptide chain are found in the asymmetric unit: chain A is more highly ordered in all PGM1 crystals, and is therefore used as the reference in following descriptions. Resolution of the diffraction data for the different variants ranged from 1.75 to 2.6 Å.

A superposition of the polypeptide backbones of the four mutants with that of WT enzyme shows that the structures are very similar overall (Figure IV-2A). For example, the overall root-mean-square-deviation (RMSD) between R503Q and WT enzyme is 0.51 Å for 540 C_α pairs. In each structure, the substituted residue (i.e., Gln, Trp, Leu) is observed in the electron density maps and makes no direct interactions with other residues in the protein. Nevertheless, all of the structures show one clear difference compared to WT enzyme: ordering of the mobile flap (residues 507-509) within the D4 loop (Figure IV-2B-E). In the two high-resolution structures of R503Q and R515W, all residues of the loop are clearly defined (Figure IV-2D-E). In the lower resolution structures, partial ordering is seen (Figure IV-S2A), which is nonetheless significant relative to the WT structure in 2F_o-F_c maps (see Figure IV-S2B for omit maps). The conformers

of the D4 loops vary somewhat in the different structures (Figure IV-2B), but include additional inter-residue interactions and regions of ordered secondary structure. For instance, in the R503Q and R515W variants, residues 509-510

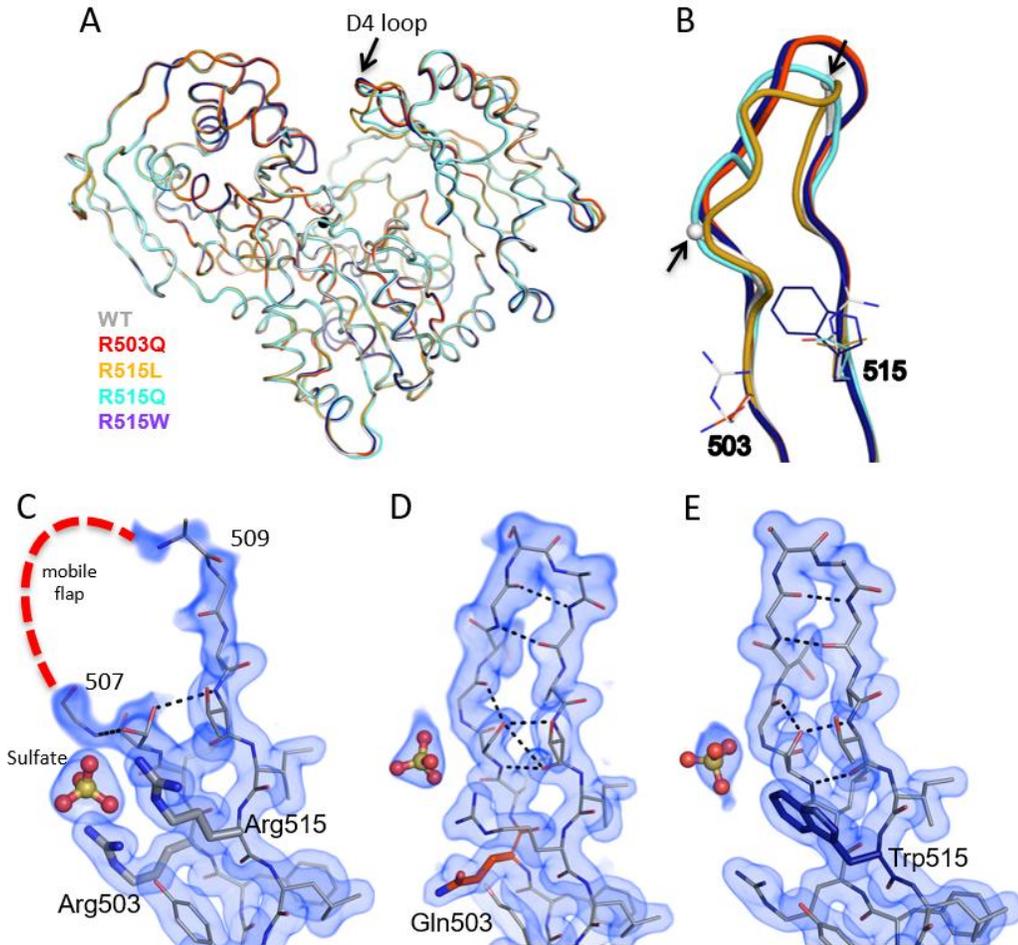


Figure IV-2. Crystal structures of missense variants of Arg503 and Arg515. (A) Backbone superposition of WT PGM1 (gray) with R503Q (red), R515L (gold), R515W (purple), R515Q (cyan), and the enzyme-substrate complex with G6P (green). (B) A close-up view of the D4 loop of each structure in (A). The break in the WT chain is indicated with spheres and highlighted by arrows. $2F_o - F_c$ electron density maps calculated from the final models and contoured at 1.0σ for the D4 loop in (C) WT PGM1, and the (D) R503Q and (E) R515Q variants. For omit maps, see Figure IV-S2B.

form a β -turn connecting two β -strands, with residues 505-507 forming a bulge within this region (Figure IV-2D,E).

A structural explanation for the reduced mobility of the flap in these missense variants is not immediately obvious. In all cases, the site of the mutation (residue 503 or 515) is outside of the flap, and none of the introduced side chains in the variants interact with residues in the flap. The ordering of the flap also occurs regardless of the type of substitution (e.g., Trp vs. Leu). Indeed, even alanine mutants at these positions result in highly impaired enzyme activity, arguing against steric effects from the introduced side chains (Table IV-1). Taken together, these observations raised the following possibility: perhaps the conformer(s) observed in the variants exist within the WT conformational ensemble of this region, but become more energetically favorable in the missense variants. A corollary to this hypothesis would be that the active site environment of the WT enzyme is poised to encourage many possible conformers of the D4 loop, presumably for functional reasons. Given the known association of the D4 loop with ligand binding, we pursued the crystal structure of a PGM1-substrate complex.

IV.3.3 The D4 loop changes conformation to bind ligand

The crystal structure of PGM1 in complex with glucose 6-phosphate (G6P) was determined at 2.3 Å resolution (Table IV-2). (Since the enzyme reaction is reversible, it is equally valid to consider G6P as either substrate or product.) G6P was soaked into crystals of WT enzyme grown in a low salt condition (Methods). Clear evidence for ligand binding was observed in initial electron

density maps; an omit map calculated from the final refined structure is shown in Figure IV-S2B. Differences between the polypeptide backbones of apo- and ligand-bound PGM1 are minor, with some small shifts in residues of domains 1 and 4 (Figure IV-3A). G6P is bound deep within the active site, with approximately 50% of its surface area inaccessible to solvent (Figure IV-3B). Notably, the flap region of the D4 loop is mostly ordered in the ligand complex, although its conformation is distinctly different from that seen in the missense variants (Figures IV-2A and IV-S2B). Significant changes in ϕ/ψ angles occur, especially for Gly506, Gly508, and Ala510, producing an overall closure of the loop. The D4 loop in the enzyme-G6P complex has no regular secondary

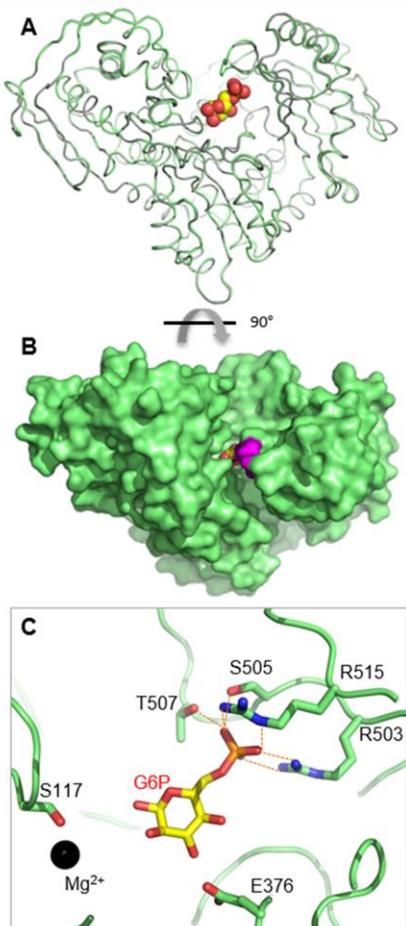


Figure IV-3. Crystal structure of the PGM1 complex with glucose 6-phosphate. (A) A backbone superposition of the apo-enzyme (gray) and enzyme-ligand complex (green). G6P is shown in a space-filling model. (B) Space-filling representation of the PGM1-G6P complex, looking down into the active site. View is a 90° rotation relative to (A). Residues in the mobile flap are in magenta. (C) A close-up view of the PGM1-G6P interactions involving residues in the D4 loop (dashed orange lines). An additional interaction with the backbone amide of Gly506 is not shown. Ser117 and the bound metal ion (black sphere) are shown for reference. For an omit map of the ligand, see Figure IV-S2B.

structure, but is rather wrapped around the phosphate group of ligand (Figure IV-3C). Unlike the R503Q and R515W complexes, no inter-residue contacts are observed between residues in the loop, highlighting the significant conformational change that would be needed to convert between the ligand-bound and ordered conformers of the variants.

Within the active site, G6P binds with its O1 hydroxyl near Ser117 (the catalytic phosphoserine), as would be required for phosphoryl transfer, and with its phosphate group proximal to the D4 loop (Figure IV-3C). As expected [17], multiple residues in the D4 loop, including one from the flap, are involved in direct interactions with G6P. These include both Arg503 and Arg515, which make bidentate interactions with the phosphate, as well as additional phosphate contacts made by Ser505, Gly506, and Thr507. Interactions are also made with the hydroxyl groups of the sugar by Arg293, Glu376, and Ser378, although these vary somewhat in the two copies of the polypeptide chain. Some of the enzyme-ligand interactions in the PGM1-G6P complex are similar to those in two unpublished structures of rabbit PGM (97% sequence identity to human enzyme) bound to ligands, but analysis of the electron density maps of these structures indicates multiple problems, including incorrectly modeled D4 loops (Figure IV-S3), so detailed comparisons are not possible.

The structure of the PGM1-G6P complex highlights the crucial role of the D4 loop in ligand binding, and also helps explain the detrimental impact of the 503 and 515 missense variants. First, the essential roles of Arg503 and Arg515 in contacting the phosphate group of the substrate are clear: these two residues

are part of an intricate network of interactions between the protein and bound ligand (Figure IV-3C), and it seems certain that substitution of either residue would disrupt binding. Support for this is evident in the changed position (or loss) of a bound sulfate ion, a mimic for the phosphate group of the substrate, seen in the crystal structures of missense variants as compared to WT enzyme (Figure IV-2C-E and IV-S2A). Second, the conformer of the D4 loop present in the G6P complex is incompatible with the conformer(s) of D4 observed in the mutant structures (Figure IV-S2B), which would present an additional impediment to ligand binding. Finally, the ordering of the D4 loop observed in the mutant structures, relative to its mobility in WT enzyme, suggests that an unfavorable conformational transition would be necessary to achieve the productive ligand-binding conformer of the loop in the variant proteins. Overall, observations from the PGM1-G6P complex support the functional importance of flap mobility, and suggest that the conformers observed in the missense variants would be unfavorable for ligand binding.

IV.3.4 Evaluation of loop mobility by molecular dynamics

To further examine the intrinsic mobility of the D4 loop, we conducted 10 ns molecular dynamics (MD) simulations of both WT PGM1 and the R503Q variant (Methods). The proteins remained stable throughout the course of the simulation, with RMSD values of the structures generally between 1-2 Å (Figure IV-S4A,B). An analysis of the C α root mean square fluctuations (RMSF), relative to the energy-minimized starting structures, shows that the D4 loop (residues 505-511) is the most highly mobile region of the protein, even in the R503Q

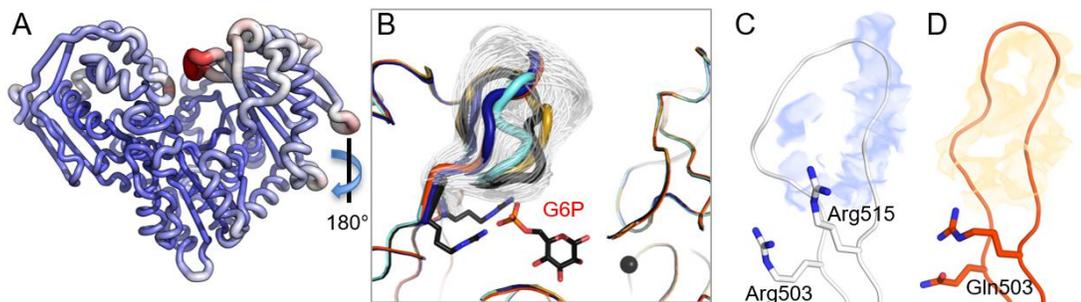


Figure IV-4. Analyses of D4 loop flexibility based on the MD simulations of human PGM1. (A) RMSF values for WT enzyme mapped onto its structure. Increasing tube radius and a change from blue to red indicate higher RMSF values. (B) A sampling of the D4 loop conformers from the MD trajectory (one structure shown per 10 ps cycle) of WT PGM1. Sampled conformers (white) are shown only for the D4 loop; other regions of the polypeptide backbone are from the crystal structures of the missense variants and PGM1-G6P complex. Colors are as in Figure IV-2. View is 180° rotation relative to (A). Residence density analysis of the MD trajectories for the D4 loop (residues 505-513) of WT PGM1 (C) and the R503Q variant (D). Maps were calculated as described in Methods. Note the discontinuity of the density for WT D4 loop (blue) versus the continuous density for the R503Q loop (orange).

variant (Figures IV-4 and IV-S4C-D). (A previous study of PGM1 mechanism by MD showed a similar result for WT enzyme [20]). We further examined the range of conformers of the D4 loop sampled during the MD simulation, and find that the conformers observed in the crystal structures of the missense variants and PGM1-G6P complex are encompassed within this range (Figure IV-4B).

The RMSF analyses suggest generally similar behavior of WT PGM1 and the R503Q variant during the MD simulations, in contrast to the obvious differences seen in the electron density maps of this region. To resolve this apparent paradox, we utilized calculations of the residence density [21], a time-averaged distribution of atoms within a particular volumetric region over the course of the simulation (Methods). These calculations reveal distinct differences between the D4 loop in these two proteins (Figure IV-4C,D). In the case of WT enzyme, the loop has discontinuous density, showing that these residues sample

a wide range of conformers during the simulations, consistent with the lack of electron density for this region in the WT structure. On the other hand, the D4 loop of the R503Q variant has well-defined residence density that generally matches the conformer observed in its crystal structure. The density calculations show that, despite the overall high RMSF of the D4 loop in the R503Q variant (Figure IV-S2B), this region tends to occupy fewer conformational states for relatively longer times, as compared to WT enzyme. Thus, the single amino change of the missense variant perturbs the conformational ensemble of the D4 loop.

IV.4 Discussion

The D4 loop is one of four highly conserved regions of the active site of phosphoglucomutases. Its sequence conservation is evident across a wide range of organisms, including bacteria, plants, and animals. In particular, the residues corresponding to Arg503 and Arg515, as well as the subsequent –Thr-Gly-Ser-, are nearly invariant across evolutionary space (Figure IV-5). This is not surprising, based on the critical role of these residues in ligand binding, as confirmed in the first PGM1-substrate complex presented here. The importance of this region in substrate recognition was previously recognized in related enzymes in the α -D-phosphohexomutase superfamily, where the corresponding loop was shown to interact with the phosphate group of the ligand regardless of whether the 1- or 6-phosphosugar is bound [22,23]. These critical protein-phosphate contacts allow enzymes in this superfamily to accommodate both 1-

and 6-phosphosugars in the same binding site, which is key to their reversible reaction (Figure IV-1).

The profound functional impairment of the Arg503 and Arg515 variants is consistent with the critical roles of these two residues in phosphate recognition. Nevertheless, the PGM1 variants at residues 503 and 515 appear to have no *direct* structural impact on the enzyme, beyond the site of substitution.

Unexpectedly, however, we find that mutations at these residue positions result in the ordering of the nearby, mobile flap of D4. This *indirect* effect, which occurs regardless of the site or nature of the substitution, suggests an additional contributing factor to enzyme dysfunction in these variants: changes to the

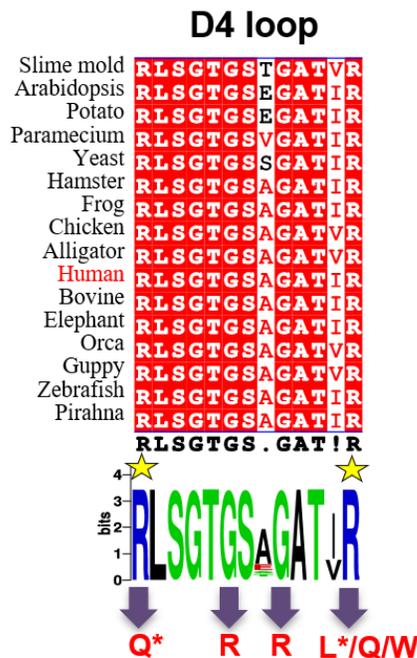


Figure IV-5. Sequences of the D4 loop of phosphoglucomutase in diverse eukaryotic organisms. Top: a multiple sequence alignment (spanning residues 503 to 515 of human PGM) highlighting identical residues with red background. Bottom: A consensus Web Logo [32] of the D4 loop. R503 and R515 are highlighted by yellow star; variants relevant to this study are indicated with arrows at bottom. Those with confirmed roles in PGM1 deficiency are marked with an asterisk.

functional dynamics of this key ligand-binding loop. This proposal is supported by the MD simulations, which show that the R503Q mutation shifts the conformational ensemble of the D4 loop, producing a favored, major conformer that is incompatible with ligand-binding. Ordering of this loop might also impede access of the substrate to the active site, the release of product, and reorientation of the intermediate (Figure IV-3C).

Such potentially compounding effects may help explain the prevalence of missense variants in the D4 loop that are associated with PGM1 deficiency [16]. In addition to those identified in patients (R503Q and R515L), we show here that other variants found in the human population (G506R, G511R, R515W, R515Q) also deleteriously impact PGM1 function *in vitro*. While these variants have not yet been identified in patients with PGM1 deficiency, our results suggest they would likely cause disease in homozygous individuals. Although not yet biochemically or structurally characterized, disease-associated variants in the D4 loop have also been identified in a related human enzyme, PGM3, consistent with the conserved functional role of this region [16,24]. Thus, this region may be a hotspot for disease-associated mutations in the entire enzyme superfamily.

The importance of missense variants in inherited human disease is well appreciated, but a molecular understanding of the effects of the mutation is often lacking [1]. Despite their functional impairment, the D4 variants described herein lack the obvious structural impacts seen in several other PGM1 missense variants, including several that cause direct and profound structural changes within the active site [15]. Instead, the effects of the D4 variants appear to be

explained by the loss of key ligand interactions, with a potentially compounding, indirect effect on loop dynamics. Dynamic changes due to mutations, which act at both short and long-range, have been previously observed, in both computational and experimental studies [25–28]. The possible contribution of a dynamic defect in PGM1 deficiency may facilitate alternative therapeutic options, such as the use of small molecules to modulate the conformational ensemble of the protein [29–31].

IV.5 Acknowledgments

We thank Abigail Graham and Zachary Addison for assistance with protein purification and crystallization, Jay Nix of the Advanced Light Source beamline 4.2.2 for assistance with data collection and processing, Ritcha Mehra-Chaudhary of the University of Missouri Molecular Interactions Core for assistance with protein purification, and Brian Mooney of the University of Missouri Charles W. Gehrke Proteomics Center for mass spectrometry. KMS was supported by National Institutes of Health training grant T32 GM008396-26 and a predoctoral fellowship 17PRE33400210 from the American Heart Association. This work was supported by grants to LJB from the University of Missouri Research Board and the National Science Foundation (MCB-0918389). Molecular dynamics studies were performed on the high-performance computing infrastructure provided by Research Computing Support Services and in part by the National Science Foundation under grant number CNS-1429294 at the University of Missouri, Columbia MO. Part of this work was performed at the Advanced Light Source. The Advanced Light Source is supported by the

Director, Office of Science, Office of Basic Energy Sciences, of the U.S.
Department of Energy under contract DE-AC02-05CH11231.

IV.6 Author Contributions

KS conducted the experiments; KS and LB designed the experiments; LB wrote the paper.

IV.7 Declaration of Interests

The authors declare that they have no conflict of interest.

IV.8 References

- [1] S. Stefl, H. Nishi, M. Petukh, A.R. Panchenko, E. Alexov, Molecular Mechanisms of Disease-Causing Missense Mutations, *J. Mol. Biol.* 425 (2013) 3919–3936. doi:10.1016/j.jmb.2013.07.014.
- [2] F. Gnad, A. Baucom, K. Mukhyala, G. Manning, Z. Zhang, Assessment of computational methods for predicting the effects of missense mutations in human cancers., *BMC Genomics.* 14 Suppl 3 (2013) S7. doi:10.1186/1471-2164-14-S3-S7.
- [3] J. Thusberg, A. Olatubosun, M. Vihinen, Performance of mutation pathogenicity prediction methods on missense variants, *Hum. Mutat.* 32 (2011) 358–368. doi:10.1002/humu.21445.
- [4] P. Yue, Z. Li, J. Moulton, Loss of Protein Structure Stability as a Major Causative Factor in Monogenic Disease, *J. Mol. Biol.* 353 (2005) 459–473. doi:10.1016/j.jmb.2005.08.020.
- [5] N. Tokuriki, D.S. Tawfik, Stability effects of mutations and protein evolvability, *Curr. Opin. Struct. Biol.* 19 (2009) 596–604. doi:10.1016/j.sbi.2009.08.003.
- [6] L.C. Tegtmeyer, S. Rust, M. van Scherpenzeel, B.G. Ng, M.-E. Losfeld, S. Timal, et al., Multiple Phenotypes in Phosphoglucomutase 1 Deficiency, *N. Engl. J. Med.* 370 (2014) 533–542. <http://www.nejm.org/doi/abs/10.1056/NEJMoa1206605>.
- [7] B. Pérez, C. Medrano, M.J. Ecaj, P. Ruiz-Sala, M. Martínez-Pardo, M. Ugarte, et al., A novel congenital disorder of glycosylation type without central nervous system involvement caused by mutations in the phosphoglucomutase 1 gene, *J. Inher. Metab. Dis.* 36 (2013) 535–542. doi:10.1007/s10545-012-9525-7.
- [8] S. Timal, A. Hoischen, L. Lehle, M. Adamowicz, K. Huijben, J. Sykut-Cegielska, et al., Gene identification in the congenital disorders of glycosylation type I by whole-exome sequencing., *Hum. Mol. Genet.* 21 (2012) 4151–4161. <http://www.hmg.oxfordjournals.org/cgi/doi/10.1093/hmg/dd123>.
- [9] N. Ondruskova, T. Honzik, A. Vondrackova, M. Tesarova, J. Zeman, H. Hansikova, Glycogen storage disease-like phenotype with central nervous system involvement in a PGM1-CDG patient, *Neuro Endocrinol Lett.* 35 (2014) 137–141. papers3://publication/uuid/CAB478E7-B251-485D-A996-127520B9E0C1.
- [10] N. Loewenthal, A. Haim, R. Parvari, E. HersHKovitz, Phosphoglucomutase-1 deficiency: Intrafamilial clinical variability and

- common secondary adrenal insufficiency, *Am. J. Med. Genet. Part A.* 167 (2015) 3139–3143. doi:10.1002/ajmg.a.37294.
- [11] S.Y.W. Wong, L.J. Beamer, T. Gadomski, T. Honzik, M. Mohamed, S.B. Wortmann, et al., Defining the Phenotype and Assessing Severity in Phosphoglucomutase-1 Deficiency, *J. Pediatr.* 175 (2015) 130–136. doi:10.1016/j.jpeds.2016.04.021.
- [12] A. Kucukcongar, L. Tumer, F. Suheyl Ezgu, C. Seher Ksapkara, J. Jaeken, G. Matthijs, et al., A case with rare type of congenital-disorder of glycosylation: PGM1-CDG, *Genet. Couns.* 26 (2015) 87–90.
- [13] T. Stojkovic, J. Vissing, F. Petit, M. Piraud, M.C. Orngreen, G. Andersen, et al., Muscle Glycogenosis Due to Phosphoglucomutase 1 Deficiency, *N. Engl. J. Med.* 361 (2009) 425–427. <http://www.nejm.org/doi/abs/10.1056/NEJMc0901158>.
- [14] K.M. Stiers, A.C. Graham, B.N. Kain, L.J. Beamer, Asp263 missense variants perturb the active site of human phosphoglucomutase 1 (PGM1), *FEBS J.* 284 (2017) 937–947.
- [15] K.M. Stiers, B.N. Kain, A.C. Graham, L.J. Beamer, Induced structural disorder as a molecular mechanism for enzyme dysfunction in phosphoglucomutase 1 deficiency, *J. Mol. Biol.* 428 (2016) 1493–1505.
- [16] A.G. Muenks, K.M. Stiers, L.J. Beamer, Sequence-structure relationships, expression profiles, and disease-associated mutations in the paralogs of phosphoglucomutase 1, *PLoS One.* 12 (2017) e0183563. doi:<https://doi.org/10.1371/journal.pone.0183563>.
- [17] L.J. Beamer, Mutations in hereditary phosphoglucomutase 1 deficiency map to key regions of enzyme structure and function, *J. Inherit. Metab. Dis.* 38 (2015) 243–256.
- [18] K.M. Stiers, A.G. Muenks, L.J. Beamer, Biology, mechanism, and structure of enzymes in the α -D-phosphohexomutase superfamily, *Adv. Protein Chem. Struct. Biol.* 109 (2017) 265–304.
- [19] Y. Lee, K.M. Stiers, B.N. Kain, L.J. Beamer, Compromised catalysis and potential folding defects in in vitro studies of missense mutants associated with hereditary phosphoglucomutase 1 deficiency, *J. Biol. Chem.* 289 (2014) 32010–32019.
- [20] N. Bras, P.A. Fernandes, M.J. Ramos, S.D. Schwartz, Mechanistic Insights on Human Phosphoglucomutase Revealed by Transition Path Sampling and Molecular Dynamics Calculations, 117 (2018) 1978–1987. doi:10.1002/chem.201705090.
- [21] N. Michaud-Agrawal, E.J. Denning, T.B. Woolf, O. Beckstein, MDAAnalysis: A toolkit for the analysis of molecular dynamics simulations, *J. Comput. Chem.* 32 (2011) 2319–2327. doi:10.1002/jcc.21787.

- [22] C. Regni, L. Naught, P.A.P.A. Tipton, L.J.L.J. Beamer, Structural Basis of Diverse Substrate Recognition by the Enzyme PMM/PGM from *P. aeruginosa*, *Structure*. 12 (2004) 55–63. doi:10.1016/j.str.2003.11.015.
- [23] Y. Nishitani, D. Maruyama, T. Nonaka, A. Kita, T.A. Fukami, T. Mio, et al., Crystal structures of N-acetylglucosamine-phosphate mutase, a member of the α -D-phosphohexomutase superfamily, and its substrate and product complexes, *J. Biol. Chem.* 281 (2006) 19740–19747.
- [24] Y. Zhang, X. Yu, M. Ichikawa, J.J. Lyons, S. Datta, I.T. Lamborn, et al., Autosomal recessive phosphoglucomutase 3 (PGM3) mutations link glycosylation defects to atopy, immune deficiency, autoimmunity, and neurocognitive impairment, *J. Allergy Clin. Immunol.* 133 (2014) 1400–1409. <http://dx.doi.org/10.1016/j.jaci.2014.02.013>.
- [25] D. Verma, D.J. Jacobs, D.R. Livesay, Changes in lysozyme flexibility upon mutation are frequent, large and long-ranged, *PLoS Comput. Biol.* 8 (2012). doi:10.1371/journal.pcbi.1002409.
- [26] A. Sekhar, J.A.O. Rumfeldt, H.R. Broom, C.M. Doyle, R.E. Sobering, E.M. Meiring, et al., Probing the free energy landscapes of ALS disease mutants of SOD1 by NMR spectroscopy, *Proc. Natl. Acad. Sci.* 113 (2016) E6939–E6945. doi:10.1073/pnas.1611418113.
- [27] G.P. Nicchia, R. Ficarella, A. Rossi, I. Giangreco, O. Nicolotti, A. Carotti, et al., D184E mutation in aquaporin-4 gene impairs water permeability and links to deafness, *Neuroscience*. 197 (2011) 80–88. doi:10.1016/j.neuroscience.2011.09.023.
- [28] B. Wang, X. Wen, X. Qin, Z. Wang, Y. Tan, Y. Shen, et al., Quantitative structural insight into human variegate porphyria disease, *J. Biol. Chem.* 288 (2013) 11731–11740. doi:10.1074/jbc.M113.459768.
- [29] T.O. Street, K.A. Krukenberg, J. Rosgen, D.W. Bolen, D.A. Agard, Osmolyte-induced conformational changes in the Hsp90 molecular chaperone, *Protein Sci.* 19 (2010) 57–65. doi:10.1002/pro.282.
- [30] A.N. Gupta, K. Neupane, N. Rezajooei, L.M. Cortez, V.L. Sim, M.T. Woodside, Pharmacological chaperone reshapes the energy landscape for folding and aggregation of the prion protein, *Nat. Commun.* 7 (2016) 1–8. doi:10.1038/ncomms12058.
- [31] M.C. Encarnación, R.J. Palomino-Morales, J.E. Fuchs, P.G. Esperanza, M.T. Noel, E. Salido, et al., Conformational dynamics is key to understanding loss-of-function of NQO1 cancer-associated polymorphisms and its correction by pharmacological ligands, *Sci. Rep.* 6 (2016) 1–14. doi:10.1038/srep20331.
- [32] G.E. Crooks, G. Hon, J.M. Chandonia, S.E. Brenner, WebLogo: A sequence logo generator, *Genome Res.* 14 (2004) 1188–1190.

- [33] M. Lek, K.J. Karczewski, E. V. Minikel, K.E. Samocha, E. Banks, T. Fennell, et al., Analysis of protein-coding genetic variation in 60,706 humans, *Nature*. 536 (2016) 285–291. doi:10.1038/nature19057.
- [34] S.A. Forbes, D. Beare, H. Boutselakis, S. Bamford, N. Bindal, J. Tate, et al., COSMIC: Somatic cancer genetics at high-resolution, *Nucleic Acids Res.* 45 (2017) D777–D783. doi:10.1093/nar/gkw1121.
- [35] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, et al., Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX*. 1–2 (2015) 19–25. doi:10.1016/j.softx.2015.06.001.
- [36] W.L. DeLano, The PyMOL Molecular Graphics System, Schrödinger LLC [Wwwpymol.org](http://www.pymol.org). Version 1. (2002) <http://www.pymol.org>. doi:citeulike-article-id:240061.
- [37] R Core Team, R: A Language and Environment for Statistical Computing, *R Found. Stat. Comput.* (2014). doi:10.1007/978-3-540-74686-7.
- [38] B.J. Grant, A.P.C. Rodrigues, K.M. ElSawy, J.A. McCammon, L.S.D. Caves, Bio3d: An R package for the comparative analysis of protein structures, *Bioinformatics*. 221 (2006) 2695–2696. doi:10.1093/bioinformatics/btl461.
- [39] G. Andreotti, M. Monticelli, M.V. Cubellis, Looking for protein stabilizing drugs with thermal shift assay., *Drug Test. Anal.* 7 (2015) 831–834. <http://doi.wiley.com/10.1002/dta.1798>.
- [40] W. Kabsch, Software XDS for image rotation, recognition and crystal symmetry assignment, *Acta Crystallogr., Sect. D Biol. Crystallogr.* 66 (2010) 125–132. doi:10.1107/S0907444909047337.
- [41] P.R. Evans, G.N. Murshudov, How good are my data and what is the resolution?, *Acta Crystallogr. Sect. D, Biol. Crystallogr.* 69 (2013) 1204–14. doi:10.1107/S0907444913000061.
- [42] E. Potterton, P. Briggs, M. Turkenburg, E. Dodson, A graphical user interface to the CCP 4 program suite, *Acta Crystallogr. Sect. D Biol. Crystallogr.* 59 (2003) 1131–1137. doi:10.1107/S0907444903008126.
- [43] P.A. Karplus, K. Diederichs, Linking crystallographic model and data quality, *Science (80-.)*. 336 (2012) 1030–1033. doi:10.1016/j.micinf.2011.07.011.Innate.
- [44] M.S. Weiss, Global indicators of X-ray data quality, *J. Appl. Crystallogr.* 34 (2001) 130–135. doi:10.1107/S0021889800018227.
- [45] P.D. Adams, P. V. Afonine, G. Bunkóczi, V.B. Chen, I.W. Davis, N. Echols, et al., PHENIX: A comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66 (2010) 213–221. doi:10.1107/S0907444909052925.

- [46] P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics, *Acta Crystallogr. Sect. D Biol. Crystallogr.* 60 (2004) 2126–2132. doi:10.1107/S0907444904019158.
- [47] V.B. Chen, W.B. Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, et al., MolProbity: all-atom structure validation for macromolecular crystallography., *Acta Crystallogr. Sect. D, Biol. Crystallogr.* 66 (2010) 12–21. doi:10.1107/S0907444909042073.
- [48] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling, *Electrophoresis.* 18 (1997) 2714–2723. doi:10.1002/elps.1150181505.
- [49] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple amber force fields and development of improved protein backbone parameters, *Proteins Struct. Funct. Genet.* 65 (2006) 712–725. doi:10.1002/prot.21123.
- [50] S. Pall, M.J. Abraham, C. Kutzner, B. Hess, E. Lindahl, Tackling exascale software challenges in molecular dynamics simulations with GROMACS, in: *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 2015: pp. 3–27. doi:10.1007/978-3-319-15976-8_1.
- [51] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, et al., GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit, *Bioinformatics.* 29 (2013) 845–854. doi:10.1093/bioinformatics/btt055.
- [52] B. Hess, C. Kutzner, D. Van Der Spoel, E. Lindahl, GRGMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation, *J. Chem. Theory Comput.* 4 (2008) 435–447. doi:10.1021/ct700301q.
- [53] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, GROMACS: Fast, flexible, and free, *J. Comput. Chem.* 26 (2005) 1701–1718. doi:10.1002/jcc.20291.
- [54] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: A message-passing parallel molecular dynamics implementation, *Comput. Phys. Commun.* 91 (1995) 43–56. doi:10.1016/0010-4655(95)00042-E.
- [55] S. Miyamoto, P.A. Kollman, Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models, *J. Comput. Chem.* 13 (1992) 952–962. doi:10.1002/jcc.540130805.
- [56] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.* 126 (2007) 014101. doi:10.1063/1.2408420.

- [57] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, *J. Chem. Phys.* 103 (1995) 8577–8593. doi:10.1063/1.470117.

IV.9 Star Methods

IV.9.1 Key Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Glucose-6-phosphate Dehydrogenase from <i>Leuconostoc mesenteroides</i>	Sigma-Aldrich	CA#: G8404-2KU
□-D-Glucose-1,6-bisphosphate	Sigma-Aldrich	CA#: G6893-25MG
DL-Dithiothreitol	Sigma-Aldrich	CA#: D9779-1G
□-D-Glucose-1-phosphate	Sigma-Aldrich	CA#: G7000-5G
Magnesium sulfate	Sigma-Aldrich	CA#: M7506-500G
NAD ⁺	Sigma-Aldrich	CA#: N0632-5G
Critical Commercial Assays		
Applied Biosystems Protein Thermal Shift Dye Kit	ThermoFisher Scientific	CA#: 4461146
Crystal Screen 1 & 2	Hampton Research	CA#: HR2-110 & HR2-112
Wizard Classic Crystallization Screen 1 & 2	Emerald BioSystems Inc.	CA#: 1009530 & 1009531
QuikChange II Site-Directed Mutagenesis Kit	Agilent	CA#: 200523
Deposited Data		
R503Q variant structure	This paper	PDB: 5VG7
R515L variant structure	This paper	PDB: 5VEC
R515Q variant structure	This paper	PDB: 5VIN
R515W variant structure	This paper	PDB: 5VBI
PGM1-G6P complex structure	This paper	PDB: 6BJ0
Wild-type PGM1	(Stiers et al., 2016)	PDB: 5EPC
Experimental Models: Organisms/Strains		
<i>Escherichia coli</i> : BL21DE3	New England Biosciences	CA#: C2527H
Recombinant DNA		
Plasmid: Wild-type human PGM1	[19]	N/A
Plasmid: PGM1-R503Q missense variant	This paper	N/A
Plasmid: PGM1-R515L missense variant	This paper	N/A
Plasmid: PGM1-R515Q missense variant	This paper	N/A
Plasmid: PGM1-R515W missense variant	This paper	N/A
Software and Algorithms		
GROMACS molecular dynamics software	[35]	http://dx.doi.org/10.1016/j.softx.2015.06.001
PyMOL	[36]	http://www.pymol.org
MDAnalysis python package	[21]	https://doi.org/10.1002/jcc.21787
R	[37]	https://cran.r-project.org/
Bio3D R-package	[38]	https://doi.org/10.1093/bioinformatics/btl461
Other		
VDX™ Plate with sealant	Hampton Research	CA#: HR3-172

IV.9.2 Contact for Reagent and Resource Sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Lesa J Beamer (beamerl@missouri.edu).

IV.9.3 Experimental Model and Subject Details

We used *E.coli* BL21(DE3) cells for recombinant expression of His-tagged PGM1 for biochemical and biophysical experiments. The cells were cultured using standard practices in LB media.

IV.10 Methods Details

IV.10.1 Mutagenesis, protein expression and purification.

Variants R503Q and R515L were identified in patients with PGM1 deficiency and previously shown to lack quantifiable activity [11]. The G508R, G511R, R515Q, and R515W mutants were identified as potentially deleterious to PGM1 function from a computational analysis of known human variants of PGM1 [16] found in the ExAc [33] or Cosmic sequence databases [34]. Each of these variants is rare in the human population. R503A and R515A were engineered for this study. All PGM1 mutants were constructed using the QuikChange kit (Agilent) and verified by automated DNA sequencing.

Missense variants were expressed recombinantly in *E. coli* and purified to homogeneity via an N-terminal His₆ affinity tag, as previously described [19]. The purified proteins were dialyzed into a solution of 50 mM MOPS, pH 7.4, with 1

mM MgCl₂, and concentrated to ~10 mg/mL or higher. If not used immediately, samples were flash-frozen in liquid nitrogen and stored at -80 °C.

IV.10.2 Biochemical studies

Dynamic light scattering—Protein samples at 1 mg/mL in 50 mM MOPS, pH 7.4, and 1 mM MgCl₂ were prepared and centrifuged prior to data collection. Data were collected on a Protein Solutions DynaPro 99 instrument at a wavelength of 8363 Å for 200 s (10 s each for 20 acquisitions) at 25 °C. Polydispersity of samples ranged from 0 to 21%.

Kinetic characterization— Phosphoglucosyltransferase activities for the missense mutants were assessed by coupling the formation of glucose 6-phosphate from glucose 1-phosphate to NADH formation via glucose 6-phosphate dehydrogenase (G6PDH). *Leuconostoc mesenteroides* G6PDH, α-D-glucose 1-phosphate, and glucose 1,6-bisphosphate were obtained from Sigma. Reactions were conducted at 25 °C in 50 mM MOPS, pH 7.4, with 1 mM dithiothreitol, 1.5 mM MgSO₄, and 0.9 mM NAD⁺. The activator glucose 1,6-bisphosphate was present at 1.0 μM and the substrate (α-D-glucose 1-phosphate) at 200 μM. Assays of variants were done in parallel with a control (WT enzyme at 4 - 7.8 nM) and monitored for at least one hour. Enzyme concentrations for the variants at 1X, 10X, and 100X of the WT control were assessed, all of which failed to produce quantifiable activity.

Assessment of phosphorylation by mass spectrometry—The phosphorylation state of the active site phosphoserine (Ser117) was analyzed

before and after treatment with glucose 1,6-bisphosphate using electrospray ionization mass spectrometry (Figure IV-S1). Protein samples at concentrations between 75 and 150 μM in 50 mM MOPS, pH 7.4, 1 mM MgCl_2 were incubated with a 6-fold molar excess of glucose 1,6-bisphosphate for 16 h at 4 °C. For mass spectrometric analyses, protein samples at 1 pmol/ μL in 1% formic acid were analyzed by NanoLC Nanospray QTOF (Agilent 6520) in positive ion mode with a Zorbax C8 trap column. Data were examined using the Qual software provided with the instrument. The mass error between samples is 0.11 Da (2.1 ppm) and quantification error is 2%. Percent phosphorylation was calculated by normalizing the sum of the dephosphorylated and phosphorylated peak heights to 1.0.

Thermal Shift Assays—Dephosphorylated WT PGM1 and missense variants were diluted to 0.5 mg/mL ($\sim 8 \mu\text{M}$) in 50 mM MOPS, pH 7.4 supplemented with 1 mM MgCl_2 . Samples were incubated with dye from the Applied Biosystems® Protein Thermal Shift kit per manufacturer's recommendation for 1 hour at 4 °C. A QuantStudio 3 Real-Time PCR System (ThermoFisher Scientific) was used to ramp from 4 °C to 95 °C in 0.3 °C increments with 10 second holds between ramping steps. Fluorescence values were normalized as in [39] and $T_{0.5}$ calculated as the midpoint of the normalized fluorescence response. Samples were run in duplicate.

IV.10.3 Crystallization

Proteins destined for crystallization trials were purified as above, with the additional step of cleavage of the histidine tag, as in [15]. Successful cleavage

was confirmed by electrospray ionization mass spectrometry. Initial crystallization screens were set up at 19 °C with Crystal Screen kits 1 and 2 (Hampton Research) and Wizard screen kits 1 and 2 (Emerald BioSystems Inc.) at a protein concentration of 9-16 mg/mL using the hanging drop vapor diffusion method at 20° C. Drops containing 2 µl protein solution and 2 µl crystallization buffer were sealed over a 0.5 mL reservoir. Crystals of R503Q used for data collection were grown in 1.8 M ammonium sulfate and 0.1 M TRIS HCl, pH 7.5. Final R515L and R515Q crystals were grown in 0.01 M cobalt (II) chloride hexahydrate, 0.1 M MES monohydrate pH 6.5, and 1.8-2.1 M ammonium sulfate. R515W crystals were grown in 0.1 M HEPES 7.5, 0.1 M sodium chloride, and 1.5-1.7 M ammonium sulfate. Crystals were cryoprotected using a solution of well buffer supplemented with 30% glycerol (v/v), mounted on Hampton loops, and flash cooled in liquid nitrogen. Crystals grew in approximately one week in space group $P4_12_12$ with ~60% solvent ($V_M = 3.04 \text{ \AA}^3/\text{Da}$). The asymmetric unit contains two copies of the polypeptide chain.

For preparation of a PGM1-substrate complex, a novel, low salt crystallization condition was identified for WT PGM1. Crystallization conditions were screened at 19°C using Intelli-Plate 96-3 well sitting drop plates (0.2 µL protein + 0.2 µL well buffer) at a protein concentration of 12 mg/mL. Crystals of WT PGM1 were obtained in condition C12 from PEG Ion Screen HT (Hampton Research), which contains 0.2 M sodium tartrate dibasic dihydrate and 20% PEG 3,350 at pH 7.3. Despite the different crystallization conditions, these crystals were isomorphous with those grown previously from ammonium sulfate [15]

(Table II-1). For ligand soaks, single crystals were extracted from drops, resuspended in well buffer, and the concentration of PEG 3,350 increased to 30%. Step-wise additions of 10 mM glucose-6-phosphate were made to the crystals over a period of five minutes to a final concentration of ~7 mM; crystals were flash-cooled directly into liquid nitrogen for storage without additional cryoprotection.

IV.10.4 X-ray diffraction data collection and refinement

Diffraction data were collected at a wavelength of 1.00003 Å from single crystals on beamline 4.2.2 of the Advanced Light Source using a Taurus-1 CMOS detector in shutterless mode. The data were processed using XDS [40] and AIMLESS [41] via CCP4i [42]. Data processing statistics are listed in Table IV-2. Values of $CC_{1/2} > 0.30$ [43] and R_{pim} [44] were used to determine the high resolution cutoff due to the large number of images (1800-3600 per data set) and high redundancy obtained with the shutterless data collection.

Crystallographic refinement calculations were initiated using coordinates of WT PGM1 (PDB code: 5EPC). Refinement was performed with PHENIX [45]; progress was monitored by following R_{free} with 5% of each data set was set aside for cross validation. The B-factor model consisted of an isotropic B-factor for each atom; TLS refinement was used as automated in PHENIX. COOT [46] was used for model building. The R_{free} data sets for the mutants were constrained to match those of the WT data. Structures were validated using MolProbity [47] and refinement statistics are listed in Table IV-2. In the R503Q and R515W structures, Ser117 was modeled as a mixture of serine and

phosphoserine, based on the electron density maps. Structural figures were prepared with PYMOL [36]. Coordinates and structure factor amplitudes have been deposited in the PDB under the accession number listed in Table IV-2.

IV.10.5 Molecular dynamics simulations and trajectory analyses

The structure of WT PGM1 (PDB ID 5EPC, chain A) and the R503Q mutant (PDB ID 5VG7, chain A) were prepared for MD simulations. Missing side chains were modeled using Swiss-PDBViewer [48]. Initial coordinates for the missing residues in the flap of WT enzyme were taken from the structure of the R503Q mutant. Topologies were generated using the Amber (amber99sbf) force-field [49] with GROMACS 2016.3 [35,50–54]. The system was solvated with the TIP3P water model and neutralized by addition of counter ions in a dodecahedral box [55]. The neutralized system was energy minimized by the steepest descent algorithm with a tolerance of force (F_{\max}) less than $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$, and both converged to an energy below $-1.0\text{e}+06 \text{ kJ mol}^{-1}$. The energy-minimized structures were subjected to consecutive 100 ps NVT and NPT equilibration steps. Each system remained stable at 300 K using a modified Berendsen thermostat, and pressure was maintained using the Parrinello-Rahman coupling method [56]. Production MD simulations were continued from the final frame of the NPT equilibration for 10 ns. Equilibration and production MD used the LINCS constraint algorithm [49] with long-range electrostatics calculated using the Particle Mesh Ewald method [57]. Trajectories were analyzed with the GROMACS energy, rms, and rmsf utilities [35,51]. RMSF

values were mapped to the B-factor field of the initial frame from production MD and visualized in PyMOL [36].

Trajectory density analysis — Raw trajectories were corrected for periodic boundary conditions, rotation, and translation with the trjconv utility of GROMACS [35]. The analysis.density submodule of the MDAnalysis python package [21] was used to generate densities from the trajectories using the default grid spacing of 1.0 Å. Density maps were visualized against the initial production MD frame in PyMOL [36] using the volume representation with carve set to 1.6 Å (default).

IV.11 Quantification and Statistical Analysis

Thermal shift analysis – Relative fluorescence units were normalized to fraction unfolded using the minimum and maximum fluorescence reading for the well as in [39] with the formula $F_{unfolded} = \frac{RFU_i - RFU_{min}}{RFU_{max} - RFU_{min}}$, where $F_{unfolded}$ is the relative fractional portion unfolded at a temperature point, RFU_i is the fluorescence reading at a temperature point, RFU_{min} is the minimum reading for that sample, and RFU_{max} is the maximum reading for that sample. The point at which the protein is half unfolded ($F_{unfolded} = 0.50$), designated $T_{0.5}$, is used for comparison of samples.

IV.12 Data and Software Availability

IV.12.1 Accession Numbers

The atomic coordinates and structure factors for the structures described in this study have been deposited to the RCSB PDB (www.rcsb.org) with the PDB ID codes: 5VG7, 5VEC, 5VIN, 5VBI, and 6BJ0.

IV.12.2 Additional Resources

N/A

IV.13 Supplemental Information

IV.13.1 Supporting Figures

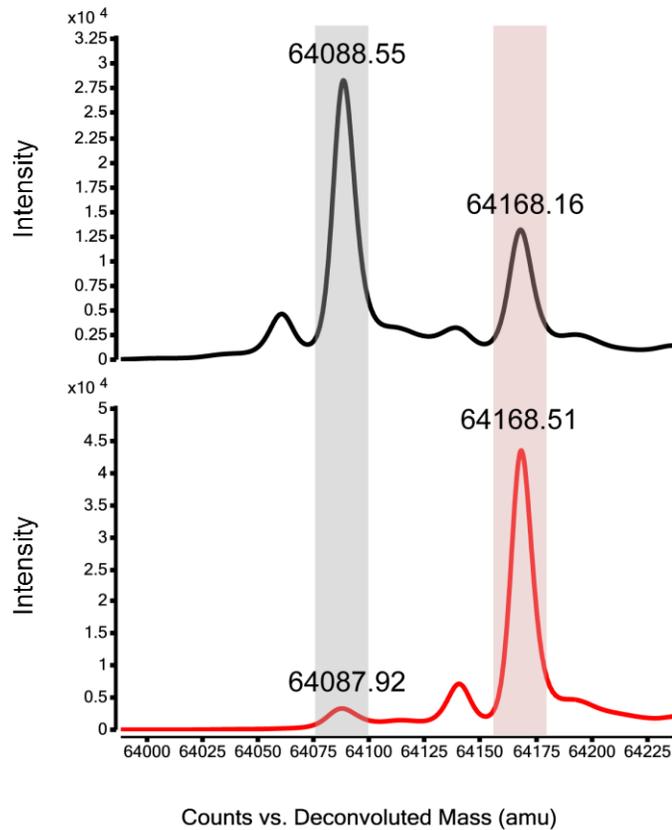


Figure IV-S1. Representative electrospray ionization mass spectra for the R503Q missense variant, related to Methods. The two major peaks are for phospho-enzyme (red shading) and dephospho-enzyme (gray shading). Spectra are shown for the protein as purified (top panel) and after treatment with glucose 1,6-bisphosphate (bottom). Spectra of other variants were highly similar.

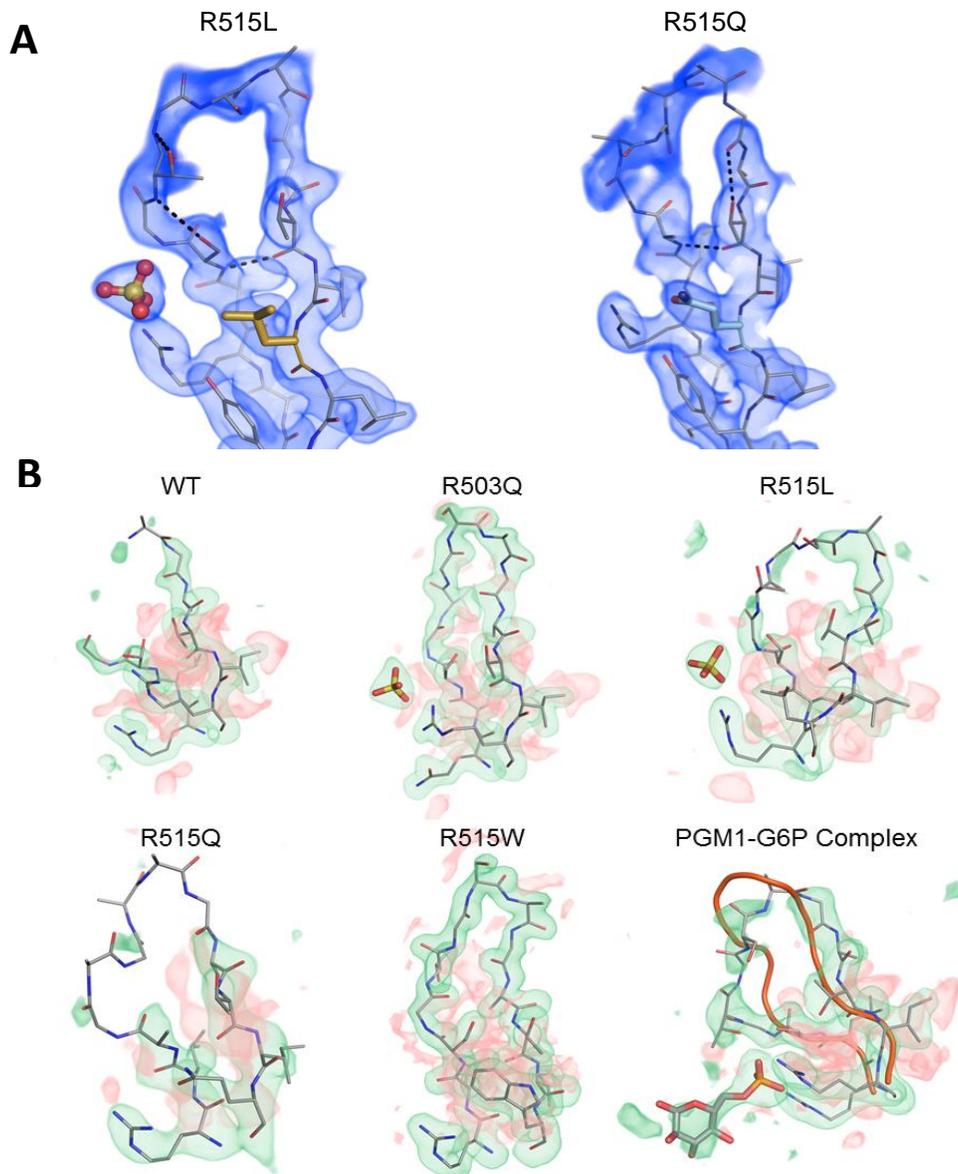


Figure IV-S2. Electron density maps in the vicinity of the D4 loop of PGM1, related to Figure IV-2 and IV-3. (A) Final $2F_o-F_c$ electron density maps contoured at 1.0 σ for the D4 loop of missense variants R515L (*left*) and R515Q (*right*); backbone interactions are indicated by dashed lines. Compare with Figure IV-2C-E in text. (B) Omit maps of the D4 loop for WT PGM1 (5EPC), the four missense variants, and the PGM1-G6P complex. Positive (green) and negative (red) electron density at 3.0 σ is shown for residues 503-515 and the ligand, when present. For the G6P complex, the D4 loop (orange) from the R503Q variant is superimposed, highlighting differences with the ligand-bound conformer. Omit map density at this contour level is clear for the two high resolution structures, R503Q and R515W, and moderate for the R515L structure. Weak density for the R515Q loop may be due to the lower resolution and higher overall B-factor of this data set (Table IV-2). Both positive and negative density is shown for completeness; no negative density is found on atoms of the structures.

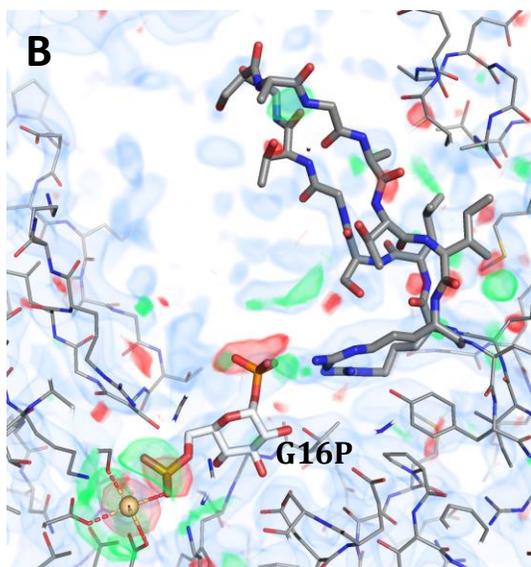
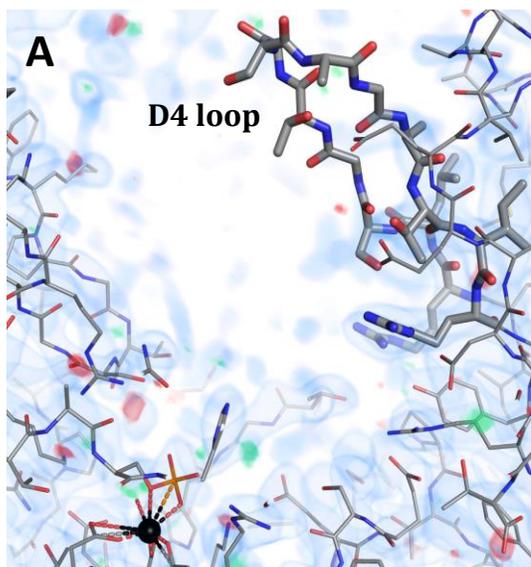
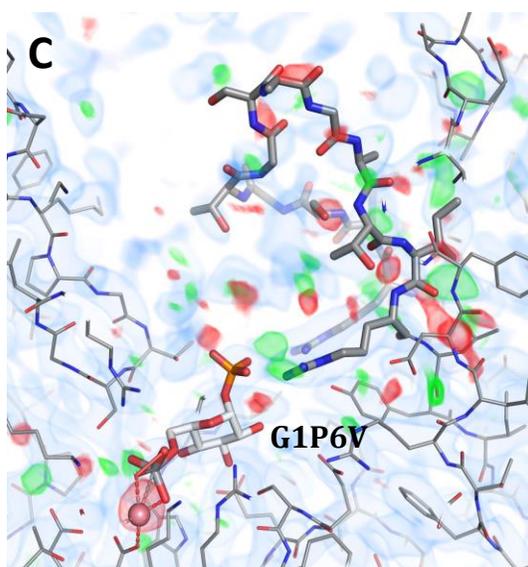


Figure IV-S3. Electron density maps of the D4 loop and active site vicinity of rabbit PGM, related to Figure IV-3. (A) Apo-enzyme (PDB ID 3PMG, 2.4 Å resolution). (B) Enzyme complex with glucose 1,6-bisphosphate (PDB ID 1C47, resolution 2.7 Å). (C) Enzyme complex with glucose 1-phosphate-6-vanadate, resolution 2.7 Å). 2Fo-Fc (light blue) and Fo-Fc (green and red) maps are shown at contours of 1.0 and $\pm 3.0 \sigma$, respectively. Ligands and D4 loop are shown in sticks. All three structures lack supporting density for the model of the D4 loop; multiple other problems in the maps are apparent, including poor density for the entire C-terminal domain of both complexes (not shown). Further examination can be done using the one-click electron density visualization tools on the structure summary



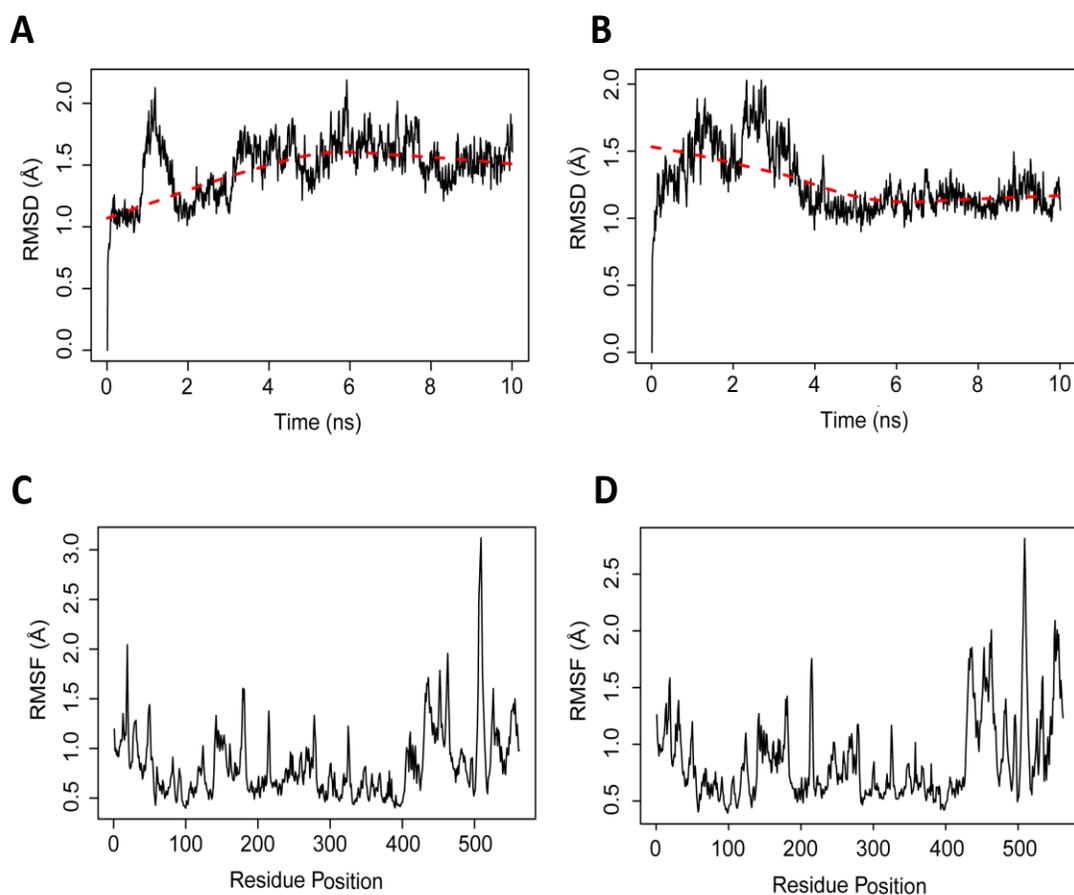


Figure IV-S4. Analysis from 10 ns MD simulations of WT PGM1 and the R503Q missense variant, related to Figure IV-4. The root mean square deviation (RMSD) of the C α atoms is shown with respect to the initial MD frame for (A) WT PGM1 and (B) the R503Q variant. The moving average is shown as a dashed red line. The per-residue root mean square fluctuations (RMSF) for (C) WT PGM1 and (D) the R503Q missense variant calculated after superposing each frames' coordinates to the first frame.

Chapter V

Principal component analysis as a method for understanding ensemble structural changes in proteins

Adapted from Structural and conformational insights into enzyme states along the catalytic cycle of a phosphohexomutase. Structural Dynamics. Submitted.

V.1 Abstract

Principal component analysis (PCA) is a powerful multidimensional reduction method for discovering patterns in complex data. Recent software developments have made PCA reduction easily applicable to many types of data, including three-dimensional (3D) coordinate sets of protein structures. To simplify the use of PCA on ensembles of structures, we have developed a script to generate principal components and useful visualizations so the user can interrogate the significance of the results in their system. No *a priori* knowledge of the system is needed – only the refined structures with matching sets of atoms. Some applications of PCA may include analyzing the effects of mutants on protein structure, binding of ligands or fragment screening, and analyses of oligomeric assemblies.

V.2 Overview of PCA

PCA is a widely used multivariate statistical technique that has found use in almost all scientific disciplines [1]. The method requires a table of observations consisting of variables that are presumably correlated in a non-trivial manner. The goal of PCA is to extract a small number of orthogonally related variables, called principal components (PCs), that represent the same information in a new

coordinate space. The newly identified PCs serve to simplify and highlight the major trends of the input observations. For more information on the PCA method see refs [1–3].

In a simple case, the requirements of PCA are straightforward. I (observations) and J (variables) can be represented as an $I \times J$ matrix, M . Within M , each column will be scaled such that the mean is zero and units are normalized on a fractional basis. (Without this, relatively small changes in one column can dominate the differences seen in the data due to magnitude differences between units.) Scaling can be done by the square root of I or $I-1$ to produce a covariance matrix, or by dividing by the square root of the sum of all the squared elements of the variable to produce a correlation matrix [1]. Either singular value decomposition or eigenvector decomposition of covariance can be computed for the resulting matrix. The scores and individual loadings (i.e., per-residue contribution to each PCA) can be visualized to inform the user of previously hidden features in the data and correlations between variables. Generally, PCA aims to maximize the variance in the data accounted for by an individual PC while minimizing correlation between different PCs.

V.3 Applying PCA to structural ensembles

To conduct PCA on 3D structural coordinates, we utilized the Bio3D [4] package implemented in the R statistical programming language. (We also suggest use of the free, open-source RStudio integrated development environment). Using the attached script (V.6.1), structures were loaded in Protein

Data Bank (PDB) format, aligned, and pairwise root mean square deviation (RMSD) values calculated using the `read.pdb`, `pdbaln`, and `rmsd` functions of the Bio3D package, respectively [4]. PCs were calculated using the `pca.xyz` function of the Bio3D R package [4]. The resulting Scree (cumulative variance) plot was inspected to determine how many components to include in the follow-up analyses. In most cases tested, >75% of the variance in the ensemble was accounted for by the first three PCs. While possible to conduct PCA on proteins of various sizes, it is most straightforward when the number of residues matches, enabling a direct comparison of residue positions. In this way, PCA of structural ensembles can be used as a fast, easy screening tool for elucidating trends in overly complex or otherwise cumbersome ensembles. Follow-up analyses and more directed studies can then be conducted to probe functional groupings or other hypotheses resulting from PCA.

Due to the nature of PCA on coordinate space data, both global and per-residue properties can be analyzed and are automatically generated by the script presented here (V.6.1). Global clustering of coordinate sets can be done using the scores (z , described below) to show structural relationships. Per residue contributions to the variance in the data (per principal component) can be accessed through the atom-wise loadings (au , described below) [4]. Loadings are especially useful as they can be readily mapped to the structure (done automatically by script V.6.1, Step 15) showing the user how much each residue contributes to the different PCs and their correlated grouping. These two

methods for visualizing trends in the 3D coordinate set are likely to be of interest in many structural investigations.

For additional types of analyses, it is useful to know the organization of the dataframe resulting from PCA. Bio3D's PCA function `pca.xyz` produces a dataframe consisting of six attributes named: L, U, z, au, sdev, and mean - corresponding to the eigenvalues, eigenvectors (x, y, and z variable loadings), scores of the coordinates on the PCs, atom-wise loadings (normalized eigenvectors), the standard deviations of the PCs, and the means that were subtracted. Assuming the dataframe was named `pca.xyz` these two variables can be accessed with `pca.xyz$z` and `pca.xyz$au` in R [4]. For a step-by-step walkthrough of the process, see the example script in the supplemental material (V-6.1).

V.3.1 PCA to cluster a structural ensemble into related groups

One application of PCA is to investigate relationships between different conformers in an ensemble of protein structures (i.e., varying conformations of the same protein). As an example of this, we used PCA to examine relationships between crystal structures determined for different states within the catalytic cycle of an enzyme of interest. We recently obtained 15 high resolution structures (Table V-1) of various enzyme states of *Xanthomonas citri* phosphoglucomutase (XcPGM), an enzyme from the α -D-phosphohexomutase superfamily. Figure V-1 illustrates the states on the catalytic cycle and maps them to numbers found in Table V-1. Among others, these include structures of the apo-enzyme, as well as complexes with substrate, product and intermediate.

Previous knowledge of the XcPGM superfamily allows us to sort these structures into functional groups. However, with PCA, an *a priori* approach can be taken, to examine structural relationships without the need for or potential bias from other kinds of information. PCA on an ensemble of XcPGM structures (states 1-12) results in three components that account for 92.7% of the total

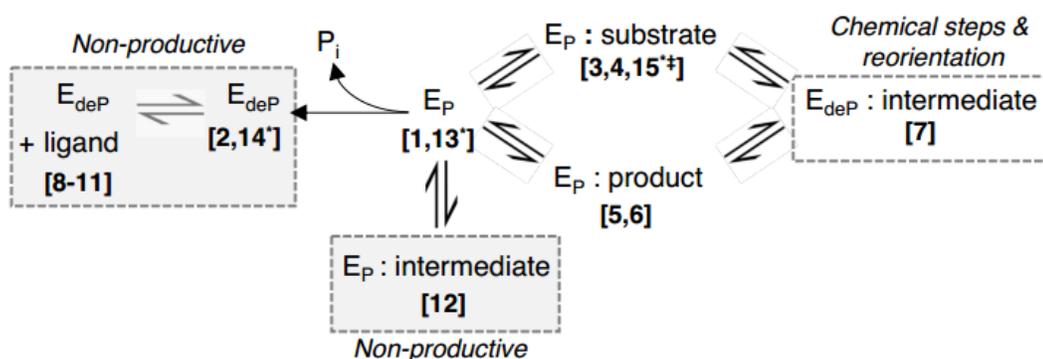


Figure V-1. Outline of the catalytic cycle of XcPGM, highlighting the various enzyme states captured in this study. Structure numbers correspond to those on Table 1. E_{deP} corresponds to enzyme lacking a phosphorylated serine, E_P corresponds to catalytically active enzyme. Structures not part of the normal catalytic cycle are labeled as “non-productive” and highlighted by gray boxes. *Room-temperature data set collected; † complex with a non-hydrolysable G1P analog.

variance in the data set. The normalized scores in each of the top three PCs were plotted (Figure V-2). In the case of PC1, which accounts for 56.3% of the variance, structures are seen to cluster into two distinct groups in PC space. These two clusters represent two functional groupings in our system: apo-enzyme structures (Figure V-2 black triangles) and enzyme-ligand complexes of XcPGM (Figure V-2 circles). These two groupings are consistent with a conformational change of an active site loop that occurs upon ligand-binding (data not shown).

Table V-1. Overview of X-ray data sets collected

	State no.	Ligand	d_{min} (Å)	Mean ADP $\langle \text{Å}^2 \rangle$	R	R_{free}	Coord. error (Å)	PDB ID
CRYO								
Apo-P	1	-	1.44	20.47	0.1708	0.2076	0.17	6NN2
Apo-deP	2	-	1.50	26.99	0.1722	0.2054	0.21	6NN1
Phospho-complexes	3	G1P	1.57	22.91	0.1699	0.2062	0.19	6NNO
	4	M1P	1.61	19.06	0.2070	0.2707	0.29	6NOQ
	5	G6P	1.45	19.73	0.1782	0.2104	0.21	6NNS
	6	M6P	1.41	20.78	0.1795	0.2131	0.23	6NP8
	12	G16P	1.46	19.99	0.1691	0.2071	0.16	6NNU
Dephospho-complexes	8	G1P	1.35	20.89	0.1743	0.2028	0.19	6NNN
	9	M1P	1.73	27.27	0.1905	0.2488	0.35	6NOL
	10	G6P	1.50	24.01	0.1792	0.2112	0.19	6NNS
	11	M6P	1.38	20.03	0.1913	0.2221	0.21	6NPX
	7	G16P	1.45	21.72	0.1689	0.2078	0.17	6NNT
RT								
Apo-P	13	-	1.90	39.94	0.1695	0.2053	0.22	6NQF
Apo-deP	14	-	1.85	35.46	0.1607	0.1873	0.19	6NQE
Complex	15	G1CP	2.05	37.94	0.1823	0.2529	0.39	6NQG

For ligand abbreviations, see text. Cryogenic data sets were collected at -170 °C and RT data sets at 25 °C. Coordinate error calculated by Phenix.

In PC2 (20.5% variance), the scoring of the XcPGM structures is dramatically different (compare positions of analogous symbols), showing the independence of this component from PC1 (Figure V-2). In PC2, the structures are distributed more evenly by score, compared to the clusters seen in PC1 or PC3. PC2 has a less obvious structural basis, but we find that complexes with substrate (1-phosphosugars) and product (6-phosphosugars) tend to fall at different ends of the distribution (bold vs. pastel colors). Moreover, the complexes with mannose phosphosugars (blue) appear to drive the variance in this PC. In other words, the structural differences between the 1- and 6-phosphosugars are more pronounced in the complexes with mannose phosphosugars than those with glucose.

Although the biochemical significance of this result remains unclear, this is an example of how PCA can uncover differences too small to be evident from RMSD values or structural superpositions with visual inspection.

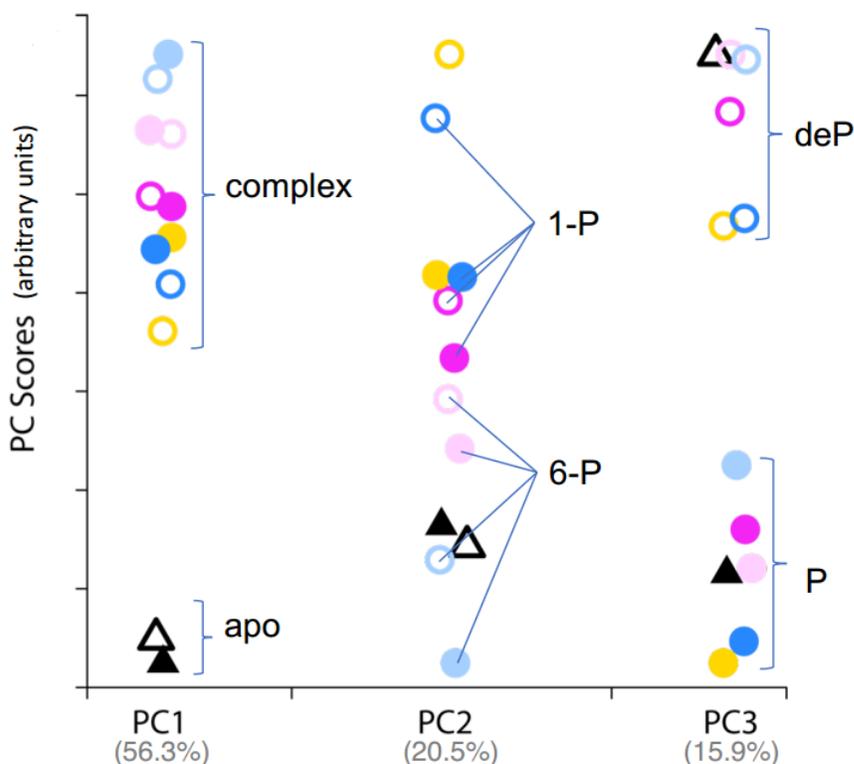


Figure V-2. A scatter plot displaying scores of the various XcPGM structures for the first three principal components of the data set. Apo structures are shown as black triangles; ligand complexes as circles. Phospho- and dephospho-enzyme structures are solid and open symbols, respectively. Complexes with glucose-based sugars are in shades of pink, mannose-based sugars in blue, and G16P complexes in yellow. Bright colors are $E_P:S$ complexes (1-phosphosugars) and pastels are $E_P:P$ complexes (6-phosphosugars). Random noise (jitter) was added to the x-dimension to separate points for ease of visualization; different components have been normalized to the same scale on the vertical axis.

In PC3 (15.9% variance), two clusters are again seen (open vs. solid symbols on Figure V-2). These two clusters separate according to the active (phosphorylated) and inactive (unphosphorylated) states of the enzyme. This is an example of how the structural impact of covalent modification can be assessed by PCA, an especially important consideration in the a-D-phosphohexomutase superfamily where phosphorylation of the catalytic serine is required for catalysis.

V.3.2 PCA to examine residue-specific correlations

In addition to assessing global structural relationships, PCA can provide information on the importance of specific residues to trends within the data set, which can subsequently be grouped into clusters of residues with covarying structural changes. As mentioned above, per residue contributions to each PC can be accessed through the atom-wise loadings. Insight into individual residues or groups of residues having correlated changes within a PC, indifferent to their proximity in 3D space, can be found by mapping these atom-wise loadings to the structure (e.g. V.1.6, Step 15) or by simply plotting the values in a standard line graph.

For this analysis, we utilized an ensemble of four XcPGM structures, all as apo-enzyme (states 1,2,13,14). Both phosphorylated and unphosphorylated states of the enzyme were included, as well as structures determined from X-ray data sets collected at room-temperature and under cryogenic conditions. The first three PCs account for 100% of the variation in the data set; Figure V-3 shows regions of XcPGM with correlated structural changes. PC1 (Figure V-3B)

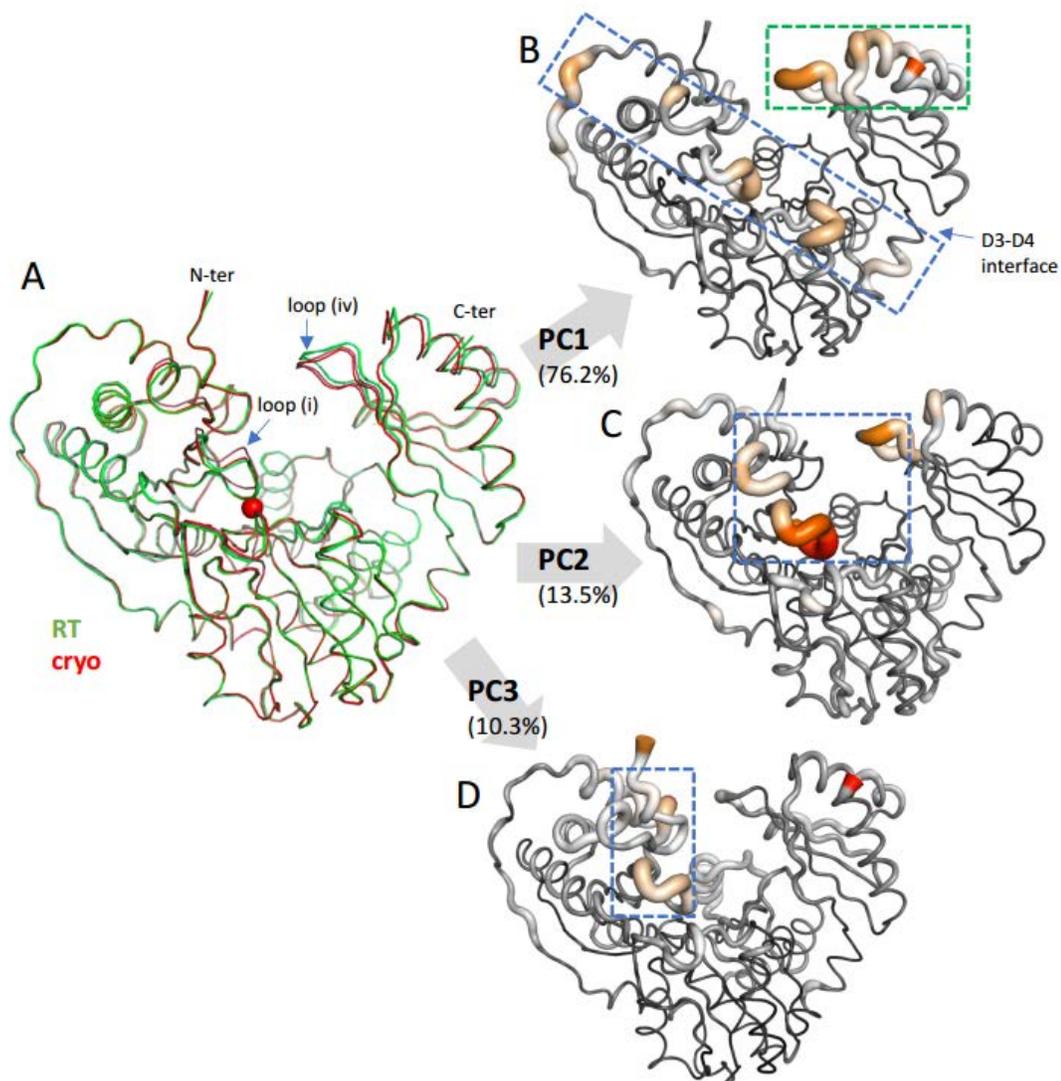


Figure V-3. PCA of an apo-enzyme ensemble of XcPGM comprised of two cryo and two RT structures. (A) A superposition of the four apo-enzyme structures, states [1,2,13,14]. (B-D) Individual residue contributions for the first three principal components plotted on the structure. Magnitude of contribution is indicated by color (low, black; high, red) and tube width. The contribution of each component to the overall variance is indicated in parentheses. Dashed boxes highlight groups of covarying residues (see text).

highlights many residues, spanning multiple domains of the enzyme and connecting key active site loops with residues on the periphery of the structure and in domain interfaces. In PC2 (Figure V-3C), a different type of co-variation between key active site loops (independent of those seen in PC1) is observed, suggesting more than one type of coupled motion in these loops. Finally, we find that the structural variations highlighted in the cryo-RT ensemble are a subset of the variations observed between the different enzyme states at cryogenic temperatures. This suggests that PCA of structures determined at both cryogenic and room temperatures can provide insight into biologically relevant protein conformers, without the need for more complicated computational analyses.

V.3.3 Potential applications of coordinate analysis by PCA

Given that PCA is blind to the type of structural ensemble supplied, virtually any perturbation can be assessed for changes as a function of that perturbation. Because of this, PCA has found widespread utility in the characterization of protein ensembles from NMR [5,6] and molecular dynamics studies [7,8]. However, PCA has not been as commonly used in crystallographic comparisons, which typically involve fewer structures. Use of PCA in 3D coordinate comparisons may have also been limited by the complexity of early implementations, but as we show here, this is no longer a barrier. Some applications of this method might include the analysis of protein oligomers, where changes in packing arrangements may occur, or of varying domain conformers in multi-domain proteins. PCA of crystallographic coordinate sets has particular utility for analysis of prohibitively large structural data sets, such as a recent

study of 41 structures examining the binding of fragments for drug discovery [9]. In such large ensembles, PCA might reveal clusters of related structures not obvious to the human examiner.

V.4 Conclusion

PCA is a well-established technique that when leveraged with recent improvements in implementation, such as the script provided here, can be used as a tool for screening and extracting trends in structural ensembles. Due to its simplification of variance in a dataset, it is capable of identifying both coarse, large-scale changes and subtle tunings of structures in an ensemble. These characteristics enable a wide range of structural questions to be probed by PCA. Furthermore, in situations where the same protein is being interrogated, e.g. binding of different ligands, it is very straightforward to identify regions that move together in 3D space and interpret residue-specific changes. The provided script (See V.6.1) aims to streamline applying PCA to 3D coordinate ensembles and to provide an entry point for working with complex systems or those otherwise too prohibitively large to inspect manually.

V.5 References

- [1] H. Abdi, L.J. Williams, Principal component analysis, Wiley Interdiscip. Rev. Comput. Stat. 2 (2010) 433–459. doi:10.1002/wics.101.
- [2] I.T.J. Springer, Principal Component Analysis, Second Edition, n.d. [http://cda.psych.uiuc.edu/statistical_learning_course/Jolliffe I. Principal Component Analysis %282ed., Springer, 2002%29%28518s%29_MVsa_.pdf](http://cda.psych.uiuc.edu/statistical_learning_course/Jolliffe%20I.%20Principal%20Component%20Analysis%202ed.,%20Springer,%202002%29%28518s%29_MVsa_.pdf) (accessed February 24, 2019).
- [3] C.R. Rao, THE USE AND INTERPRETATION OF PRINCIPAL COMPONENT ANALYSIS IN APPLIED RESEARCH*, n.d. <https://about.jstor.org/terms> (accessed February 24, 2019).
- [4] B.J. Grant, A.P.C. Rodrigues, K.M. ElSawy, J.A. McCammon, L.S.D. Caves, Bio3d: an R package for the comparative analysis of protein structures, Bioinformatics. 22 (2006) 2695–2696. doi:10.1093/bioinformatics/btl461.
- [5] P.W.A. Howe, Principal components analysis of protein structure ensembles calculated using NMR data, J. Biomol. NMR. 20 (2001) 61–70. doi:10.1023/A:1011210009067.
- [6] D.M.A. Gendoo, P.M. Harrison, The Landscape of the Prion Protein's Structural Response to Mutation Revealed by Principal Component Analysis of Multiple NMR Ensembles, PLoS Comput. Biol. 8 (2012) e1002646. doi:10.1371/journal.pcbi.1002646.
- [7] C.C. David, D.J. Jacobs, Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins, in: Humana Press, Totowa, NJ, 2014: pp. 193–226. doi:10.1007/978-1-62703-658-0_11.
- [8] A. Romero-Rivera, M. Garcia-Borràs, S. Osuna, Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity, ACS Catal. 7 (2017) 8524–8532. doi:10.1021/acscatal.7b02954.
- [9] N. Radeva, J. Schiebel, X. Wang, S.G. Krimmer, K. Fu, M. Stieler, F.R. Ehrmann, A. Metz, T. Rickmeyer, M. Betz, J. Winquist, A.Y. Park, F.U. Huschmann, M.S. Weiss, U. Mueller, A. Heine, G. Klebe, Active Site Mapping of an Aspartic Protease by Multiple Fragment Crystal Structures: Versatile Warheads To Address a Catalytic Dyad, (2016). doi:10.1021/acs.jmedchem.6b01195.

V.6 Supplemental Material

V.6.1 PCA script using Bio3D in R

```
#####
##### PCA of multiple structures with Bio3D #####
#####
##### Script written by: Kyle M. Stiers #####
#####
#####
#####
#STEP 1 - Set up work environment
setwd(dir = "C:/Users/username/Desktop/")
require(bio3d)
#STEP 2 - Load in all relevant structures
#Note: This can be done in multiple ways: doing each individually (shown here) or by
looping through a directory
and loading all files that match a pattern (*.pdb), loading many structures based on PDB
ID, etc
apo_dep <- read.pdb(file = "apo-dep_done.pdb",multi = T, ATOM.only = T, rm.insert = T)
apo_pho <- read.pdb(file = "apo-pho_done.pdb",multi = T, ATOM.only = T, rm.insert = T)
g1p_dep <- read.pdb(file = "g1p-dep_done.pdb",multi = T, ATOM.only = T, rm.insert = T)
g1p_pho <- read.pdb(file = "g1p-pho_done.pdb",multi = T, ATOM.only = T, rm.insert = T)
g6p_dep <- read.pdb(file = "g6p-dep_done.pdb",multi = T, ATOM.only = T, rm.insert = T)
g6p_pho <- read.pdb(file = "g6p-pho_done.pdb",multi = T, ATOM.only = T, rm.insert = T)
g16p_dep <- read.pdb(file = "g16p-dep_done.pdb",multi = T, ATOM.only = T, rm.insert =
T)
g16p_pho <- read.pdb(file = "g16p-pho_done.pdb",multi = T, ATOM.only = T, rm.insert =
T)
m6p_dep <- read.pdb(file = "m6p-dep_done.pdb",multi = T, ATOM.only = T, rm.insert =
T)
m6p_pho <- read.pdb(file = "m6p-pho_done.pdb",multi = T, ATOM.only = T, rm.insert =
T)
m1p_dep <- read.pdb(file = "m1p-dep_done.pdb",multi = T, ATOM.only = T, rm.insert =
T)
m1p_pho <- read.pdb(file = "m1p-pho_done.pdb",multi = T, ATOM.only = T, rm.insert =
T)
#####
#Example of how to load structures based on PDB ID
# **THIS IS NOT RUN, IT IS AN EXAMPLE**
#ids <- c("6NN1_A","6NN2_A","6NNN_A","6NNO_A","6NNP_A",
#raw.files <- get.pdb(ids)
#files <- pdbsplit(raw.files, ids)
#####
#STEP 3 - create a list of the loaded structures
current_list <- list(apo_dep, apo_pho, g1p_dep, g1p_pho, g6p_dep, g6p_pho,
g16p_dep, g16p_pho, m6p_dep,
m6p_pho, m1p_dep, m1p_pho)
#STEP 4 - Align the structures sequences
#http://www.drive5.com/muscle/downloads.htm to download the MUSCLE alignment
```

```

binary file for your system.
#Change the path of the exefile argument to match the path to your MUSCLE
executable file
aligned_xanths <- pdbaln(current_list, exefile="C:/Users/kmskvf/Desktop/muscle.exe")
#STEP 5 - Extract names to be used for the files and calculate sequence identity
#Note: change the next line range (1 or 8) to change the length of the substring
extracted
aligned_xanths$id <- substr(basename(aligned_xanths$id),1,8)
aligned_xanths$id
seqidentity(aligned_xanths)
#The function above will print a matrix of the sequence identities to the console
#STEP 6 - Calculate an RMSD matrix for all the structures and write it to a CSV called
"rmsd.csv"
rmsd_matrix <- rmsd(aligned_xanths, fit = T)
rmsd_matrix
write.csv(rmsd_matrix,file = "rmsd.csv")
#STEP 7 - Optional heatmap with clustering in the row direction, and labels for
convenience. Requires the NMF
package (or you to use a different heatmap function)
#If you would like to create the heatmap install the NMF package and uncomment the
next two lines
#library(NMF)
#heatmap(rmsd(aligned_xanths, fit = T), Colv = NA,labRow = c("apo-dep (1)","apo-pho
(2)","g1p-dep (3)","g1ppho (4)", "g6p-dep (5)","g6p-pho (6)", "g16p-dep (7)", "g16p-pho
(8)", "m6p-dep (9)", "m6p-pho (10)", "m1p-dep
(11)", "m1p-pho (12)"))
#STEP 8 - Calculate the 'core' (unchanging portion) of the aligned structures and plot it
in two ways, plain black and
colored red gradient to pink
core <- core.find(aligned_xanths)
plot.core(core)
print.core(core)
col=rep("black", length(core$volume))
col[core$volume<2]="pink"; col[core$volume<1]="red"
plot(core, col=col)
#STEP 9 - (Re)Align the PDBs in sequence and coordinates (in case you skipped to this
point) and calculate the
density histogram as well as a dendrogram (OPTIONAL)
aligned_xanths <- pdbaln(current_list, exefile="C:/Users/kmskvf/Desktop/muscle.exe")
xyz <- pdbfit(aligned_xanths, outpath = cat("\",path,"", "fitted.pdb\"", sep = ""))
rd <- rmsd(xyz)
hist(rd, breaks=10, xlab="RMSD (Å)", main="Histogram of RMSD")
hc.rd <- hclust(as.dist(rd))
aligned_xanths$id <- substr(basename(aligned_xanths$id), 1, 6)
hclustplot(hc.rd, labels=aligned_xanths$id, cex=1.0, colors = c("black","black","green",
"green", "blue","blue",
"orange","orange", "red","red", "purple", "purple"),ylab="RMSD (Å)", main="RMSD
Cluster Dendrogram",
fillbox=FALSE)
#STEP 10 - Calculate the RMSF of the structures and plot it
rf <- rmsf(xyz)

```

```

plot.bio3d(rf,ylab="RMSD (Å)", xlab="Residue No.", typ="l")
#STEP 11 - Perform PCA and plot the overall result
pc.xray <- pca.xyz(xyz)
pc.xray
plot(pc.xray)
#STEP 12 - Plot the principal components against each other (and optionally
interactively identify the points with
the structure name)
plot(pc.xray$z[,1], pc.xray$z[,2], xlab="PC1", ylab="PC2", pch=19, cex=1.75)
abline(h=0,v=0,lty=2,col="gray")
#Uncomment the next line if you'd like to interactively add labels to the points (RStudio is
best for this). Simply
click each point you'd like to label after running the identify() command and click finished
when done.
#identify(pc.xray$z[,1],pc.xray$z[,2], labels=basename.pdb(aligned_xanths$id))
#Repeat process here for PC3 vs PC2
plot(pc.xray$z[,3], pc.xray$z[,2], xlab="PC3", ylab="PC2", pch=19, cex=1.75)
abline(h=0,v=0,lty=2,col="gray")
#identify(pc.xray$z[,3],pc.xray$z[,2], labels=basename.pdb(aligned_xanths$id))
#Repeat process here for PC1 vs PC3
plot(pc.xray$z[,1], pc.xray$z[,3], xlab="PC1", ylab="PC3", pch=19, cex=1.75)
abline(h=0,v=0,lty=2,col="gray")
#identify(pc.xray$z[,1],pc.xray$z[,3], labels=basename.pdb(aligned_xanths$id))
#STEP 13 - Write out CSV files for individual atom-wise loadings of PC1-PC3 (e.g.
"pc1.csv")as well as scores of each
structure on the PCs ("pca_zs.csv")
write.csv(x = pc.xray$au[,1],file = "pc1.csv")
write.csv(x = pc.xray$au[,2],file = "pc2.csv")
write.csv(x = pc.xray$au[,3],file = "pc3.csv")
write.csv(x = pc.xray$z,file = "pca_zs.csv")
#STEP 14 - Plot the contribution of each (Ca) residue position to an individual
component, x-axis are the residue
positions, y is the magnitude of the contribution
plot(pc.xray$au[,1], type="l", lwd=2, col="black", xlab="Residue Position", ylab="PC1
Contribution")
plot(pc.xray$au[,2], type="l", lwd=2, col="red", xlab="Residue Position", ylab="PC2
Contribution")
plot(pc.xray$au[,3], type="l", lwd=2, col="blue", xlab="Residue Position", ylab="PC3
Contribution")
#STEP 15 - Map the atom-wise loadings (that were just plotted) to the B-factor column of
a structure for 3d
analysis. This uses the apo-dep structure, but you can use any you want.
#Open the resulting file in PyMOL and run the "cartoon putty" command and then color
by B-factor
ca.inds <- atom.select(apo_dep, "calpha")
write.pdb(xyz=apo_dep$xyz[1,ca.inds$xyz], file="pc1.pdb", b = pc.xray$au[,1])
write.pdb(xyz=apo_dep$xyz[1,ca.inds$xyz], file="pc2.pdb", b = pc.xray$au[,2])
write.pdb(xyz=apo_dep$xyz[1,ca.inds$xyz], file="pc3.pdb", b = pc.xray$au[,3])

```

Appendix

1. Compromised catalysis and potential folding defects in in vitro studies of missense mutants associated with hereditary phosphoglucomutase 1 deficiency

Yingying Lee, **Kyle M. Stiers**, Bailee N. Kain, and Lesa J. Beamer

Recent studies have identified phosphoglucomutase 1 (PGM1) deficiency as an inherited metabolic disorder in humans. Affected patients show multiple disease phenotypes, including dilated cardiomyopathy, exercise intolerance, and hepatopathy, reflecting the central role of the enzyme in glucose metabolism. We present here the first in vitro biochemical characterization of 13 missense mutations involved in PGM1 deficiency. The biochemical phenotypes of the PGM1 mutants cluster into two groups: those with compromised catalysis and those with possible folding defects. Relative to the recombinant wild-type enzyme, certain missense mutants show greatly decreased expression of soluble protein and/or increased aggregation. In contrast, other missense variants are well behaved in solution, but show dramatic reductions in enzyme activity, with k_{cat}/K_m often $< 1.5\%$ of wild-type. Modest changes in protein conformation and flexibility are also apparent in some of the catalytically impaired variants. In the case of the G291R mutant, severely compromised activity is linked to the inability of a key active site serine to be phosphorylated, a prerequisite for catalysis. Our results complement previous in vivo studies, which suggest that both protein misfolding and catalytic impairment may play a role in PGM1 deficiency.

2. Synchrotron-based macromolecular crystallography module for an undergraduate biochemistry laboratory course

Kyle M Stiers, Christopher B Lee, Jay C Nix, John J Tanner, Lesa J Beamer

This paper describes the introduction of synchrotron-based macromolecular crystallography (MX) into an undergraduate laboratory class. An introductory 2 week experimental module on MX, consisting of four laboratory sessions and two classroom lectures, was incorporated into a senior-level biochemistry class focused on a survey of biochemical techniques, including the experimental characterization of proteins. Students purified recombinant protein samples, set up crystallization plates and flash-cooled crystals for shipping to a synchrotron. Students then collected X-ray diffraction data sets from their crystals via the remote interface of the Molecular Biology Consortium beamline (4.2.2) at the Advanced Light Source in Berkeley, CA, USA. Processed diffraction data sets were transferred back to the laboratory and used in conjunction with partial protein models provided to the students for refinement and model building. The laboratory component was supplemented by up to 2 h of lectures by faculty with expertise in MX. This module can be easily adapted for implementation into other similar undergraduate classes, assuming the availability of local crystallographic expertise and access to remote data collection at a synchrotron source.

3. Biology, mechanism, and structure of enzymes in the α -D-phosphohexomutase superfamily

Kyle M Stiers, Andrew G Muenks, Lesa J Beamer

Enzymes in the α -D-phosphohexomutases (PHM) superfamily catalyze the reversible conversion of phosphosugars, such as glucose 1-phosphate and glucose 6-phosphate. These reactions are fundamental to primary metabolism across the kingdoms of life, and are required for a myriad of cellular processes, ranging from exopolysaccharide production to protein glycosylation. The subject of extensive mechanistic characterization during the latter half of the twentieth century, these enzymes have recently benefitted from biophysical characterization, including X-ray crystallography, NMR, and hydrogen-deuterium exchange studies. This work has provided new insights into the unique catalytic mechanism of the superfamily, shed light on the molecular determinants of ligand recognition, and revealed the evolutionary conservation of conformational flexibility. Novel associations with inherited metabolic disease and the pathogenesis of bacterial infections have emerged, spurring renewed interest in the long-appreciated functional roles of these enzymes.

4. Structure and characterization of a class 3B proline utilization A: Ligand-induced dimerization and importance of the C-terminal domain for catalysis

David A Korasick, Thameesha T Gamage, Shelbi Christgen, **Kyle M Stiers**, Lesa J Beamer, Michael T Henzl, Donald F Becker, John J Tanner

The bifunctional flavoenzyme proline utilization A (PutA) catalyzes the two-step oxidation of proline to glutamate using separate proline dehydrogenase (PRODH) and L-glutamate-- semialdehyde dehydrogenase active sites. Because PutAs catalyze sequential reactions, they are good systems for studying how metabolic enzymes communicate via substrate channeling. Although mechanistically similar, PutAs vary widely in domain architecture, oligomeric state, and quaternary structure, and these variations represent different structural solutions to the problem of sequestering a reactive metabolite. Here, we studied PutA from *Corynebacterium freiburgense* (CfPutA), which belongs to the uncharacterized 3B class of PutAs. A 2.7 Å resolution crystal structure showed the canonical arrangement of PRODH, L-glutamate- -semialdehyde dehydrogenase, and C-terminal domains, including an extended interdomain tunnel associated with substrate channeling. The structure unexpectedly revealed a novel open conformation of the PRODH active site, which is interpreted to represent the non-activated conformation, an elusive form of PutA that exhibits suboptimal channeling. Nevertheless, CfPutA exhibited normal substrate-channeling activity, indicating that it isomerizes into the active state under assay conditions. Sedimentation-velocity experiments provided insight into the isomerization process, showing that CfPutA dimerizes in the presence of a proline analog and NAD. These results are consistent with the morpheein model of enzyme hysteresis, in which substrate binding induces conformational changes that promote assembly of a high-activity oligomer. Finally, we used domain deletion analysis to investigate the function of the C-terminal domain. Although this domain contains neither catalytic residues nor substrate sites, its removal impaired both catalytic activities, suggesting that it may be essential for active-site integrity.

5. Sequence-structure relationships, expression profiles, and disease-associated mutations in the paralogs of phosphoglucomutase 1

Andrew G Muenks, **Kyle M Stiers**, Lesa J Beamer

The key metabolic enzyme phosphoglucomutase 1 (PGM1) controls glucose homeostasis in most human cells. Four proteins related to PGM1, known as PGM2, PGM2L1, PGM3 and PGM5, and referred to herein as paralogs, are encoded in the human genome. Although all members of the same enzyme superfamily, these proteins have distinct substrate preferences and different functional roles. The recent association of PGM1 and PGM3 with inherited enzyme deficiencies prompts us to revisit sequence-structure and other relationships among the PGM1 paralogs, which are understudied despite their importance in human biology. Using currently available sequence, structure, and expression data, we investigated evolutionary relationships, tissue-specific expression profiles, and the amino acid preferences of key active site motifs. Phylogenetic analyses indicate both ancient and more recent divergence between the different enzyme sub-groups comprising the human paralogs. Tissue-specific protein and RNA expression profiles show widely varying patterns for each paralog, providing insight into function and disease pathology. Multiple sequence alignments confirm high conservation of key active site regions, but also reveal differences related to substrate specificity. In addition, we find that sequence variants of PGM2, PGM2L1, and PGM5 verified in the human population affect residues associated with disease-related mutants in PGM1 or PGM3. This suggests that inherited diseases related to dysfunction of these paralogs will likely occur in humans.

6. Phosphorylation-Dependent Effects on the Structural Flexibility of Phosphoglucosamine Mutase from *Bacillus anthracis*

Kyle M Stiers, Jia Xu, Yingying Lee, Zachary R Addison, Steven R Van Doren, Lesa J Beamer

Phosphoglucosamine mutase (PNGM) is an evolutionarily conserved bacterial enzyme in the peptidoglycan biosynthetic pathway, catalyzing the reversible conversion between glucosamine 1- and 6-phosphate. Previous structural studies of PNGM from the pathogen *Bacillus anthracis* revealed its dimeric assembly and highlighted the rotational mobility of its C-terminal domain. Recent studies of two other enzymes in the same superfamily have demonstrated the long-range effects on the conformational flexibility associated with phosphorylation of the conserved, active site phosphoserine involved in phosphoryl transfer. Building on this work, we use a combination of experimental and computational studies to show that the active, phosphorylated version of *B. anthracis* PNGM has decreased flexibility relative to its inactive, dephosphorylated state. Limited proteolysis reveals an enhanced and accelerated cleavage of the dephosphorylated enzyme. ¹⁵N transverse relaxation-optimized NMR spectra corroborate a conformational adjustment with broadening and shifts of peaks relative to the phospho-enzyme. Electrostatic calculations indicate that residues in the mobile, C-terminal domain are linked to the phosphoserine by lines of attraction that are absent in the dephosphorylated enzyme. Phosphorylation-dependent changes in protein flexibility appear linked with the conformational change and enzyme mechanism in PNGM, establishing this as a conserved theme in multiple subgroups of the diverse α -d-phosphohexomutase superfamily.

7. Assessment and Impacts of Phosphorylation on Protein Flexibility of the α -d-Phosphohexomutases

Kyle M Stiers, Lesa J Beamer

Enzymes in the α -d-phosphohexomutase (PHM) superfamily catalyze a multistep reaction, entailing two successive phosphoryl transfers. Key to this reaction is a conserved phosphoserine in the active site, which serves alternately as a phosphoryl donor and acceptor during the catalytic cycle. In addition to its role in the enzyme mechanism, the phosphorylation state of the catalytic phosphoserine has recently been found to have widespread effects on the structural flexibility of enzymes in this superfamily. These effects must be carefully accounted for when assessing other perturbations to these enzymes, such as mutations or ligand binding. In this chapter, we focus on methods for assessing and modulating the phosphorylation state of the catalytic serine, as well as straightforward ways to probe the impacts of this modification on protein structure/flexibility. This knowledge is essential for producing homogeneous and stable samples of these proteins for biophysical studies. The methods described herein should be widely applicable to enzymes across the PHM superfamily and may also be useful in characterizing the effects of posttranslational modifications on other proteins.

8. Synthesis, derivatization and structural analysis of phosphorylated mono-, di- and tri-fluorinated D-gluco-heptuloses by glucokinase: tunable phosphoglucomutase inhibition

Jian-She Zhu, **Kyle M. Stiers**, Sherany M. Winter, Anthony D. Garcia, Antoine F. Versini, Lesa J. Beamer, and David L. Jakeman

Glucokinase phosphorylated a series of C-1 fluorinated α -D-gluco-heptuloses. These phosphorylated products were discovered to be inhibitors of α -Phosphomannomutase/phosphoglucomutase (α PMM/PGM) and β -phosphoglucomutase (β PGM). Inhibition potency with both mutases inversely correlated to the degree of fluorination. Structural analysis with α PMM demonstrated inhibitor binding to the active site, with the phosphate in the phosphate binding site, and the anomeric hydroxyl directed to the catalytic site.

9. Inhibitory evaluation of α PMM/PGM from *Pseudomonas aeruginosa*: chemical synthesis, enzyme kinetic and protein crystallography study

Jian-She Zhu, **Kyle M. Stiers**, Ebrahim Soleimani, Brandon Groves, Lesa J. Beamer, and David L. Jakeman

α -Phosphomannomutase/phosphoglucomutase (α PMM/PGM) from *P. aeruginosa* is involved in bacterial cell wall assembly and is implicated in *P. aeruginosa* virulence, an important Gram-negative human pathogen, yet a few studies have addressed α PMM/PGM inhibition. Four structurally different α G1P derivatives including C1 fluoromethylated analogues (1-3), 1,2-cyclic phosph(on)ate analogues (4-6), isosteric methylene phosphono analogues (7-8) and 6-fluoro- α G1P (9), were synthesized and assessed as potential time-dependent or reversible α PMM/PGM inhibitors. The resulting kinetic data were consistent with the crystallographic structures of the homologous *Xanthomonas citri* PGM with inhibitors 3, 7, 8, and 9 binding to the enzyme active site (1.65 – 1.9 Å). Inhibitors with more bulky and electronegative substituents (1 and 2) at C1 and the cyclic derivatives were not observed crystallographically, consistent with weak inhibition data. These structural and kinetic insights will help facilitate the design of improved α PMM/PGM inhibitors

10. Galactose restores glycosylation through replenishing activated sugars in PGM1-CDG

Silvia Radenkovic, Matthew J Bird, Sunnie Y Wong, Catarina Felgueira, Kyle M Stiers, Leila Sabbagh, Nastassja Himmelreich, Gernot Poschet, Petra Windmolders, Ingrid Vander Elst, Jan Verheijen, Peter Witters, Ruqaiyah Altassan, Tomas Honzik, Tuba Eminoglu, Phillip M James, Andrew C Edmondson, Jozef Hertecant, Gert Matthijs, Tim Emmerzaal, Tamas Kozicz, Christian Thiel, Pieter Vermeersch⁴, Lesa Beamer, David Cassiman, Eva Morava, Bart Ghesquière

Congenital disorders of glycosylation (CDGs) are a diverse group of diseases affecting a critical form of post-translational modifications: glycosylation¹. Glycosylation is the attachment of sugar chains to proteins and lipids, and plays crucial roles in both the stability and function of the target molecules². Phosphoglucomutase 1 (PGM1: EC; 5.4.2.2) interconverts Glucose(Glc)-6-P and Glc-1-P. Disruption impairs both glycogen metabolism³, and glycosylation, the latter manifesting as a congenital disorder of glycosylation (PGM1-CDG, OMIM: 614921)⁴. Based on decreased galactosylation in truncated glycan chains, D-galactose (D-Gal for in vivo, later Gal for in vitro) was administered to PGM1-CDG patients and was shown biochemically to markedly reverse many of the disease symptoms⁵⁻⁸. The treatment mechanism(s) however have not been understood. Here, principally using tracer-based metabolomics^{9,10}, we report that D-Gal treatment of PGM1-CDG fibroblasts metabolically re-wires their sugar metabolism and replenishes depleted levels of the activated sugars UDP-Glc and UDP-Gal, so required for ER and Golgi linked glycosylation respectively. To this end, we further show that ¹³C₆ Gal is incorporated into mature de novo glycans. We also probed other potential mechanisms of galactose's action, such as through altered PGM1 activity, central carbon metabolism or mitochondrial function, but did not observe any effect of the treatment.

VITA

Kyle Stiers was born in Hermann, Missouri to Randy and Susan Stiers. Kyle went to the University of Missouri – Columbia for his undergraduate degree and obtained a B.S. in Biochemistry with minors in Information Technology, Computer Science, and Chemistry in 2014. He began graduate school immediately after and joined Dr. Lesa J. Beamer's laboratory in the spring of 2015 and completed his graduate work in the spring of 2019.