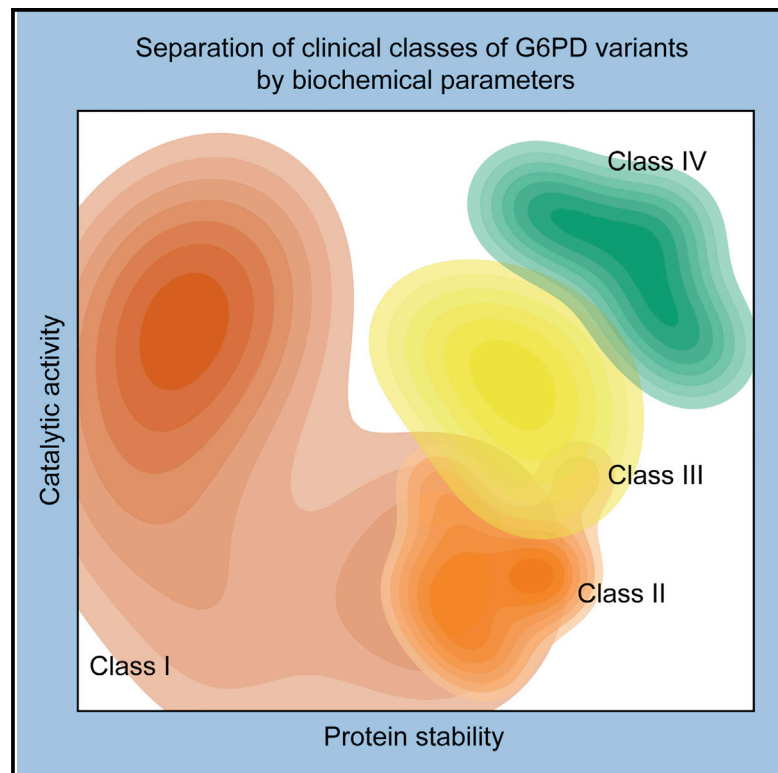


## Coupling between Protein Stability and Catalytic Activity Determines Pathogenicity of G6PD Variants

### Graphical Abstract



### Authors

Anna D. Cunningham, Alexandre Colavin, Kerwyn Casey Huang, Daria Mochly-Rosen

### Correspondence

mochly@stanford.edu

### In Brief

G6PD deficiency is one of the most common human enzymopathies, but the relationship between amino acid variant and clinical phenotype is poorly understood. Cunningham et al. find that clinical severity of a G6PD variant is determined by coupling between catalytic activity and protein stability.

### Highlights

- Structural distribution of G6PD variants identifies critical structural regions of G6PD
- Five previously uncharacterized G6PD variants from ExAC are likely nonpathogenic
- Coupling between catalytic activity and stability determines variant phenotype
- Severe G6PD variants affect functions not captured by standard biochemistry protocols

# Coupling between Protein Stability and Catalytic Activity Determines Pathogenicity of G6PD Variants

Anna D. Cunningham,<sup>1</sup> Alexandre Colavin,<sup>2</sup> Kerwyn Casey Huang,<sup>2,3,4</sup> and Daria Mochly-Rosen<sup>1,5,\*</sup>

<sup>1</sup>Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA 94305, USA

<sup>2</sup>Biophysics Program, Stanford University, Stanford, CA 94305, USA

<sup>3</sup>Department of Bioengineering, Stanford University, Stanford, CA 94305, USA

<sup>4</sup>Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305, USA

<sup>5</sup>Lead Contact

\*Correspondence: [mochly@stanford.edu](mailto:mochly@stanford.edu)

<http://dx.doi.org/10.1016/j.celrep.2017.02.048>

## SUMMARY

G6PD deficiency, an enzymopathy affecting 7% of the world population, is caused by over 160 identified amino acid variants in glucose-6-phosphate dehydrogenase (G6PD). The clinical presentation of G6PD deficiency is diverse, likely due to the broad distribution of variants across the protein and the potential for multidimensional biochemical effects. In this study, we use bioinformatic and biochemical analyses to interpret the relationship between G6PD variants and their clinical phenotype. Using structural information and statistical analyses of known G6PD variants, we predict the molecular phenotype of five uncharacterized variants from a reference population database. Through multidimensional analysis of biochemical data, we demonstrate that the clinical phenotypes of G6PD variants are largely determined by a trade-off between protein stability and catalytic activity. This work expands the current understanding of the biochemical underpinnings of G6PD variant pathogenicity and suggests a promising avenue for correcting G6PD deficiency by targeting essential structural features of G6PD.

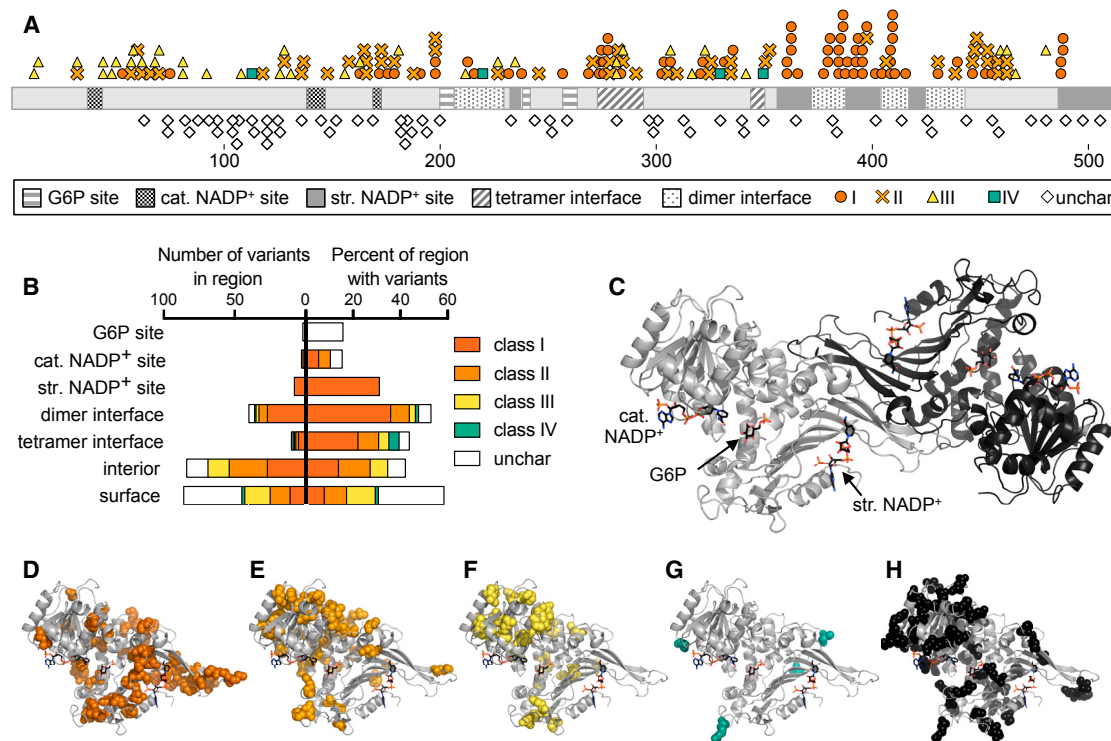
## INTRODUCTION

As the rate-limiting enzyme in the pentose phosphate pathway, glucose-6-phosphate dehydrogenase (G6PD) catalyzes the oxidation of glucose-6-phosphate (G6P) and concomitant reduction of nicotinamide adenine dinucleotide phosphate (NADP<sup>+</sup>) to NADPH (Cappellini and Fiorelli, 2008). NADPH then regenerates the essential antioxidant glutathione and is therefore important in maintaining redox homeostasis, especially in red blood cells, which lack mitochondria (Cappellini and Fiorelli, 2008). Certain single-amino-acid variants in G6PD lead to G6PD deficiency, one of the most common Mendelian diseases (Cappellini and Fiorelli, 2008). Roughly 7% of the world population is affected, with a geographic distribution that is strongly correlated with malaria prevalence, as G6PD deficiency protects

against malaria (Nkhoma et al., 2009; Vulliamy et al., 1992). G6PD deficiency is typically characterized by hemolytic episodes after acute oxidative insults; in rare severe cases, G6PD-deficient patients suffer from chronic non-spherocytic hemolytic anemia (CNSHA) (Cappellini and Fiorelli, 2008).

More than 160 unique missense variants in G6PD have been identified to cause G6PD deficiency, although their effects on G6PD biochemistry and disease phenotype vary widely (Luzzatto, 2006). In an attempt to address the biochemical, clinical, and genetic heterogeneity of G6PD deficiency, the World Health Organization (WHO) stratified patients with G6PD deficiency into four classes based on clinical presentation and G6PD activity in patient blood samples: class I (<10% activity and CNSHA), class II (<10% activity and hemolytic episodes), class III (10%–60% activity and hemolytic episodes), and class IV (60%–150% activity and no clinical manifestations) (Luzzatto, 2006; World Health Organization, 1967). However, these classifications are often determined via measurement of G6PD activity in the blood of single subjects, and they are possibly influenced by additional genetic, temporal, and environmental factors (Minucci et al., 2009; von Seidlein et al., 2013).

The diverse clinical presentation of G6PD deficiency motivates an equally diverse understanding of the molecular effects of G6PD variants. However, the molecular mechanisms of pathogenic G6PD variants remain largely unknown. Biochemical characterization of G6PD variants has revealed that pathogenic variants exhibit a range of complex multidimensional effects, including changes in kinetic activity, thermostability, and protein folding (Boonyuen et al., 2016; Gómez-Manzo et al., 2014, 2015, 2016; Huang et al., 2008; Wang and Engel, 2009; Wang et al., 2005, 2006). Crystal structures of human G6PD (Au et al., 2000; Kotaka et al., 2005) identified a dimeric or tetrameric enzyme with two bound NADP<sup>+</sup> molecules per subunit, one in the catalytic site and another in an allosteric site, named the structural NADP<sup>+</sup> for its importance in the thermostability and long-term stability of G6PD (Wang et al., 2008). Class I variants often fall near the structural NADP<sup>+</sup> site and exhibit decreased thermostability, suggesting that CNSHA associated with G6PD deficiency may result from G6PD instability and subsequent depletion of G6PD in red blood cells (Gómez-Manzo et al., 2014; Wang and Engel, 2009). However, class I variants are also found in many other structural regions of G6PD, and the relationship between



**Figure 1. G6PD Variant Classes Are Distributed Differently across the G6PD Structure**

(A) Linear representation of G6PD showing location of variants and structural regions. Structural regions shown are the G6P, catalytic (cat.) NADP<sup>+</sup>, and structural (str.) NADP<sup>+</sup> binding sites and dimer and tetramer interfaces. Variants shown are class I–IV and uncharacterized (unchar) variants from the ExAC database.

(B) Quantification of the number of variants in each structural region (left) and the percentage of amino acids in each region for which a variant has been identified (right).

(C) Crystal structure of dimeric G6PD, assembled from PDB: 2BH9 and 2BHL.

(D–H) Variant locations are shown in spheres on the monomeric structure of G6PD: class I (D), class II (E), class III (F), class IV (G), and uncharacterized variants from the ExAC database (H).

the structural or biochemical effects of a G6PD variant and its clinical phenotype remains poorly understood.

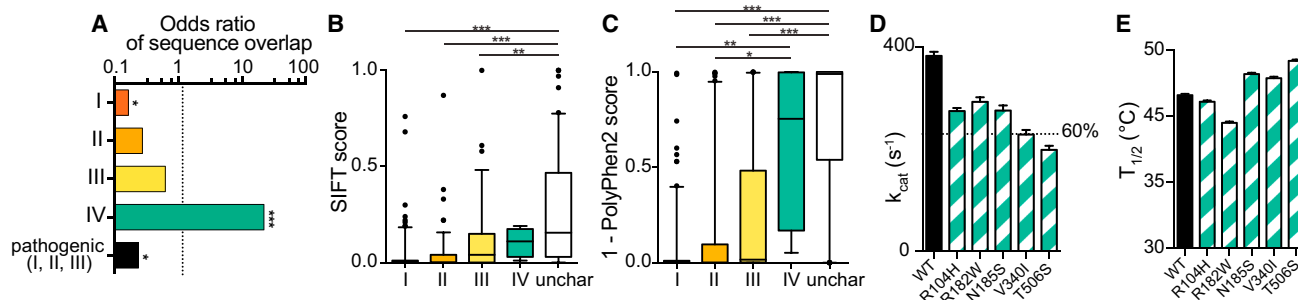
To elucidate the biochemical mechanisms underpinning the diverse phenotypes of G6PD variants, we combine statistical analyses with biochemical characterization of clinically relevant G6PD variants and variants identified from the Exome Aggregation Consortium (ExAC), a sequencing database of multiple large cohorts. We find highly significant relationships among the structural location of a G6PD variant, its effects on enzyme activity and stability, and its clinical outcome. This work provides insight into how competing evolutionary pressures and biological requirements have shaped the biochemical landscape of G6PD variants, predicts the phenotype of uncharacterized G6PD variants that appear in reference population databases, and suggests a promising avenue for the treatment of severe G6PD deficiency.

## RESULTS

### Structural Distribution of G6PD Variants

We defined structural regions (G6P and NADP<sup>+</sup> binding sites and oligomer interfaces) by calculating solvent-accessible surface

area using the three available crystal structures of human G6PD (PDB: 1QKI, 2BH9, 2BHL; [Experimental Procedures](#)) (Au et al., 2000; Kotaka et al., 2005). Reported variants (Benmansour et al., 2013; Chaves et al., 2016; García-Magallanes et al., 2014; Jang et al., 2015; Minucci et al., 2012; Warny et al., 2015) and structural regions were then mapped onto a linear representation of G6PD (Figures 1A and 1B) and onto the crystal structure (Figures 1C–1H). As previously speculated (Wang et al., 2008), the structural NADP<sup>+</sup> binding site is significantly enriched in class I variants ( $p < 0.05$  by Fisher's exact test). A lack of pathogenic variants at the substrate and cofactor binding sites ( $p < 0.05$ ) is consistent with the finding that complete loss of G6PD activity is embryonic lethal (Longo et al., 2002). The dimer interface is significantly enriched in pathogenic variants (class I, II, and III,  $p < 0.005$ ), especially class I variants ( $p < 0.001$ ). However, there is no significant enrichment of any variants at the tetramer interface. This pattern of variation suggests that while loss of dimerization is detrimental to G6PD activity, tetramerization may not be necessary for enzyme function. Taken together, the structural distribution of pathogenic variants highlights the importance of the dimer interface and structural NADP<sup>+</sup> binding site for G6PD function.



**Figure 2. Structural Distribution and Pathogenicity Prediction of Uncharacterized Variants from the ExAC Database Are Similar to Class IV Variants**

(A) Odds ratio of the sequence positional overlap between uncharacterized variants and known variants (class I, II, III, or IV) or all pathogenic (I, II, and III) variants. An odds ratio of 1 indicates expected overlap, a ratio below 1 indicates less overlap than expected, and a ratio above 1 indicates more overlap than expected. \* $p < 0.05$ ; \*\* $p < 0.005$ ; \*\*\* $p < 0.0005$ .

(B and C) Prediction of variant severity using two prediction algorithms: (B) SIFT, which assigns a score between 0 (damaging) and 1 (benign); and (C) PolyPhen2, which assigns a score between 1 (damaging) and 0 (benign). Unchar, uncharacterized variants from ExAC.

(D)  $k_{cat}$  measurements of wild-type (WT) G6PD and the five uncharacterized variants examined in this study. A line is shown at 60% activity, which delineates the separation between class III and class IV. Data represent mean  $\pm$  SD.

(E)  $T_{1/2}$  measurements of WT G6PD and the five uncharacterized variants.  $k_{cat}$  and  $T_{1/2}$  measurements both support class IV characterization. Data represent mean  $\pm$  SD.

See also Tables S1 and S2 and Figure S1.

### Structural Distribution and Pathogenicity Prediction of Uncharacterized Variants from the ExAC Database Are Similar to Class IV Variants

The recent availability of large sequencing databases provides an excellent opportunity for examining enzyme variation across multiple populations. The ExAC database, which catalogs exome sequences from over 60,000 unrelated individuals (Lek et al., 2016), contains 101 single missense variants in G6PD, of which 64 have not been previously reported (Table S1). The ExAC database generally excludes individuals with pediatric illnesses, so as expected, none of the 37 previously reported variants in ExAC were class I variants (Table S2). We therefore surmised that the 64 uncharacterized variants are also unlikely to be class I. We also examined allele frequencies of the previously known and uncharacterized variants, as variants of high allele frequency are often benign (Salgado et al., 2016). However, we found high allele frequencies of known pathogenic G6PD mutations (Figure S1; Tables S1 and S2), indicative of the selective advantage of G6PD variants against malaria, making allele frequency of the uncharacterized variants difficult to interpret.

Among these uncharacterized variants identified in the ExAC database, we observed significant enrichment on the surface of the protein ( $p < 0.001$ ) and depletion in the interior ( $p < 0.01$ ) and on the dimer interface ( $p < 0.05$ ). This structural distribution is most similar to the four known class IV mutations, of which two are on the surface and none are in the interior, as well as the class III mutations, of which half are on the surface. Indeed, mutations on the protein surface are less likely to be deleterious than mutations buried in the interior (Ng and Henikoff, 2006); therefore, the uncharacterized variants on the surface of G6PD are likely nonpathogenic.

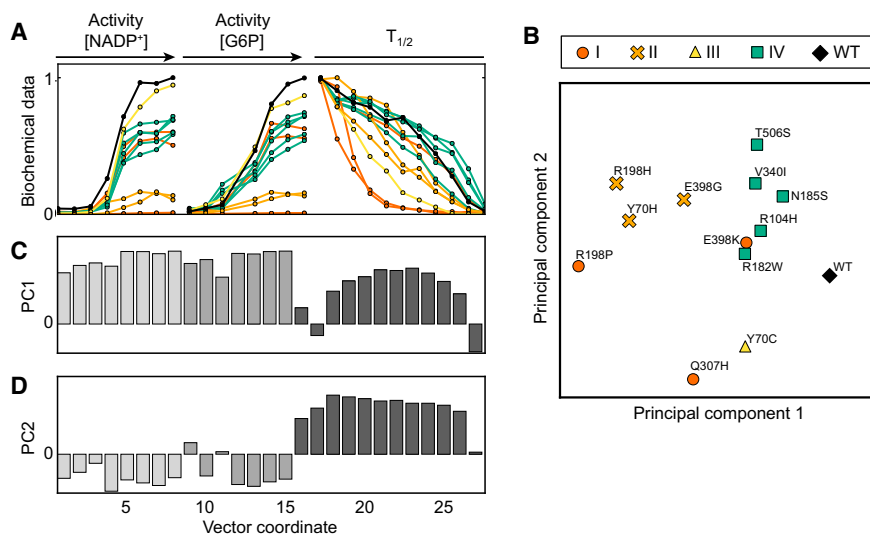
Because structural location (and therefore sequence position) is a major contributor to mutation severity (Adzhubei et al., 2010; Kumar et al., 2014; Ng and Henikoff, 2003), we examined how

many of these uncharacterized variants occur at the same amino acid position as known pathogenic variants. Overlap between uncharacterized and pathogenic variants would suggest that the overlapping uncharacterized variants are also likely to cause pathology. We found only one-quarter as much overlap between uncharacterized and pathogenic variants as expected ( $p < 0.001$  by Fisher's exact test, Figure 2A). Interestingly, uncharacterized variants overlapped significantly with three of the four known class IV variants ( $p < 0.001$ ).

To further evaluate the pathogenicity of these uncharacterized variants, we used two prediction algorithms: SIFT, which uses evolutionary conservation (Ng and Henikoff, 2003), and PolyPhen2, a machine learning method that combines chemical similarity, sequence information, and 3D structural information (Adzhubei et al., 2010). To test the reliability of these prediction algorithms, we included analysis of the 166 variants for which clinical classification has been previously reported (Figures 2B and 2C). Both algorithms showed a trend toward predicting class I variants to be more damaging and class IV variants to be more benign. Using both prediction algorithms, the uncharacterized variants were predicted to be more benign than class I, II, and III variants ( $p < 0.001$  by one-way ANOVA followed by Tukey's multiple comparison test; Figures 2B and 2C) and not significantly different from the class IV variants. However, all of the class groups contained outliers, showing that these prediction algorithms are limited or that the clinical manifestations of G6PD variants are likely structurally and biochemically complex and may be affected by additional genetic and environmental factors.

### Biochemical Characterization of G6PD Variants Reveals Diverse Effects of Different Amino Acid Substitutions

Although sequence position of a protein variant is a major determinant of variant severity, the chemical difference between the



**Figure 3. PCA of Biochemical Data Reveals Biochemical Separation between G6PD Variant Classes**

(A) Representation of data vectors used for PCA. Normalized median curves of kinetic and biochemical measurements for each G6PD variant are shown in black (wild-type), red (class I), orange (class II), yellow (class III), or green (predicted class IV).

(B) Biochemical characterization of 13 G6PD variants projected onto PCs 1 and 2.

(C and D) Values of principal components (PCs) 1 (C) and 2 (D), which represent correlation and anticorrelation, respectively, between activity and stability.

See also [Table S3](#).

original and substituted amino acid is also important (Adzhubei et al., 2010). Of G6PD variant pairs in which two different amino acid variants have been found at the same sequence position, 32 pairs yield the same class of G6PD deficiency, while 19 pairs yield different classes of G6PD deficiency. To further understand the biochemical consequence of amino acid substitutions, we expressed, purified, and biochemically characterized several clinically relevant G6PD variants, focusing on four pairs of variants in which the same amino acid position was changed to two different amino acids. In particular, we chose variant pairs that were assigned the same prediction by SIFT and PolyPhen2 yet yielded two different classes of G6PD deficiency: Y70H (II) and Y70C (III), R198P (I) and R198H (II), E398K (I) and E398G (II), and Q307H (I) and Q307P (III) ([Table S3](#)). Q307P yielded too little protein for biochemical characterization and was excluded from subsequent analyses. For each variant, we measured activity parameters ( $K_M$ ,  $k_{cat}$ ) and stability ( $T_{1/2}$ , the temperature at which half of G6PD activity is retained) using WHO standard protocols ([Table S3](#); [Supplemental Experimental Procedures](#)).

To further predict pathogenicity of uncharacterized variants from the ExAC database, we also biochemically characterized five of these variants, sampling a wide range of SIFT and PolyPhen2 prediction scores ([Table S3](#)). We found that  $T_{1/2}$  of all five variants was as high as or higher than that of wild-type (WT) G6PD, and  $k_{cat}$  of four of these variants was within the class IV activity range of 60%–150% ([Figures 2D and 2E](#)); the fifth had  $k_{cat}$  slightly below this range. Thus, in subsequent analysis we treat these five uncharacterized variants as class IV.

### Principal-Component Analysis of Biochemical Data Reveals a Trade-Off between Activity and Stability

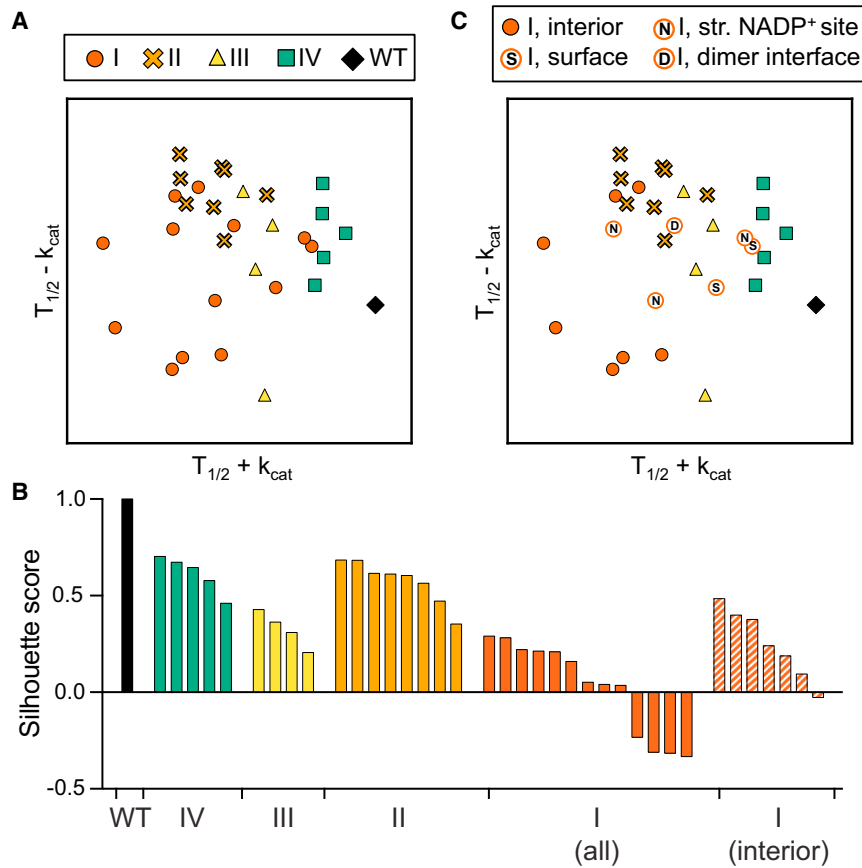
Initial inspection of the biochemical data from the selected G6PD variants did not reveal obvious trends, suggesting that there are couplings between biochemical parameters that might be revealed through dimensional reduction. We subjected the data to principal-component analysis (PCA), using our kinetic and sta-

bility measurements. To avoid bias from curve fitting, rather than using fitted kinetic and stability parameters, we used a vector containing the raw data from activity and stability measurements (27 data points, [Figure 3A](#); [Experimental Procedures](#)). Principal components (PCs) 1 and 2 accounted for 50% and 24% of the variance in the data, respectively. Projecting the 13 variants onto these two components, we found that each pair of pathogenic variants at the same amino acid position were biochemically distinct from each other, highlighting the importance of amino acid chemical properties in determining the severity of a protein variant.

Interestingly, the five uncharacterized variants that we predicted to be class IV clustered together near WT G6PD ([Figure 3B](#)), further validating that these variants are biochemically more similar to WT G6PD than are pathogenic variants. The three class II variants also visually clustered together and separated from WT/class IV ([Figure 3B](#)). To probe the basis of this separation, we examined the biochemical signatures encoded by PCs 1 and 2 ([Figures 3C and 3D](#)). PC 1 consisted of mainly positive values, reflecting correlation between activity and stability (i.e., a variant having high activity and high stability, or low activity and low stability). Interestingly, PC 2 contained negative values for the activity measurements and positive values for the stability measurements, reflecting anticorrelation between activity and stability (i.e., a variant having high activity and low stability or low activity and high stability). This analysis suggests that the clinical phenotype of a G6PD variant is determined by both its overall performance and by a trade-off between catalytic activity and protein stability.

### Generalized Principal-Component Axes Separate G6PD Variants into Clusters by Class

We then generalized the PCs to other previously characterized G6PD variants by plotting the variants from this study and 20 variants from previous work (Boonyuen et al., 2016; Gómez-Manzo et al., 2014, 2015, 2016; Huang et al., 2008; Wang et al., 2005, 2008; Wang and Engel, 2009) ([Table S4](#)) onto axes roughly corresponding to the PCs. To reflect correlation and anticorrelation between stability and activity, we compared normalized values of  $T_{1/2} + k_{cat}$  and  $T_{1/2} - k_{cat}$ , respectively. We found that class



**Figure 4. PCA-like Visualization Suggests Additional Functionality of Some Class I Variants**

(A) Analysis of 13 G6PD variants from this study and 20 additional variants from previous studies, represented by normalized  $T_{1/2} + k_{cat}$  and  $T_{1/2} - k_{cat}$ .

(B) Silhouette scores of each variant cluster from (A) (WT, IV, II, III, and I) and (C) (class I, excluding variants that are not in the protein interior). A silhouette score ranges from  $-1$  to  $1$ , with a higher score indicating that a point is matched well to its cluster and matched poorly to other clusters.

(C) PCA-like plot with class I variants labeled corresponding to their structural location (S, surface; N, structural NADP<sup>+</sup>; D, dimer interface). Class I variants in solid red circles are located in the protein interior.

See also [Table S4](#) and [Figures S2](#) and [S3](#).

structural NADP<sup>+</sup> site, or dimer interface are excluded from the clustering analysis, the remaining class I variants form a tighter cluster as quantified by silhouette score ([Figures 4B](#) and [4C](#)).

## DISCUSSION

G6PD is ubiquitously expressed and essential for maintaining redox homeostasis in all tissues and cell types. As a result, there are conflicting evolutionary

pressures that shape its mutational landscape. Although complete loss of enzyme function is lethal, mild loss of function is advantageous as it protects against malaria. This loss of function can occur through different biochemical mechanisms leading to different clinical outcomes. Based on our PCA of biochemical characterization, unsurprisingly, the overall fitness of the enzyme (good stability and activity) is the primary determinant of the clinical outcome of a G6PD variant (PC1; [Figure 3C](#)). Interestingly, as shown by PC2 ([Figure 3D](#)), the clinical outcome is also largely determined by a trade-off between stability and activity. The importance of this trade-off is consistent with the necessity for G6PD to retain NADPH-producing activity in all cell types while also remaining stable in red blood cells, which contain no translational machinery, for the lifetime of a red blood cell (up to 110 days).

Taken together, our biochemical characterization and meta-analysis of G6PD variants has clarified the biochemical underpinning that determines the severity of G6PD deficiency. We found that activity or stability alone does not determine or predict the phenotype of a G6PD variant, but a combination of both yields significant separation of variants by class. This finding is a crucial advance that informs previous work in which a quantitative model of G6PD kinetics, which did not include protein stability as a parameter, was unable to segregate class I variants into a biochemical space consistent with the CNSHA phenotype ([Coelho et al., 2010](#)).

II, III, and IV variants segregated visually by class ([Figures 4A](#) and [S2](#)) and were quantitatively clustered by silhouette scoring ([Figure 4B](#)). This clustering was not recapitulated by taking into account  $k_{cat}$  ([Figure S3A](#)) or  $T_{1/2}$  ([Figure S3B](#)) alone, highlighting the importance of trade-offs between them in determining clinical severity.

In particular, previous work has suggested that class I variants cause CNSHA due to decreased enzyme stability and hence lower enzyme levels in red blood cells. In this study, we observed several class I variants that confirm this finding, but we also show that not all class I variants display reduced stability or enzyme activity in vitro. In the generalized PCA-like plot ([Figure 4A](#)), the class I variants did not cluster significantly; we speculate several possible reasons for this. Because class I mutations are rare, most documented class I mutations are single case study reports. Thus, it is possible that some class I mutations may only mildly disrupt G6PD activity but could nevertheless contribute to CNSHA when combined with a patient's unique genetic background. Another possibility is that some class I variants disrupt the function of G6PD in a manner not captured by the assay conditions used in vitro. For example, the high concentrations of Mg<sup>2+</sup>, NADP<sup>+</sup>, or G6PD enzyme in vitro may mask the effects of class I variants on dimerization or structural NADP<sup>+</sup> binding under physiological conditions. Class I variants on the surface may also disrupt an essential protein-protein interaction in vivo. Indeed, when class I variants located on the protein surface,

We found that the structural distribution of the uncharacterized and known variants is similar to that of class IV variants (Figure 2A) and that the predicted pathogenicity of these variants is not significantly different from the class IV variants (Figures 2B and 2C). Furthermore, based on the biochemical signatures of five uncharacterized variants from the ExAC database, we predict that these five variants are likely nonpathogenic (class IV), even though one variant (R182W) was predicted to be pathogenic by SIFT (Table S3). However, other uncharacterized variants that were predicted to be severe by SIFT or PolyPhen2, or that lie in the dimer interface, may be outliers in this group and should be further characterized biochemically to predict their phenotype.

Consistent with previous work, we found that G6PD variants with the lowest stability often result in a worse clinical outcome (Figure 4A). Therefore, we propose that efforts toward identifying a therapy to treat G6PD deficiency should focus on increasing enzyme stability. The structural NADP<sup>+</sup> site is known to be important for enzyme stability, and we found significant enrichment of class I variants in this site, suggesting that therapies that improve structural NADP<sup>+</sup> binding may rescue severe G6PD deficiency. Additionally, we found that the dimer interface is significantly enriched in pathogenic variants, especially class I variants. This suggests enzyme dimerization as a promising target for rescuing G6PD deficiency of any type. Nonetheless, our data indicate that the efficacy of treatment for patients with severe symptoms may be variant dependent.

Beyond acute and chronic anemia, there are many other pathologies associated with reduced G6PD activity. G6PD deficiency increases the risk of kernicterus and death from neonatal jaundice (Cunningham et al., 2016) and has also been associated with bipolar and schizoaffective disorders (Raj et al., 2014), erectile dysfunction (Morrison et al., 2014), and vitiligo (Namazi, 2015). Because G6PD plays an essential role in maintaining healthspan by protection against oxidative damage (Nóbrega-Pereira et al., 2016), the effects of G6PD deficiency on human health have likely been underestimated, and it is thus expected that additional consequences of G6PD deficiency will be identified in the future (Spencer and Stanton, 2017; Stanton, 2012). Our biochemical and informatics-based study suggests a promising avenue for treatment of G6PD deficiency and its sequelae by targeting enzyme stability, structural NADP<sup>+</sup> binding, or dimerization.

## EXPERIMENTAL PROCEDURES

### Definition of Structural Regions

For each available crystal structure of human G6PD (PDB: 1QKI, 2BH9, and 2BHL), solvent-accessible surface area (SASA) was calculated using AREAIMOL (Lee and Richards, 1971; Saff and Kuijlaars, 1997) in the CCP4 suite (Winn et al., 2011). Structural regions were defined by changes in SASA; for example, an amino acid was included in the dimer interface if the SASA was reduced in the dimeric structure compared to the monomeric structure. Amino acids that were not at an oligomeric interface or G6P- or NADP<sup>+</sup>-binding site were designated as “surface” (SASA > 25%) or “interior” (SASA < 25%) (Levy, 2010). For ease of visualization in Figure 1A, structural regions were approximated by blocks spanning the densest clusters of amino acids in each region. Structure images were generated using PyMol version 1.7.6.6.

### Analysis of ExAC Database Mutations

High-quality missense variants (genotype quality  $\geq 20$  and depth  $\geq 10$ ) in the G6PD transcript ENST00000393564, as of October 2016, were collected

from the ExAC browser (Lek et al., 2016) and compared with G6PD variants reported in the literature (Benmansour et al., 2013; Chaves et al., 2016; García-Magallanes et al., 2014; Jang et al., 2015; Minucci et al., 2012; Warny et al., 2015). Variants in ExAC that were not previously reported were designated “uncharacterized.” Uncharacterized variants and variants from the literature were submitted by batch query to SIFT (Ng and Henikoff, 2003) (Ensembl: ENSP00000377194) and PolyPhen2 (Adzhubei et al., 2010) (Uniprot: P11413) using the default settings.

### PCA

For each G6PD variant, a vector was assembled using the median value for each data point measured: activity measurements varying [NADP<sup>+</sup>] (eight data points), activity measurements varying [G6P] (seven data points), and stability measurements (12 data points) (see “Enzyme activity and stability measurements” in Supplemental Experimental Procedures for further description of biochemical measurements). The data were then normalized by subtracting the mean and dividing by the SD at each position in the vector. PCA was performed using the Python scikit-learn module (sklearn.decomposition.PCA, version 0.17.1) (Pedregosa et al., 2011).

### Generation of PCA-like Plot

$T_{1/2}$  and  $k_{cat}$  data from G6PD variants previously purified and biochemically characterized following WHO standards (Boonyuen et al., 2016; Gómez-Manzo et al., 2014, 2015, 2016; Huang et al., 2008; Wang et al., 2005, 2008; Wang and Engel, 2009) were combined with data from this study and then normalized by subtracting the mean and dividing by the SD. The PCA-like plot ( $T_{1/2} - k_{cat}$  versus  $T_{1/2} + k_{cat}$ ) was then generated using these normalized values.

### Silhouette Scoring

Silhouette scoring was calculated using a custom Python script. Each cluster was defined as variants in the same clinical class. For each cluster, the nearest cluster was identified by comparing the mean center of each cluster. Then, for each point in the cluster, the silhouette score ( $s$ ) was defined as:

$$s = \frac{(b - a)}{\max(a, b)}$$

where  $a$  is the mean distance from the point to all other points in the cluster and  $b$  is the mean distance from the point to all other points in the nearest cluster. A silhouette score ranges from  $-1$  to  $1$ , with a higher score indicating that a point is matched well to its cluster and matched poorly to other clusters.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and four tables and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2017.02.048>.

## AUTHOR CONTRIBUTIONS

Conceptualization, A.D.C. and A.C.; Methodology, A.D.C., A.C., K.C.H. and D.M.-R.; Investigation, A.D.C.; Resources, D.M.-R.; Writing – Original Draft, A.D.C.; Writing – Review & Editing, A.C., K.C.H., and D.M.-R.; Visualization, A.D.C.; Supervision, K.C.H. and D.M.-R.; Funding Acquisition, A.D.C., A.C., K.C.H., and D.M.-R.

## ACKNOWLEDGMENTS

This work was supported in part by NIH grant R01 HD08442 (to D.M.-R.), NSF CAREER Award MCB-1149328 (to K.C.H.), a seed grant from the Stanford Center for Systems Biology under grant P50-GM107615, Stanford Graduate Fellowships (to A.D.C. and A.C.), a Gerald J. Lieberman Fellowship (to A.C.), and an NSF Graduate Research Fellowship DGE-114747 (to A.D.C.). We thank C. Liu, S. Hwang, and A. Raub for their critical reading of the manuscript.

Received: November 11, 2016

Revised: January 8, 2017

Accepted: February 15, 2017

Published: March 14, 2017

## REFERENCES

- Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249.
- Au, S.W., Gover, S., Lam, V.M., and Adams, M.J. (2000). Human glucose-6-phosphate dehydrogenase: the crystal structure reveals a structural NADP(+) molecule and provides insights into enzyme deficiency. *Structure* **8**, 293–303.
- Benmansour, I., Moradkhani, K., Mourni, I., Wajcman, H., Hafsia, R., Ghanem, A., Abbès, S., and Préhu, C. (2013). Two new class III G6PD variants [G6PD Tunis (c.920A>C: p.307Gln>Pro) and G6PD Nefza (c.968T>C: p.323 Leu>Pro)] and overview of the spectrum of mutations in Tunisia. *Blood Cells Mol. Dis.* **50**, 110–114.
- Boonyuen, U., Chamchoy, K., Swangsri, T., Saralamba, N., Day, N.P.J., and Imwong, M. (2016). Detailed functional analysis of two clinical glucose-6-phosphate dehydrogenase (G6PD) variants, G6PDViangchan and G6PDViangchan+Mahidol: decreased stability and catalytic efficiency contribute to the clinical phenotype. *Mol. Genet. Metab.* **118**, 84–91.
- Cappellini, M.D., and Fiorelli, G. (2008). Glucose-6-phosphate dehydrogenase deficiency. *Lancet* **371**, 64–74.
- Chaves, A., Eberle, S.E., Defelipe, L., Pepe, C., Milanesio, B., Aguirre, F., Fernandez, D., Turjanski, A., and Feliú-Torres, A. (2016). Two novel DNA variants associated with glucose-6-phosphate dehydrogenase deficiency found in Argentine pediatric patients. *Clin. Biochem.* **49**, 808–810.
- Coelho, P.M., Salvador, A., and Savageau, M.A. (2010). Relating mutant genotype to phenotype via quantitative behavior of the NADPH redox cycle in human erythrocytes. *PLoS ONE* **5**, e13031.
- Cunningham, A.D., Hwang, S., and Mochly-Rosen, D. (2016). Glucose-6-phosphate dehydrogenase deficiency and the need for a novel treatment to prevent kernicterus. *Clin. Perinatol.* **43**, 341–354.
- García-Magallanes, N., Luque-Ortega, F., Aguilar-Medina, E.M., Ramos-Payán, R., Galaviz-Hernández, C., Romero-Quintana, J.G., Del Pozo-Yauner, L., Rangel-Villalobos, H., and Arámbula-Meraz, E. (2014). Glucose-6-phosphate dehydrogenase deficiency in northern Mexico and description of a novel mutation. *J. Genet.* **93**, 325–330.
- Gómez-Manzo, S., Terrón-Hernández, J., De la Mora-De la Mora, I., González-Valdez, A., Marcial-Quino, J., García-Torres, I., Vanoye-Carlo, A., López-Velázquez, G., Hernández-Alcántara, G., Oria-Hernández, J., et al. (2014). The stability of G6PD is affected by mutations with different clinical phenotypes. *Int. J. Mol. Sci.* **15**, 21179–21201.
- Gómez-Manzo, S., Marcial-Quino, J., Vanoye-Carlo, A., Enríquez-Flores, S., De la Mora-De la Mora, I., González-Valdez, A., García-Torres, I., Martínez-Rosas, V., Sierra-Palacios, E., Lazcano-Pérez, F., et al. (2015). Mutations of glucose-6-phosphate dehydrogenase Durham, Santa-Maria and A+ variants are associated with loss functional and structural stability of the protein. *Int. J. Mol. Sci.* **16**, 28657–28668.
- Gómez-Manzo, S., Marcial-Quino, J., Vanoye-Carlo, A., Serrano-Posada, H., González-Valdez, A., Martínez-Rosas, V., Hernández-Ochoa, B., Sierra-Palacios, E., Castillo-Rodríguez, R.A., Cuevas-Cruz, M., et al. (2016). Functional and biochemical characterization of three recombinant human glucose-6-phosphate dehydrogenase mutants: Zacatecas, Vanua-Lava and Viangchan. *Int. J. Mol. Sci.* **17**, 787.
- Huang, Y., Choi, M.Y., Au, S.W.N., Au, D.M.Y., Lam, V.M.S., and Engel, P.C. (2008). Purification and detailed study of two clinically different human glucose 6-phosphate dehydrogenase variants, G6PD(Plymouth) and G6PD(Mahidol): Evidence for defective protein folding as the basis of disease. *Mol. Genet. Metab.* **93**, 44–53.
- Jang, M.-A., Kim, J.-Y., Lee, K.-O., Kim, S.-H., Koo, H.H., and Kim, H.-J. (2015). A novel de novo mutation in the G6PD gene in a Korean boy with glucose-6-phosphate dehydrogenase deficiency: case report. *Ann. Clin. Lab. Sci.* **45**, 446–448.
- Kotaka, M., Gover, S., Vandeputte-Rutten, L., Au, S.W.N., Lam, V.M.S., and Adams, M.J. (2005). Structural studies of glucose-6-phosphate and NADP+ binding to human glucose-6-phosphate dehydrogenase. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 495–504.
- Kumar, A., Rajendran, V., Sethumadhavan, R., Shukla, P., Tiwari, S., and Purohit, R. (2014). Computational SNP analysis: current approaches and future prospects. *Cell Biochem. Biophys.* **68**, 233–239.
- Lee, B., and Richards, F.M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400.
- Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291.
- Levy, E.D. (2010). A simple definition of structural regions in proteins and its use in analyzing interface evolution. *J. Mol. Biol.* **403**, 660–670.
- Longo, L., Vanegas, O.C., Patel, M., Rosti, V., Li, H., Waka, J., Merghoub, T., Pandolfi, P.P., Notaro, R., Manova, K., and Luzzatto, L. (2002). Maternally transmitted severe glucose 6-phosphate dehydrogenase deficiency is an embryonic lethal. *EMBO J.* **21**, 4229–4239.
- Luzzatto, L. (2006). Glucose 6-phosphate dehydrogenase deficiency: from genotype to phenotype. *Haematologica* **91**, 1303–1306.
- Minucci, A., Giardina, B., Zuppi, C., and Capoluongo, E. (2009). Glucose-6-phosphate dehydrogenase laboratory assay: How, when, and why? *IUBMB Life* **61**, 27–34.
- Minucci, A., Moradkhani, K., Hwang, M.J., Zuppi, C., Giardina, B., and Capoluongo, E. (2012). Glucose-6-phosphate dehydrogenase (G6PD) mutations database: review of the “old” and update of the new mutations. *Blood Cells Mol. Dis.* **48**, 154–165.
- Morrison, B.F., Thompson, E.B., Shah, S.D., and Wharfe, G.H. (2014). Ischaemic priapism and glucose-6-phosphate dehydrogenase deficiency: a mechanism of increased oxidative stress? *West Indian Med. J.* **63**, 658–660.
- Namazi, M.R. (2015). What is the important practical implication of detecting decreased G6PD levels in vitiligo? *Adv. Biomed. Res.* **4**, 89.
- Ng, P.C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814.
- Ng, P.C., and Henikoff, S. (2006). Predicting the effects of amino acid substitutions on protein function. *Annu. Rev. Genomics Hum. Genet.* **7**, 61–80.
- Nkhoma, E.T., Poole, C., Vannappagari, V., Hall, S.A., and Beutler, E. (2009). The global prevalence of glucose-6-phosphate dehydrogenase deficiency: a systematic review and meta-analysis. *Blood Cells Mol. Dis.* **42**, 267–278.
- Nóbrega-Pereira, S., Fernandez-Marcos, P.J., Briocche, T., Gomez-Cabrera, M.C., Salvador-Pascual, A., Flores, J.M., Viña, J., and Serrano, M. (2016). G6PD protects from oxidative damage and improves healthspan in mice. *Nat. Commun.* **7**, 10894.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830.
- Raj, V., Chism, K., Minckler, M.R., and Denysenko, L. (2014). Catatonia and glucose-6-phosphate dehydrogenase deficiency: a report of two cases and a review. *Psychosomatics* **55**, 92–97.
- Saff, E.B., and Kuijlaars, A.B.J. (1997). Distributing many points on a sphere. *Math. Intell.* **19**, 5–11.
- Salgado, D., Bellgard, M.I., Desvignes, J.P., and Bérout, C. (2016). How to identify pathogenic mutations among all those variations: variant annotation and filtration in the genome sequencing era. *Hum. Mutat.* **37**, 1272–1282.
- Spencer, N.Y., and Stanton, R.C. (2017). Glucose 6-phosphate dehydrogenase and the kidney. *Curr. Opin. Nephrol. Hypertens.* **26**, 43–49.



- Stanton, R.C. (2012). Glucose-6-phosphate dehydrogenase, NADPH, and cell survival. *IUBMB Life* 64, 362–369.
- von Seidlein, L., Auburn, S., Espino, F., Shanks, D., Cheng, Q., McCarthy, J., Baird, K., Moyes, C., Howes, R., Ménard, D., et al. (2013). Review of key knowledge gaps in glucose-6-phosphate dehydrogenase deficiency detection with regard to the safe clinical deployment of 8-aminoquinoline treatment regimens: a workshop report. *Malar. J.* 12, 112.
- Vulliamy, T., Mason, P., and Luzzatto, L. (1992). The molecular basis of glucose-6-phosphate dehydrogenase deficiency. *Trends Genet.* 8, 138–143.
- Wang, X.-T., and Engel, P.C. (2009). Clinical mutants of human glucose 6-phosphate dehydrogenase: impairment of NADP(+) binding affects both folding and stability. *Biochim. Biophys. Acta* 1792, 804–809.
- Wang, X.-T., Lam, V.M.S., and Engel, P.C. (2005). Marked decrease in specific activity contributes to disease phenotype in two human glucose 6-phosphate dehydrogenase mutants, G6PD(Union) and G6PD(Andalus). *Hum. Mutat.* 26, 284.
- Wang, X.-T., Lam, V.M.S., and Engel, P.C. (2006). Functional properties of two mutants of human glucose 6-phosphate dehydrogenase, R393G and R393H, corresponding to the clinical variants G6PD Wisconsin and Nashville. *Biochim. Biophys. Acta* 1762, 767–774.
- Wang, X.-T., Chan, T.F., Lam, V.M.S., and Engel, P.C. (2008). What is the role of the second “structural” NADP+-binding site in human glucose 6-phosphate dehydrogenase? *Protein Sci.* 17, 1403–1411.
- Warny, M., Lausen, B., Birgens, H., Knabe, N., and Petersen, J. (2015). Severe G6PD deficiency due to a new missense mutation in an infant of Northern European descent. *J. Pediatr. Hematol. Oncol.* 37, e497–e499.
- Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* 67, 235–242.
- World Health Organization (1967). Standardization of procedures for the study of glucose-6-phosphate dehydrogenase: report of a WHO Scientific Group. *World Health Organ. Tech. Rep. Ser* 366, 1–53.

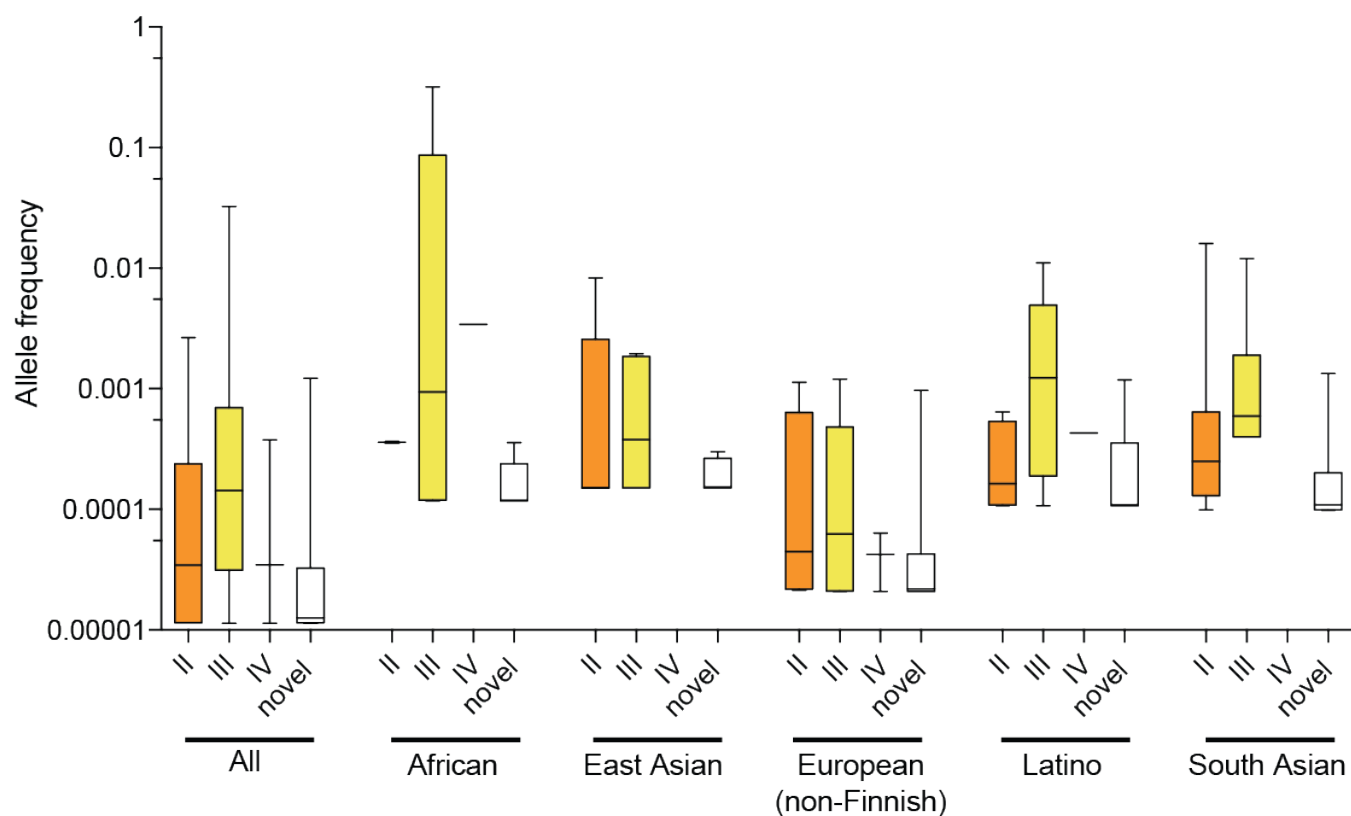
**Cell Reports, Volume 18**

**Supplemental Information**

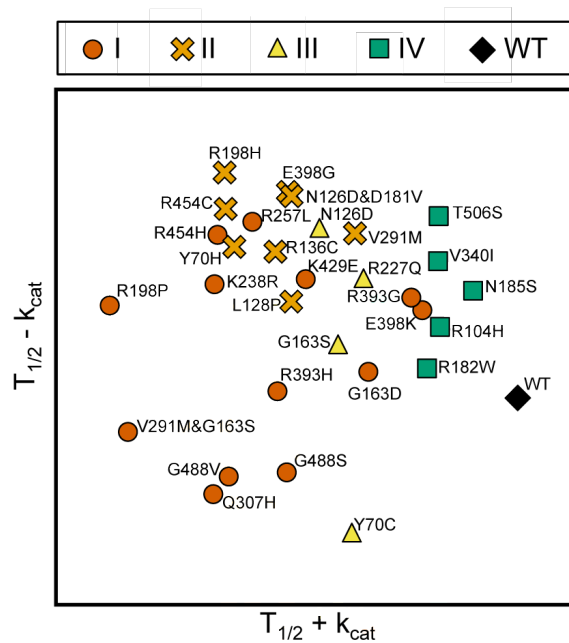
**Coupling between Protein Stability and Catalytic  
Activity Determines Pathogenicity of G6PD Variants**

**Anna D. Cunningham, Alexandre Colavin, Kerwyn Casey Huang, and Daria Mochly-Rosen**

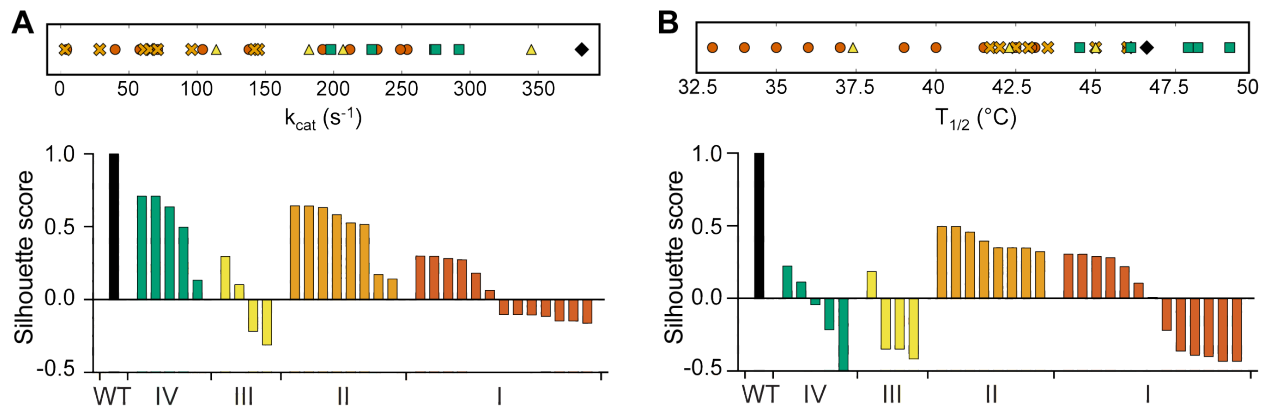
## Supplemental Figures and Tables



**Figure S1: Allele frequencies of previously known and uncharacterized G6PD variants in the ExAC database. Related to Figure 2.** Allele frequencies of G6PD variants in each population (Tables S1, S2) are plotted, separated by class. The Finnish population is omitted due to the low number of variants observed in this population.



**Figure S2: Labeled PCA-like plot. Related to Figure 4.** PCA-like plot of variants from this study and other studies, labeled corresponding to each variant (data are the same as in Fig. 4A).



**Figure S3: Separation of G6PD variants along individual biochemical parameters. Related to Figure 4.**

- A)  $k_{cat}$  values of variants from this study and other studies (top) and silhouette scores resulting from clustering (bottom).
- B)  $T_{1/2}$  values of variants from this study and other studies (top) and silhouette scores resulting from clustering (bottom).

**Table S1: Uncharacterized G6PD variants in the ExAC database. Related to Figure 2.** Variants characterized in this study are highlighted in yellow.

*See separate Excel file.*

**Table S2: Previously identified G6PD variants in the ExAC database. Related to Figure 2.**

*See separate Excel file.*

**Table S3: Biochemical parameters of G6PD variants in this study. *h* is the fitted Hill coefficient.** Related to Figure 3A.

Variant	Class	SIFT	PolyPhen2	$K_M$ NADP <sup>+</sup> ( $\mu$ M)	<i>h</i> NADP <sup>+</sup>	$K_M$ G6P ( $\mu$ M)	<i>h</i> G6P	$k_{cat}$ (s <sup>-1</sup> )	$T_{1/2}$ (°C)
WT	--	--	--	16.3 ± 0.6	2.3	106 ± 4	1.6	382 ± 8	46.6 ± 0.1
Y70H	II	0	0.999	34 ± 4	2.2	143 ± 18	1	61 ± 5	41.7 ± 0.2
Y70C	III	0	1	21 ± 2	2.5	92 ± 5	3	345 ± 12	37.4 ± 0.3
R198P	I	0	1	38 ± 2	1.5	26 ± 4	1	4.4 ± 0.2	36.0 ± 0.2
R198H	II	0	1	48 ± 3	1.5	81 ± 19	0.7	3.2 ± 0.2	43.5 ± 0.2
Q307H	I	0	1	14.7 ± 0.6	2.3	83 ± 3	2.3	212 ± 5	34.0 ± 0.2
Q307P	III	0	1	--	--	--	--	--	--
E398K	I	0	1	16.3 ± 0.3	2.7	78 ± 3	2.8	249 ± 7	46.0 ± 0.2
E398G	II	0	1	11 ± 1	1.7	47 ± 5	1.6	65 ± 4	45.0 ± 0.1
R104H	--	0.62	0.106	16.1 ± 0.5	3.9	98 ± 4	1.5	274 ± 6	46.1 ± 0.1
R182W	--	0	0.437	20 ± 1	2	145 ± 10	1	292 ± 9	44.5 ± 0.1
N185S	--	0.91	0	17 ± 1	3.3	94 ± 7	1	275 ± 10	48.2 ± 0.1
V340I	--	0.66	0.047	20 ± 1	2.7	119 ± 12	1	228 ± 9	47.9 ± 0.1
T506S	--	0.74	0.008	18.9 ± 0.9	2.8	107 ± 12	1	198 ± 8	49.2 ± 0.1



**Table S4: Biochemical parameters of G6PD variants from other studies that are used in this study.** Related to Figure 4.

<b>Variant</b>	<b>Class</b>	<b><math>K_M</math> NADP<sup>+</sup> (<math>\mu</math>M)</b>	<b><math>K_M</math> G6P (<math>\mu</math>M)</b>	<b><math>k_{cat}</math> (<math>s^{-1}</math>)</b>	<b><math>T_{1/2}</math> (<math>^{\circ}</math>C)</b>	<b>Reference</b>
<b>K429E</b>	I	6.4	39.9	138	43.1	(Gómez-Manzo et al., 2014)
<b>R136C</b>	II	3.6	21.5	96	42.9	(Gómez-Manzo et al., 2014)
<b>R227Q</b>	III	9.1	24.9	182	45	(Gómez-Manzo et al., 2014)
<b>R393H</b>	I	31.2	90.6	119	44.3	(Gómez-Manzo et al., 2014)
<b>K238R</b>	I	6.96	24.77	71	40	(Gómez-Manzo et al., 2015)
<b>N126D &amp; D181V</b>	II	9.06	15.35	71	45	(Gómez-Manzo et al., 2015)
<b>N126D</b>	III	12.97	56.44	114	45	(Gómez-Manzo et al., 2015)
<b>R393G</b>	I	9.31	67.1	232	46	(Wang and Engel, 2009)
<b>R393H</b>	I	16.5	190	192	39	(Wang and Engel, 2009)
<b>G488V</b>	I	10.7	53.8	212	35	(Wang and Engel, 2009)
<b>G488S</b>	I	11.8	49.1	254	37	(Wang and Engel, 2009)
<b>R454C</b>	II	2.76	9.53	28.6	42.5	(Wang et al., 2005)
<b>R454H</b>	I	2.38	9.71	39.9	41.5	(Wang et al., 2005)
<b>G163D</b>	I	4.8	44.4	249	42.5	(Huang et al., 2008)
<b>G163S</b>	III	6.46	50.7	207	42.3	(Huang et al., 2008)
<b>R257L</b>	I	24	111	58	43	(Gómez-Manzo et al., 2016)
<b>V291M</b>	II	17	42	145	46	(Gómez-Manzo et al., 2016)
<b>L128P</b>	II	18	34	142	42	(Gómez-Manzo et al., 2016)
<b>V291M</b>	II	34.1	56.3	116	38	(Boonyuen et al., 2016)
<b>V291M &amp; G163S</b>	I	55.9	54.3	104	33	(Boonyuen et al., 2016)

## Supplemental Experimental Procedures

### *Cloning and mutagenesis of plasmid constructs*

Wild-type G6PD cDNA was inserted into the pET-28a vector with a 6xHis tag linked by a thrombin cleavage site, using NdeI and Sall restriction sites. Mutagenesis to obtain G6PD variants was performed by the GenScript USA Inc (Piscataway, NJ) mutagenesis service.

### *Enzyme purification*

Proteins were expressed in *E. coli* C43(DE3). Bacteria were grown at 37 °C until  $OD_{600} = 0.4-0.5$ , then expression was induced with 0.5 M IPTG for 4 h. Bacterial pellets were suspended in lysis buffer (50 mM Tris, 300 mM NaCl, 5% glycerol, 0.4 mM PMSF, 1 mg/mL lysozyme, 0.1% Triton X-100, protease inhibitor cocktail (Sigma, St Louis, MO), pH 7.8) and lysed by sonication. Lysate was clarified by spinning at 30,000g for 30 min, and supernatant was loaded onto 1 mL TALON Superflow beads (GE Healthcare, Pittsburgh, PA) for 1 h at 4 °C with rocking. Beads were washed with 30 mL wash buffer (50 mM Tris, 300 mM NaCl, 20 mM imidazole, pH 7.8) and resuspended in 2 mL PBS. 50 Units thrombin (R-BioPharm, Darmstadt, Germany) were added and beads were incubated overnight at 4 °C with rocking. The eluate was then purified on Superdex 75 gel filtration column (GE Healthcare) in buffer: 50 mM Tris, 150 mM NaCl, 0.5 mM EDTA, pH 7.4. Fractions containing G6PD were concentrated using a Vivaspin-6 30 kD MWCO spin column (GE Healthcare) and stored in 40% glycerol, 1 mM EDTA, 1 mM DTT at -80 °C. Protein concentration was measured using the Protein Assay Dye Reagent (Bio-Rad, Hercules, CA).

### *Enzyme activity and stability measurements*

Enzyme activity was determined following the WHO protocol (Organization, 1989), using 0.1–0.5 µg/mL G6PD and 100 mM Tris, 0.5 mM EDTA, 10 mM MgCl<sub>2</sub>, pH 8.0, and varying concentrations of G6P (Sigma) and NADP<sup>+</sup> (98% pure, Sigma).

For  $K_M$  and  $k_{cat}$  measurements, G6P concentration was kept constant at 500 µM while NADP<sup>+</sup> was varied from 1 to 200 µM, or NADP<sup>+</sup> concentration was kept constant at 200 µM while G6P was varied from 5 µM to 1 mM. Production of NADPH was determined by measuring fluorescence on a Flexstation II microplate reader (Molecular Devices Corporation, Sunnyvale, CA) using the NADPH-diaphorase-resazurin coupled reaction (Zhu et al., 2009). Relative Fluorescence Unit (RFU) values were then converted to NADPH concentration using a standard NADPH curve (NADPH 97% pure, Sigma).

For  $T_{1/2}$  measurements, 2 µg/mL enzyme was incubated with 5 µM NADP<sup>+</sup> for 20 min at temperatures ranging from 25 °C to 65 °C, then placed on ice for 5 min. Activity was then assayed by measuring NADPH via absorbance at 340 nm on a Bio-Rad Benchmark Plus microplate reader (Bio-Rad Laboratories, Inc., Hercules, CA).

Data were analyzed using GraphPad Prism v. 6.0a (GraphPad Software, La Jolla, California USA). To obtain kinetic parameters, the data were fitted using the preset nonlinear regression equations *Michaelis-Menten* and *Allosteric sigmoidal*. The best fit was selected using the *Compare* feature, and the following parameters were reported:  $k_{cat}$ ,  $K_M$ , and  $h$  (Hill coefficient, equal to 1 when *Michaelis-Menten* was the best fit). For  $T_{1/2}$ , the data were fit using the *Boltzmann sigmoidal* equation and the *Top* and *Bottom* values were set to 0 and 100, respectively. All values are reported as mean ± standard error.

Interestingly, we found that WT G6PD exhibited allosteric sigmoidal kinetics rather than Michaelis-Menten kinetics, which could be explained by the presence of multiple independent active sites in dimeric or tetrameric G6PD (Sweeny and Fisher, 1968). Furthermore, G6PD contains two binding sites for NADP<sup>+</sup> per monomer, so allosteric sigmoidal kinetics may be explained by binding cooperativity between these two sites.

## **Supplemental References**

Organization, W.H. (1989). Glucose-6-phosphate dehydrogenase deficiency. WHO Working Group. Bulletin of the World Health Organization 67, 601-611.

Sweeny, J.R., and Fisher, J.R. (1968). Alternative to allosterism and cooperativity in the interpretation of enzyme kinetic data. Biochemistry 7, 561-565.

Zhu, A., Romero, R., and Petty, H.R. (2009). An enzymatic fluorimetric assay for glucose-6-phosphate: application in an in vitro Warburg-like effect. Analytical biochemistry 388, 97-101.