Molecular mechanisms of functional impairment for active site mutations in glucose-6-phosphatase catalytic subunit 1 linked to glycogen storage disease type 1a.

Matt Sinclair[1], Richard A Stein[2], Jonathan H Sheehan[3,4], Emily M Hawes[2], Richard M O'Brien[2], Emad Tajkhorshid[1] and Derek P Claxton[2]*

[1]Theoretical and Computational Biophysics Group, NIH Center for Macromolecular Modeling and Visualization, Beckman Institute for Advanced Science and Technology, Department of Biochemistry, and Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

[2]Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN 37232

[3]Center for Structural Biology, Vanderbilt University, Nashville, TN 37240

[4]Department of Internal Medicine, Division of Infectious Diseases, Washington University School of Medicine, St Louis, MO 63110

*Corresponding Author
Email: derek.p.claxton@vanderbilt.edu

Competing Interest Statement: The authors declare no competing interest.

Classification: BIOLOGICAL SCIENCES; Biophysics and Computational Biology.

Keywords: glycogen storage disease, G6Pase, AlphaFold2, Rosetta, molecular dynamics

This PDF file includes:

Main Text
Figures 1 to 6
Table 1
Supplementary Figures 1 to 7

Abstract

Mediating the terminal reaction of gluconeogenesis and glycogenolysis, the integral membrane protein G6PC1 regulates hepatic glucose production by catalyzing hydrolysis of glucose-6-phosphate within the lumen of the endoplasmic reticulum. Because G6PC1 function is essential for blood glucose homeostasis, inactivating mutations cause glycogen storage disease (GSD) type 1a, which is characterized by severe hypoglycemia. Despite its physiological importance, the structural basis of G6P binding to G6PC1 and the molecular disruptions induced by missense mutations within the active site that give rise to GSD type 1a are unknown. Exploiting a computational model of G6PC1 derived from the groundbreaking structure prediction algorithm AlphaFold2 (AF2), we combine molecular dynamics (MD) simulations and computational predictions of thermodynamic stability with a robust *in vitro* screening platform to define the atomic interactions governing G6P binding within the active site as well as explore the energetic perturbations imposed by disease-linked variants. From over 15 μs of MD simulations, we identify a collection of side chains, including conserved residues from the signature phosphatidic acid phosphatase motif, that contribute to a hydrogen bonding and van der Waals network that stabilize G6P in the active site. Introduction of GSD type 1a mutations into the G6PC1 sequence causes changes in G6P binding energy, thermodynamic stability and structural properties, suggesting multiple mechanisms of catalytic impairment. Our results, which corroborate the high quality of the AF2 model as a guide for experimental design and to interpret outcomes, not only confirm active site structural organization but also suggest novel mechanistic contributions of catalytic side chains.

Introduction

Interconversion of glucose between its free and phosphorylated forms is among the most central reactions in metabolism, controlling its transport between the cytoplasmic and extracellular regions, and thereby its storage and homeostasis. For example, glucose-6-phosphatase (G6Pase) couples the uptake of intracellular glucose-6-phosphate (G6P) into the endoplasmic reticulum (ER) lumen (1) with G6P hydrolysis into glucose and inorganic phosphate (Pi) (2, 3). The hydrolysis reaction is catalyzed by one of three membrane-embedded G6Pase catalytic subunit (G6PC) family members that differ in their tissue expression patterns (4). The terminal reaction of gluconeogenesis and glycogenolysis is mediated by G6PC1, which is predominantly expressed in the liver, kidney and small intestines (5-7). Consequently, G6PC1 serves as the gatekeeper for hepatic glucose production which stabilizes blood glucose between meals. By residing at the crux of glucose homeostasis, G6PC1 is a potential therapeutic target for disorders of glucose metabolism. Whereas upregulation of *G6PC1* gene expression contributes to diabetes by increasing fasting blood glucose concentration (8-17), aberrant mutations in G6PC1 that impair activity cause glycogen storage disease (GSD) type 1a (18, 19).

With an approximate incidence of 1 in 100,000 births, GSD type 1a is an autosomal recessive disorder characterized primarily by severe hypoglycemia following a fast (20). Due to inactive G6PC1, the accumulation of intracellular G6P stimulates flux through alternate metabolic pathways including glycogenesis that promotes excessive glycogen storage in the liver. In addition to hepatomegaly, GSD type 1a is also associated with a host of other metabolic complications including nephromegaly, hyperlipidemia and growth retardation. The disorder is juvenile lethal if left untreated, yet therapeutic interventions are limited presently to dietary restrictions supplemented with a drug regimen to treat the disease complications (21). According to the Human Gene Mutation Database, more than 100 pathogenic missense/nonsense mutations scattered throughout the *G6PC1* coding sequence have been identified from GSD type 1a patients (22). However, the structural and mechanistic basis for catalytic dysfunction of these mutations remains to be elucidated.

Despite considerable effort since its discovery in the early 1950s (7), G6PC1 extracted from its native environment has been recalcitrant to *in vitro* investigations due to structural and catalytic instability (23-26), precluding detailed analysis of the structure/function paradigm. In 2021, DeepMind revolutionized the field of structural biology with the introduction of AlphaFold2 (AF2), a deep learning platform that integrates knowledge

of protein folds with residue co-evolution encoded within deep multiple sequence alignments to predict remarkably accurate structures of proteins (27). Highlighted in one of two seminal publications (28), the AF2 model of G6PC1 displayed a fold consistent with the type 2 phosphatidic acid phosphatase (PAP2) superfamily that includes membrane-integrated lipid phosphatases and water-soluble haloperoxidases (29) as well as the predicted topology of G6PC1 from biochemical experiments (30, 31). The defining feature of this superfamily is a signature tripartite sequence motif (KX_6RP---PSGH---SRX_5HX_3D/Q) that participates in active site formation (32) and includes sites of GSD type 1a mutations (18). Although the AF2 model recasts G6PC1 and disease-linked mutations in a previously absent structural context, notably absent from the computational prediction is an understanding of fundamental substrate binding and interactions critical to the catalytic mechanism.

In this work, we use molecular dynamics (MD) simulations to identify the G6P binding mode and interactions with the G6PC1 AF2 model immersed in a representative ER membrane. Moreover, we predict the molecular consequences to G6P binding energetics and G6PC1 thermodynamics by introducing a select panel of GSD type 1a mutations into the active site *in silico*. Capitalizing on our recently published *in vitro* methods that overcome previous experimental limitations with purified enzyme (33), we complement the computational evaluation of the AF2 model with biochemical and biophysical studies of G6PC1 variants to outline changes in expression, activity, and stability. We find that the AF2-generated model is remarkably stable and binds both α- and β-G6P anomers with a nearly identical pattern of electrostatic and van der Waals interactions within the predicted active site. Perturbation of these interactions through disease-linked mutations not only reveals changes to G6P binding energetics but also to G6PC1 backbone dynamics and intrinsic stability. The integrated analysis uncovers novel side chain contributions to Pi release and suggests that GSD type 1a mutations disrupt catalysis via distinct molecular pathways.

Results

Structure and stability of G6PC1 in a model lipid bilayer

We chose to exploit mouse G6PC1 as an archetype G6Pase catalytic subunit for the structural and mechanistic questions addressed here. Mouse G6PC1 bears 89% sequence identity with the human ortholog and is predicted by AF2 to possess a nearly identical tertiary fold (root mean squared deviation (RMSD) ~0.125

Å) emphasized by a conserved active site organization. Globally, 65 of the 69 amino acids at positions of the naturally occurring GSD type 1a missense/nonsense mutations, including all of those at the active site, are strictly conserved between hG6PC1 and mG6PC1. Moreover, a murine model of GSD type 1a (G6PC1$^{-/-}$) recapitulates a similar phenotype as human patients (34, 35). Importantly, detergent-solubilized mouse G6PC1 demonstrates elevated structural and catalytic stability relative to human G6PC1 in biochemical assays (33), supporting an unequivocal interpretation of our *in vitro* studies. Thus, mouse G6PC1, subsequently referred to simply as G6PC1, is an appropriate model for computational studies of substrate binding and experimental evaluation of structural and functional perturbations associated with disease-linked mutations.

Because the AF2 modeling pipeline does not account explicitly for the membrane, the G6PC1 model was placed into an approximate ER membrane to assess its relative stability (Fig 1a). G6PC1 was almost entirely buried within the membrane with the active site residing at the membrane-water interface. This model, which was only gently relaxed from its initial state during the minimization phase, was then subjected to equilibrium MD simulations, which converged within 100 ns in multiple iterations to arrive at a common structural model suitable for ligand docking and thermodynamic analysis. The aggregate simulation data yielded a final all-atom RMSD of 1.76 Å relative to the original AF2 model (Fig 1b), consistent with a very stable structure within experimental error (27, 36). Regions of higher internal flexibility are highlighted by root mean squared fluctuation (RMSF) analysis of the simulation dataset (Fig 1c) and correlate well with unstructured loops and regions of lower confidence from the AF2 prediction.

Modeling the G6P-bound active site

We leveraged available crystallographic structures from one vanadium-containing chloroperoxidase (37) and three bacterial PAP2 type lipid phosphatases (38-40) that contain surrogate phosphate molecules (SO_4^{2-}, WO_4^{2-}, and VO_4^{3-}) to guide docking of the G6P substrate into the putative G6PC1 active site. These protein structures, when aligned with the AF2 model of G6PC, highlight a strongly conserved active site organization composed of key residues in the phosphatase sequence motif (SI Appendix, Fig. S1a). We mapped the phosphate moiety of both α- and β-G6P anomers, derived from 95 crystal structures, to the AF2 G6PC1 model by targeting the atomic coordinates belonging to the phosphate analogs. Although placement of the phosphate

group for each modeled G6P was well defined, a “cloud” of possible conformations was showcased by the glucose moiety because of dihedral flexibility (SI Appendix, Fig. S1b, e).

Two conformations representing an α- and a β-G6P anomer were chosen for preliminary docking onto the bilayer-relaxed G6PC1 structure (SI Appendix, Fig. S1c). These two G6P-bound models were subjected to additional 250 ns equilibrium MD simulations performed in triplicate to assess stability. Although repacking of critical active-site residues was observed early in the simulations (SI Appendix, Fig. S1d), RMSD analysis indicated stability and robustness similar to the Apo state (Fig 2a). Both G6P anomers induced only marginal structural perturbations relative to the Apo model (< 2 Å RMSD). However, free energy perturbation calculations indicated that binding of the β anomer is more favorable than α-G6P by ~5 kcal/mol, which may reflect more stabilizing interactions with G6PC1 facilitated by the lower energy configuration of the anomeric carbon (SI Appendix, Fig 2). Of note, the coordinating residues for both anomers, constituting electrostatic and van der Waals interactions, were largely similar with the exception that A120, A175, and R170 make preferentially contacts with β-G6P during the simulations (Fig 2b). Fig 2c-d captures a representative binding mode sampled for each G6P anomer over the course of the simulations as determined by a minimal sum of distances between the catalytic K76, R83/170, and H119/176 side chains. Collectively, the structural convergence of multiple independent simulations as well as the stable and consistent pattern of interactions with G6P strongly support the predicted active site identity of G6PC1.

Energetic perturbations arising from clinically relevant mutations

Nine of the 17 residues that interact with G6P in our model are sites of naturally occurring GSD type 1a mutations in human G6PC1 (Table 1). Previous *in vitro* characterization of these missense variants in the context of ER microsomes reported that G6P catalysis was abrogated strongly (19), but the molecular basis of compromised function was unclear. Based on the G6P-bound model alone, mutation of coordinating residues may affect substrate binding energy. Alternatively, side chain substitutions may destabilize critical packing interactions. Indeed, calculation of changes in Gibbs energy (ΔΔG) by the Rosetta macromolecular modeling suite (41-44) predicted that six of these variants (D38V, K76N, R83H, E110K, P113L and T255I) promoted thermodynamic instability greater than 2 kcal/mol in the bilayer-relaxed Apo model (Table 1, SI Appendix, Fig

S3). By contrast, H119L and R170Q enhanced stability. The significance of these findings was emphasized by ΔΔG analysis of two uncharacterized missense variants, Q14R and E319K, listed in the Genome Aggregation Database (45) that are found outside the active site on the protein surface and expected to be well tolerated (Table 1). These predicted thermodynamic perturbations suggest that compromising native side chain energetic interactions could disrupt active site stability, potentially contributing to local or global misfolding. We further explored these possibilities by a series of computational and biochemical assays.

Residue-specific interaction energies with both G6P anomers were calculated using NAMD (46, 47). The time-series of energy distributions were binned from three independent 250-ns replicates for a total of 54 mutant simulations accumulating 13.5 µs of sampling. The consequences of the *in silico* mutagenesis on G6P binding are shown in Figure 3, which can be categorized into three distinct classes. In one class, the enthalpic component to G6P interactions was largely abolished by non-conservative substitution of key phosphatase residues K76, R83 or H119 incurring ~20-50 kcal/mol loss of interaction energy in each. According to the model, these positively-charged native side chains are positioned optimally to coordinate the negatively-charged phosphate moiety of G6P (Fig 2). In a second class, the introduction of the D38V, P113L, G118S or T255I mutations was far less impactful toward G6P binding energetics, suggesting that impaired function of these variants originates from a distinct mechanism. Interestingly, a third class consisting of E110K and R170Q was defined by a left shift in the interaction energy distribution indicating greater stabilization of the substrate in the binding pocket. In a proposed hydrolysis model, R83 and R170 donate H-bonds to stabilize the phosphate moiety and the phosphoenzyme transition state (32, 48). Recently, E110 was postulated to serve as a surrogate substrate by stabilizing a closed conformation of the active site through salt bridges with the nearby R83 and R170 (28). However, the MD simulations suggested that E110 and R170 are more destabilizing to G6P binding than K110 or Q170. This observation likely indicates that while the native Glu (E) and Arg (R) side chains are critical for product release, the E110K and R170Q variants may favor trapping of G6P in the active site to reduce or eliminate hydrolysis.

Expression and activity profiles of GSD type 1a variants

To explore the hypotheses derived from our computational analysis, we tested the biological outcomes of these active-site GSD type 1a mutations as well as the Q14R and E319K variants by employing the enzyme expression and activity screens that we recently developed (33). Specifically, WT and mutant G6PC1 were transfected into adherent HEK293SG cells as a fusion with EGFP on the C-terminus to facilitate assay analyses. G6PC1 expression was confirmed via cell epifluorescence (SI Appendix, Fig S4). Relative quantities and homogeneity of G6PC1 fusion protein were determined by EGFP fluorescence detection size exclusion chromatography (FSEC) (49) following whole cell solubilization with 5mM lauryl maltose neopentyl glycol (LMNG) detergent. LMNG was identified previously from a detergent screen to support extraction of active and homogeneous G6PC1 (33).

Representative FSEC traces in Figure 4a illustrate the spectrum of chromatographic behaviors observed for the variants. The full complement of traces is shown in Fig S5 (SI Appendix). In general, expression of disease-linked variants was either unperturbed or greatly reduced relative to WT. Additionally, the elution profiles, which indicate the presence of either polydisperse entities or folded monodisperse protein, reported a variable propensity for aggregation. Notably, the R83H and P113L mutations displayed highly compromised G6PC1 expression with elution profiles consistent with a large degree of structural disorder. Although D38V and E110K expression levels were attenuated strongly, the peak shapes suggested that these mutants largely retained the structural properties of the WT. Interestingly, K76N, G118S, H119L, and R170Q behaved similarly to WT, Q14R, and E319K. These features were captured by integration of the elution peak for multiple independent trials, translating the FSEC traces into an expression bar graph for each variant relative to WT (Fig 4b). The observed trends were comparable to the whole cell epifluorescence pattern (SI Appendix, Fig S4). Statistical analyses of these data supported the conclusion that expression of Q14R, K76N, G118S, H119L, R170Q and E319K was unchanged relative to WT.

Apart from D38V, *in vitro* hydrolysis measurements of LMNG-solubilized G6PC1 with a saturating concentration of G6P revealed a binary enzyme activity profile (Fig 4c). Not surprisingly, the absence of G6P hydrolysis in R83H and P113L was directly linked to either absent or poor expression of folded protein. Even though E110K expressed better than D38V without gross misfolding, catalysis was abolished as expected. Likewise, K76N, G118S, H119L and R170Q were catalytically inactive despite similar expression as WT. We

further confirmed the impaired activity of K76N and H119L with an *in situ* luciferase reporter assay that indirectly measures G6Pase activity within intact cells (50, 51). This assay capitalized on a cell line derived from rat islets, 832/13, which displays limited endogenous G6Pase activity. These cells express a transfected rat *G6PC1-luciferase* gene fusion when robustly induced with glucose, stimulating G6P production via glucokinase activity. In the absence of hydrolysis, G6P binds to the luciferase promoter to drive fusion gene expression (Mock, Fig 4d). As expected, co-transfection of functional WT or Q14R blunted glucose-induced fusion gene expression by mediating G6P hydrolysis and reducing glycolytic flux (Fig 4d). On the other hand, cells co-transfected with the K76N and H119L variants behaved similarly to the mock control, consistent with these variants being catalytically impaired.

Trends in backbone dynamic fluctuations induced by GSD type 1a variants

The particularly striking observation that K76N, G118S, H119L, and R170Q retained WT-like expression properties demonstrated that these substitutions are structurally tolerated, precluding a clear correlation with the predicted $\Delta\Delta G$. Given that energetic disruptions to G6P binding were unequal among these variants (Fig 3), we considered that the molecular consequences of these substitutions may be conferred to the protein backbone. We parsed the available 13.5 μs of combined simulation data between the two G6P anomers to determine if there were observable trends in the pattern of backbone dynamics across the nine variants. Comparison of the root mean squared fluctuations (RMSF) in the protein backbone when the active site is occupied with either G6P anomer identified potential hot spots along the primary sequence for dynamic perturbations relative to the WT (Fig 5a). Heat maps of ΔRMSF depicted on the G6PC1 model highlighted the nature and magnitude of these fluctuations observed within the MD simulation time course, which displayed some dependence on the G6P anomer (Fig 5 and SI Appendix, Fig S6). The location of these perturbations mostly tracked with the site of substitution, although the appearance of longer-range fluctuations were observed especially within the large luminal loops that contribute to active site formation, even within the relatively short simulation time. In particular, ΔRMSF was notably different within the large luminal loops that contribute to active site formation. Interestingly, the K76N, G118S, H119L and R170Q variants reduced the apparent flexibility of the backbone in multiple regions (blue shading, Fig 5b-g). As a corollary, these changes in backbone dynamics were accompanied by both local

(proximal to the substitution) and global (distal to the substitution) structural distortions, especially within the large luminal loops (SI Appendix, Fig S7). Calculation of the average pairwise residue-residue distances uncovered pockets of either constriction or expansion to accommodate side chain repacking in the active site. Taken together, the computational analysis emphasizes the potential for active-site variants to drive structural alterations, including within the large luminal loops, that may be causative or degenerate with other mechanisms of catalytic impairment, such as reduced thermodynamic stability.

Correlation of hydrolysis kinetics with thermostability of the D38V variant

Despite substantially reduced expression (~10% of WT), D38V was the only disease-linked variant that retained catalytic activity significantly above background (Fig 4c). Yet the hydrolysis screen indicated that this variant retained only ~30% of the Pi release capacity from the equivalent concentration of WT enzyme. Although the ΔRMSF calculation suggested changes in backbone dynamics near position 38, global fluctuations were more similar to WT or Q14R than other variants (SI Appendix, Fig S6). Moreover, the pairwise distance matrices implied limited repacking induced by the substitution (SI Appendix, Fig S7). We therefore sought to define the kinetic and structural properties of D38V when purified into LMNG detergent micelles. Unlike Q14R, which displayed nearly identical saturation kinetics as the WT (33), the specific activity of D38V was reduced three-fold without a shift in K_M (Fig 6a). Thus, the diminished hydrolytic capacity of D38V was a direct consequence of a lower Pi release rate.

We reasoned that the origin of reduced turnover velocity for D38V may be associated with compromised structural stability as predicted by the ΔΔG calculations (Table 1). To probe this hypothesis, we measured the thermostability of D38V and compared it to that obtained for Q14R. Variant thermostability was determined by a chromatography-based approach wherein purified protein samples subjected to a thermocycler heating protocol were analyzed for soluble protein content by FSEC monitoring intrinsic Trp fluorescence (33, 52). Plotting the elution peak integral as a function of temperature produced a curve characterized by a melting temperature, T_M. As shown in Figure 6b, the D38V melting curve was left-shifted relative to Q14R, which was indicative of a loss of heat tolerance. Non-linear least squares fitting of the data confirmed a 5 °C T_M differential between the two variants. Interpreted as a reporter of $\Delta G_{unfolding}$, this phenomenological analysis of the D38V variant was

consistent with compromised thermodynamic stability. Collectively, the reduced thermostability of D38V rationalizes both the lower abundance of folded G6PC1 and impaired catalysis.

Discussion

It is widely recognized that systematic exploration of protein structure and function is essential to defining pathogenic mechanisms arising from missense mutations and that reliable protein models are a prerequisite for contextualizing such biochemical data (53-55). Homology models derived from high resolution structures of related proteins have been proposed to serve as sufficient templates for the interpretation of disease-causing variants (56, 57), although AF2 models are expected to further improve the accuracy of the predictions (58). Due to the integration of deep multiple sequence alignments (MSAs) in the AF2 prediction pipeline, it is unclear if strict assessment of missense mutations on protein structure will uncover either strong correlations or predictive outcomes between the WT and variant sequences for a given metric (59-61). Indeed, rational manipulation of the MSA through subsampling (62, 63) or introduction of strategic point mutant combinations (64) have been necessary to explore structural heterogeneity of AF2 model predictions. However, a recent evaluation of AF2 applications showed that high accuracy AF2 models (pLDDT > 90) at least matched or improved the correlation between experiment-derived $\Delta\Delta G$ and $\Delta\Delta G$ from structure-based predictors, such as Rosetta, relative to a structure obtained from experiment (58). We show here that combining a high accuracy AF2 model of G6PC1 with orthogonal computational and experimental approaches supports mechanistic insight of critical catalytic interactions and characterization of the molecular disruptions imposed by known disease-linked missense mutations.

We find that the AF2 G6PC1 model is highly stable in the simulated *in vivo* environment and that the two stereoisomers of G6P make favorable electrostatic and van der Waals interactions with a network of side chains in the putative active site. This active site, outlined by the consensus phosphatidic acid phosphatase sequence motif, demonstrates similar packing of conserved side chains observed in crystal structures of evolutionary-distant PAP2 superfamily members with bound quaternary ions. Despite subtle repacking of active site side chains in the MD simulations to optimize G6P interactions, the overall conformation of G6PC1 remained intact, suggesting that the AF2 model represents a relevant intermediate in the catalytic cycle. Numerous

phosphorylated substrates for G6PC1 have been identified (7), so the discovery that binding of the β-G6P anomer would be favored by up to three orders of magnitude under equilibrium conditions ($K \approx 3300$ at 37 °C) was unexpected. Since spontaneous mutarotation of G6P can occur in solution, selectivity for the β-anomer was thought to be exclusively conferred by the cognate SLC37A4 transporter (65). We speculate that preference for β-G6P is associated in part with three residues, namely A120, R170 and A175, that show difference in their contact probabilities with the two anomers (Fig 2), which may be influenced by sugar pucker angle sampling (SI appendix, Fig S1e).

The *in silico* mutagenesis, facilitating both non-bonded side chain interaction energy analysis and ΔΔG predictions, largely confirms the anticipated contribution of conserved active-site residues in substrate coordination, but also suggest multiple modes of inhibition that are not necessarily mutually exclusive. For instance, the charge reversal and charge neutralizing variants E110K and R170Q substantially increase the binding energy with G6P, which may imply that these variants trap the substrate within the positively charged active site. Moreover, this observation also predicts that the native Glu and Arg residues, though bearing opposite charges, are involved in Pi release. This could be achieved mechanistically through dynamics of salt bridge formation and rupture between E110 and R170 over the course of a catalytic cycle, functioning as a modulator of active site polarity to trigger product release.

Interestingly, the Rosetta protein stability predictor estimates divergent effects of E110K and R170Q on $\Delta G_{unfolding}$ (Table 1). In concert with the predictions, the destabilizing E110K variant displays highly attenuated expression of folded protein whereas the stabilizing R170Q reports similar expression as WT. For most variants, a general pattern emerged in which destabilizing substitutions caused either reduced expression or misfolding while neutral or stabilizing substitutions retained WT-like expression profiles. However, this pattern was not explicitly predictive of expression levels with respect to the value of ΔΔG, nor could ΔΔG be assigned as a predictor for activity. For example, while K76N is expected to be moderately destabilizing (+2.8 kcal/mol), ΔΔG demonstrates no correlation with expression and the variant is completely non-functional likely due to impaired G6P binding (Fig 3) combined with restricted backbone dynamics (Fig 5). On the other hand, D38V is predicted to be similarly destabilizing (+3.7 kcal/mol), and expression is reduced markedly, but it retains a fractional yet significant level of activity relative to WT.

The collective body of work highlights the complexity for which missense mutations found in the active site of G6PC1 conspire to sabotage function via multiple mechanisms. In summary, the disruption of D38V, R83H and P113L variants are driven predominately by thermodynamic instability leading to reduced G6P turnover and expression levels or cause gross misfolding. The remaining variants (K76N, E110K, G118S, H119L and R170Q) appear primarily to alter energetic barriers to G6P binding or access to functional intermediates along the reaction coordinate ($\Delta G°$) generating catalytically trapped states, which may be reflected in changes to backbone fluctuations, induced structural distortions or impaired conformational dynamics. In contrast, Q14R and E319K, variants of uncertain clinical significance, were found to have a benign phenotype in this study and so would be expected to not confer disease. Importantly, the relationship between the missense mutation and dysfunction required integration of multiple computational approaches, including the availability of a highly accurate and stable 3D model, with recently developed experimental protocols to measure biological outcomes because no single determinant explained all aspects contributing to catalytic inhibition. Since the variants explored here are but a small subset of the 72 known GSD type1a missense mutations in G6PC1, this study supports further characterization of the full complement of variants using the AF2 structural model as a tractable template.

Materials and Methods

Modeling of Apo G6PC1 in a simulated ER membrane bilayer

The structural model of mouse G6PC1 (Uniprot P35576) was acquired from the publicly available AF2 database (https://alphafold.ebi.ac.uk/). The initial equilibrium simulation was constructed using the CHARMM-GUI web server. An approximate ER lipid bilayer composition was chosen consisting of 55% POPC, 20% POPE, 10% POPI, 5% POPS, 5% cholesterol, and 5% sphingomyelin for both membrane leaflets. The position of G6PC1 relative to the bilayer was computed using the Orientation of Proteins in Membranes server (OPM) being facilitated by CHARMM-GUI to ensure proper placement. The pKa of titratable residues was calculated by PropKa3.1. H119 was protonated (HSP) as predicted by its catalytic role as a proton donor in the reaction mechanism. All other histidine residues were assigned neutral (HSD).

The model was relaxed from its initial state over a series of equilibration steps with decreasing restraints using NAMD2 and the CHARMM36 force field. First a 10,000 step minimization was performed after which a slow heating protocol of increasing the system temperature by 25 K every 10 ps until 310 K was reached. During heating and the first 2.5 ns of equilibration there was a 1 kcal/mol/Å^2 harmonic restraint was placed on lipid head group and protein backbone atoms to allow the lipid tails to melt. This was followed by 2.5 ns of simulation with a 1 kcal/mol/Å^2 restraint only on the protein backbone, and a final 2.5 ns of simulation with a 0.5 kcal/mol/Å^2 restraint on protein alpha carbons. Five nanoseconds of unrestrained equilibration was then performed to allow for switching to NPT and subsequent relaxation of the system.

Simulation replicas were generated after equilibration using the Membrane Mixer tool in VMD in order to sufficiently sample the initial lipid configuration of the system. Production MD runs were carried out using NAMD3 with a timestep of 2 fs and a long-range interaction cutoff of 14 Å. In production runs, each replica was simulated for 100 ns resulting in a cumulative total of 300 ns of trajectory data.

In silico mutagenesis and thermodynamic predictions

ΔΔG calculations were performed to predict the thermodynamic effects of missense variants using Rosetta following the method described in Ref (41). Briefly, the model was oriented in an implicit membrane environment by aligning it to the OmpLA structure provided in $ROSETTA/main/demos/public/mp_ddg/inputs/1qd6_tr_C.pdb using UCSF Chimera 1.16 (66). The spanfile was generated using the consensus server at TOPCONS.net (67) and edited to optimally specify the membrane-spanning helices. The structure was relaxed into Rosetta's franklin2019 score function using 1000 independent runs of the rosetta_scripts executable version 2020.37.61417 with the membrane_relax.xml protocol. The lowest-scoring relaxed decoy was then used as the input for the predict_ddG.py script with a repack radius of 8.0 Å. The predict_ddG.py script was adjusted to run under pyrosetta-2021.49 release r305 (Python 3.8.12, Conda version 5.0.1) inside a Slurm array of 357 jobs on Vanderbilt's ACCRE cluster to assess all 20 possible mutations at each position in the protein. The ΔΔG values of disease-associated variants were compiled from the output files using GNU Awk 4.0.2, mapped onto the structural model using Pymol 2.5.4 (68) (SI Appendix, Fig S3) and reported in Table 1. The error associated with these predictions is expected to be within 1 kcal/mol

on average (42, 44, 69). Therefore, any value greater than 2 kcal/mol was considered destabilizing, whereas any value less than -2 kcal/mol was considered stabilizing.

Preliminary docking of G6P

Crystal structures of bacterial PAP2 type lipid phosphatases (PDB 6EBU, 6FMX and 5JKI) and vanadium containing chloroperoxidase (PDB 1IDQ) were aligned with the AF2 G6PC1 model using the catalytic Arg, His and Lys side chains from the signature phosphatase sequence motif. In MATLAB (MathWorks), the PO_4^{3-} moiety of the α- and β-G6P anomers (also obtained from crystal structures) were transformed onto the atomic coordinates of the quaternary ions of SO_4^{2-}, WO_4^{2-}, and VO_4^{3-} from the aligned structures yielding 95 potential configurations of G6P. Any G6P model within 1.2 Å of the AF2 model, whether in side chains or backbone atoms, was discarded to prevent van der Waals clashes. From the remaining G6P models, two were chosen as representative α- and β-G6P for further MD simulations.

Equilibrium simulations with docked G6P

A similar equilibration protocol to the Apo simulations was followed with the notable exception of including G6P restraints with the same force constant as the protein heavy atoms throughout restrained equilibration. Parameters for G6P were obtained from phosphorylated sugar parameters in the CHARMM36 force field. The same initial lipid configuration as in the Apo replicas was used to seed the docked simulations and each was simulated for 250 ns of production MD for each G6P anomer.

From the initial WT state, nine mutants were also generated (Table 1) using the Mutagenesis tool in VMD. In each mutant state, the system charge was adjusted accordingly by reionization to ensure a net charge of 0. Each replica was equilibrated as above for the WT systems and simulated for 250 ns resulting in a cumulative total of 15 μs of trajectory data.

Free Energy Perturbation calculations

Free energy perturbation (FEP) was performed using a representative model for both G6P anomers. In each system the substrate molecule was duplicated and placed over 25 Å away from the protein in the bulk

solvent. The resulting system was re-minimized and then gently equilibrated for 30 ns to ensure accurate results. The protein backbone was restrained as well as the G6P phosphorous atom throughout the simulation with a force constant of 1 kcal/mol/$Å^2$ to allow for side chain rearrangement and substrate internal sampling to still occur. Each FEP protocol consisted of 50 windows of lambda values ranging from 0 to 1 and incrementing/decrementing by 0.02 each window. Each window was simulated for 1 ns for a total of 50 ns per FEP simulation set in each direction.

Free energies were extracted from the trajectories using the Analyze FEP plugin in VMD. Gram-Charlier expansion was set to 0 and the Bennet Acceptance Ratio (BAR) estimator was used to compute error. Forward and backward paths were plotted to ensure convergence of the simulation set. Relative binding free energy was then calculated and reported.

Screening G6PC1 expression and activity *in vitro*

Measurements of expression and G6P hydrolysis following transfection of adherent HEK293SG cells (*N*-acetylglucosaminyl-transferase I-negative; ATCC CRL-3022) followed the previously published methods (33). Briefly, WT mouse DNA (accession number NM_008061) was cloned into the pJPA5 MOD expression vector. EGFP was fused to the C-terminus of G6PC1 with a linker that included a thrombin protease recognition sequence. Mutations were introduced via site directed mutagenesis using complementary oligonucleotide primers, and DNA sequencing confirmed the presence of the desired mutation and the absence of unwanted changes. Adherent semiconfluent HEK293SG cells cultured in DMEM:F12 medium supplemented with 10% FBS were transfected in 6-well plates with 2μg/well of plasmid DNA complexed with Lipofectamine 2000 reagent (Invitrogen). The cells were incubated for 48 hrs at 37 °C under 7% CO_2. Expression was confirmed by visualizing cell epifluorescence on a Zeiss Axio Zoom.V16 fluorescence stereo microscope. Epifluorescence intensity was quantified for mock (H_2O), WT and variants across multiple independent transfection experiments (n=3-7) using the ImageJ application (70) running the StarDist plugin.

Following harvest, the cells were solubilized in 300 μL buffer containing 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 5 mM (0.5% w/v) LMNG and 5mM PMSF. Insoluble material was removed by ultracentrifugation at 105,000 rcf for 20 mins. Supernatant containing solubilized G6PC1-EGFP was injected

onto a Superose6 Increase 10/300 GL column equilibrated in 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA and 0.01% (w/v) LMNG buffer. The column was attached to an Agilent 1260 Infinity II chromatography system equipped with a fluorescence detector and a temperature-controlled multi-sampler. Sample elution profiles monitored EGFP fluorescence (Ex 475 nm, Em 515 nm). The enzyme concentration was estimated by integration of the elution peak obtained from 15-17 mL and normalized to the WT for each transfection experiment (n=4-9). Variability in the calculated expression level (Fig 4) was a function of transfection efficiency compounded with detergent solubilization efficiency for each experimental iteration. One-way ANOVA and the Tukey test were used to assess statistical variation in the mean epifluorescence and FSEC peak area datasets at the 0.05 level.

Phosphohydrolase activity was measured by diluting 20 μL of LMNG-solubilized enzyme into a final volume of 150 μL 50 mM Tris/Mes pH 6.5, 50 mM NaCl, 0.2mM LMNG in the presence or absence of 1.5 mM G6P and mixed on ice. The reaction was transferred to a 30 °C water bath for five minutes to stimulate G6P hydrolysis. The reaction was quenched with 150 μL of 12% (w/v) SDS and vortexed. The amount of Pi released was determined from a colorimetric assay relative to a Pi standard curve as previously described. Absolute Pi released was normalized to enzyme concentration determined from the FSEC peak area and then scaled relative to the WT. Reactions were performed in triplicate per biological repeat (n=4-9), and the means subjected to ANOVA and Tukey tests.

Measurement of glucose-6-phosphatase (G6Pase) activity *in situ*

G6Pase activity was measured *in situ* as previously described (50). Briefly, semi-confluent 832/13 cells in 3.5 cm diameter dishes were co-transfected with 2 μg of a *G6pc1*-firefly *luciferase* fusion gene construct, 0.5 μg of SV40-*Renilla luciferase* (Promega, Madison, WI) and 1 μg of an expression vector encoding WT or mutated G6PC1-EGFP, using the lipofectamine reagent (InVitrogen, Waltham, MA). Following transfection, cells were incubated for 18-20 hours in serum-free medium supplemented with 2 or 30 mM glucose. Cells were then harvested using passive lysis buffer (Promega, Madison, WI) and both firefly and *Renilla* luciferase activity were assayed using the Dual Luciferase Assay kit (Promega, Madison, WI). To correct for variations in transfection efficiency, the results were calculated as a ratio of firefly activity to protein concentration in the cell lysate. The

four G6PC1 constructs (WT, Q14R, K76N and H119L) were measured in duplicate with n=3 independent biological replicates. The means were subjected to ANOVA and Tukey tests.

Expression and purification of G6PC1 from Sf9 insect cells

Heterologous expression and purification of Q14R and D38V followed the previously published protocols with minor adjustments (33). Briefly, G6PC1 in the pFastBac1 vector was expressed as an EGFP fusion that contains a C-terminal His_8 tag via baculovirus transduction of Sf9 insect cells for 72 hrs at 27 °C. Isolated membranes were solubilized with 5 mM (0.5% w/v) LMNG in 50 mM Tris pH 8, 100 mM NaCl, 10% (v/v) glycerol buffer for 1 hr on ice. Insoluble material was removed by ultracentrifugation at 185,000 rcf for 50 min. The supernatant was filtered through a 0.45 μm syringe filter and mixed with Ni^{2+}-NTA Superflow (Qiagen) equilibrated in the same buffer supplemented with 25 mM imidazole for 4 hrs at 4 °C with gentle shaking. The resin was washed with 10 column volumes of buffer containing 100 mM imidazole followed by elution of G6PC1-EGFP with 300 mM imidazole buffer. EGFP was cleaved via thrombin digestion (0.04 NIH units/μg of fusion protein) at 15 °C for 16 hrs. EGFP and thrombin was removed from G6PC1 by gel filtration through a Superose6 Increase 10/300 GL column equilibrated with 50 mM Tris pH 7.5, 150 mM NaCl, 0.2 mM LMNG. Enzyme concentration was determined from absorbance at 280 nm assuming a 90,340 M^{-1} cm^{-1} molar extinction coefficient calculated from the protein sequence. Sample purity was assessed by SDS-PAGE using a 13% acrylamide gel. Chromatographic behavior was assayed with FSEC by injecting 7 μg of the purified enzyme onto the Superose6 Increase 10/300 GL column and monitoring fluorescence (Ex 280 nm, Em 320 nm).

Measurement of G6P hydrolysis and thermostability with purified G6PC1 variants

The Pi release rate was determined from titration of 0.1 μg (2.4 pmol) G6PC1 with G6P at 30 °C for 1 min at pH 6.5 as previously described (33). A linear [G6P]-dependent curve collected on ice was subtracted from reaction carried out at 30 °C. Catalytic parameters were determined from fits of the saturation curve assuming a Michaelis-Menten model. The reported mean values and standard deviations of K_M and V_{max} were derived from analysis of duplicate hydrolysis curves acquired from two independent enzyme preparations for each variant.

The thermostability of each variant was described by the temperature-dependent, irreversible depletion of soluble enzyme as observed by FSEC (33, 52). Purified Q14R or D38V (0.1 mg/mL) was transferred to standard PCR tubes and incubated in a thermocycler at defined temperatures for 10 min intervals. Precipitated material was removed by ultracentrifugation at 105,000 rcf and the supernatant injected on a Superose6 Increase 10/300 GL column equilibrated in 50 mM Tris pH 7.5, 150 mM NaCl, 0.2 mM LMNG buffer. Enzyme elution was monitored by fluorescence (Ex 280 nm, Em 320 nm). Peak area, representing the remaining soluble population, was plotted as a function of increasing temperature and the data fit with a dose-response curve to quantify the melting temperature (T_M).

Acknowledgments

This work was supported by the National Institutes of Health grants P41-GM104601 and R24-GM145965 (M.S. and E.T.) and R01-DK132259 (R.O'B.), the Vanderbilt Molecular Endocrinology Training program 5T32-DK07563 (E.M.H.), the American Heart Association 23PRE1017904 (E.M.H.), the Juvenile Diabetes Research Foundation 1-INO-2022-1121-A-N and Vanderbilt University's Program in the Molecular Basis of Genetic Diseases. We also acknowledge computing resources provided by eXtreme Science and Engineering Discovery Environment (XSEDE) (grant MCA06N060 to E.T.). The authors wish to thank Mr. Michael Mohan for assistance with baculovirus production and protein expression and Dr. Chris Newgard for providing the 832/13 cell line. We thank Dr. Hassane Mchaourab and Dr. Benjamin Brown for a critical reading of the manuscript.

References

1. I. Gerin, M. Veiga-da-Cunha, Y. Achouri, J. F. Collet, E. Van Schaftingen, Sequence of a putative glucose 6-phosphate translocase, mutated in glycogen storage disease type Ib. *FEBS Lett* **419**, 235-238 (1997).
2. K. J. Lei, L. L. Shelly, C. J. Pan, J. B. Sidbury, J. Y. Chou, Mutations in the glucose-6-phosphatase gene that cause glycogen storage disease type 1a. *Science* **262**, 580-583 (1993).
3. E. van Schaftingen, I. Gerin, The glucose-6-phosphatase system. *Biochem J* **362**, 513-532 (2002).
4. J. C. Hutton, R. M. O'Brien, Glucose-6-phosphatase catalytic subunit gene family. *J Biol Chem* **284**, 29241-29245 (2009).
5. G. Mithieux, New knowledge regarding glucose-6 phosphatase gene and protein and their roles in the regulation of glucose metabolism. *Eur J Endocrinol* **136**, 137-145 (1997).
6. G. Mithieux *et al.*, Glucose-6-phosphatase mRNA and activity are increased to the same extent in kidney and liver of diabetic rats. *Diabetes* **45**, 891-896 (1996).
7. R. C. Nordlie, P. D. Boyer, "22 Glucose-6-Phosphatase, Hydrolytic and Synthetic Activities" in The Enzymes. (Academic Press, 1971), vol. 4, pp. 543-610.
8. N. Barzilai, L. Rossetti, Role of glucokinase and glucose-6-phosphatase in the acute and chronic regulation of hepatic glucose fluxes by insulin. *J Biol Chem* **268**, 25019-25025 (1993).

9. G. Boden, P. Cheung, T. P. Stein, K. Kresge, M. Mozzoli, FFA cause hepatic insulin resistance by inhibiting insulin suppression of glycogenolysis. *Am J Physiol Endocrinol Metab* **283**, E12-19 (2002).
10. A. Consoli, Role of liver in pathophysiology of NIDDM. *Diabetes Care* **15**, 430-441 (1992).
11. R. M. O'Brien, R. S. Streeper, J. E. Ayala, B. T. Stadelmaier, L. A. Hornbuckle, Insulin-regulated gene expression. *Biochem Soc Trans* **29**, 552-558 (2001).
12. L. Rossetti, A. Giaccari, R. A. DeFronzo, Glucose toxicity. *Diabetes Care* **13**, 610-630 (1990).
13. K. Sharabi, C. D. Tavares, A. K. Rines, P. Puigserver, Molecular pathophysiology of hepatic glucose production. *Mol Aspects Med* **46**, 21-33 (2015).
14. K. Y. Trinh, R. M. O'Doherty, P. Anderson, A. J. Lange, C. B. Newgard, Perturbation of fuel homeostasis caused by overexpression of the glucose-6-phosphatase catalytic subunit in liver of normal rats. *J Biol Chem* **273**, 31615-31620 (1998).
15. G. van de Werve *et al.*, New lessons in the regulation of glucose metabolism taught by the glucose 6-phosphatase system. *Eur J Biochem* **267**, 1533-1549 (2000).
16. D. K. Granner, R. M. O'Brien, Molecular physiology and genetics of NIDDM. Importance of metabolic staging. *Diabetes Care* **15**, 369-395 (1992).
17. M. C. Petersen, D. F. Vatner, G. I. Shulman, Regulation of hepatic glucose metabolism in health and disease. *Nat Rev Endocrinol* **13**, 572-587 (2017).
18. J. Y. Chou, B. C. Mansfield, Mutations in the glucose-6-phosphatase-alpha (G6PC) gene that cause type Ia glycogen storage disease. *Hum Mutat* **29**, 921-930 (2008).
19. J. J. Shieh *et al.*, The molecular basis of glycogen storage disease type 1a: structure and function analysis of mutations in glucose-6-phosphatase. *J Biol Chem* **277**, 5047-5053 (2002).
20. J. Y. Chou, H. S. Jun, B. C. Mansfield, Type I glycogen storage diseases: disorders of the glucose-6-phosphatase/glucose-6-phosphate transporter complexes. *J Inherit Metab Dis* **38**, 511-519 (2015).
21. J. Y. Chou, H. S. Jun, B. C. Mansfield, Glycogen storage disease type I and G6Pase-beta deficiency: etiology and therapy. *Nat Rev Endocrinol* **6**, 676-688 (2010).
22. P. D. Stenson *et al.*, The Human Gene Mutation Database (HGMD((R))): optimizing its use in a clinical diagnostic or research setting. *Hum Genet* **139**, 1197-1207 (2020).
23. G. F. Bickerstaff, B. Burchell, Studies on the purification of glucose 6-phosphatase from rabbit liver microsomal fraction [proceedings]. *Biochem Soc Trans* **8**, 389-390 (1980).
24. A. Burchell, B. Burchell, Identification and purification of a liver microsomal glucose 6-phosphatase. *Biochem J* **205**, 567-573 (1982).
25. M. Speth, H. U. Schulze, Is thermostability of glucose-6-phosphatase indeed dependent on a stabilizing protein? *FEBS Lett* **202**, 32-36 (1986).
26. M. Speth, H. U. Schulze, The purification of a detergent-soluble glucose-6-phosphatase from rat liver. *Eur J Biochem* **208**, 643-650 (1992).
27. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
28. K. Tunyasuvunakool *et al.*, Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590-596 (2021).
29. Y. J. Sigal, M. I. McDermott, A. J. Morris, Integral membrane lipid phosphatases/phosphotransferases: common structure and diverse functions. *Biochem J* **387**, 281-293 (2005).
30. A. Krogh, B. Larsson, G. von Heijne, E. L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**, 567-580 (2001).
31. C. J. Pan, K. J. Lei, B. Annabi, W. Hemrika, J. Y. Chou, Transmembrane topology of glucose-6-phosphatase. *J Biol Chem* **273**, 6144-6148 (1998).
32. J. Stukey, G. M. Carman, Identification of a novel phosphatase sequence motif. *Protein Sci* **6**, 469-472 (1997).
33. D. P. Claxton, E. M. Overway, J. K. Oeser, R. M. O'Brien, H. S. McHaourab, Biophysical and functional properties of purified glucose-6-phosphatase catalytic subunit 1. *J Biol Chem* **298**, 101520 (2021).
34. K. J. Lei *et al.*, Glucose-6-phosphatase dependent substrate transport in the glycogen storage disease type-1a mouse. *Nat Genet* **13**, 203-209 (1996).
35. S. V. Salganik *et al.*, A detailed characterization of the adult mouse model of glycogen storage disease Ia. *Lab Invest* **89**, 1032-1042 (2009).
36. C. Lee, B. H. Su, Y. J. Tseng, Comparative studies of AlphaFold, RoseTTAFold and Modeller: a case study involving the use of G-protein-coupled receptors. *Brief Bioinform* **23** (2022).

37. A. Messerschmidt, L. Prade, R. Wever, Implications for the catalytic mechanism of the vanadium-containing enzyme chloroperoxidase from the fungus Curvularia inaequalis by X-ray structures of the native and peroxide form. *Biol Chem* **378**, 309-315 (1997).
38. M. E. Ghachi *et al.*, Crystal structure and biochemical characterization of the transmembrane PAP2 type phosphatidylglycerol phosphate phosphatase from Bacillus subtilis. *Cell Mol Life Sci* **74**, 2319-2332 (2017).
39. C. Y. Huang *et al.*, In situ serial crystallography for rapid de novo membrane protein structure determination. *Commun Biol* **1**, 124 (2018).
40. J. Zhao *et al.*, The Lipid A 1-Phosphatase, LpxE, Functionally Connects Multiple Layers of Bacterial Envelope Biogenesis. *mBio* **10** (2019).
41. R. F. Alford, P. J. Fleming, K. G. Fleming, J. J. Gray, Protein Structure Prediction and Design in a Biologically Realistic Implicit Membrane. *Biophys J* **118**, 2042-2055 (2020).
42. L. Gerasimavicius, X. Liu, J. A. Marsh, Identification of pathogenic missense mutations using protein stability predictors. *Sci Rep* **10**, 15387 (2020).
43. E. H. Kellogg, A. Leaver-Fay, D. Baker, Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* **79**, 830-838 (2011).
44. H. Park *et al.*, Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *J Chem Theory Comput* **12**, 6201-6212 (2016).
45. K. J. Karczewski *et al.*, The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).
46. J. C. Phillips *et al.*, Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802 (2005).
47. J. C. Phillips *et al.*, Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J Chem Phys* **153**, 044130 (2020).
48. A. Ghosh, J. J. Shieh, C. J. Pan, M. S. Sun, J. Y. Chou, The catalytic center of glucose-6-phosphatase. HIS176 is the nucleophile forming the phosphohistidine-enzyme intermediate during catalysis. *J Biol Chem* **277**, 32837-32842 (2002).
49. T. Kawate, E. Gouaux, Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673-681 (2006).
50. K. A. Boortz *et al.*, Functional Analysis of Mouse G6pc1 Mutations Using a Novel In Situ Assay for Glucose-6-Phosphatase Activity and the Effect of Mutations in Conserved Human G6PC1/G6PC2 Amino Acids on G6PC2 Protein Expression. *PLoS One* **11**, e0162439 (2016).
51. E. M. Overway *et al.*, Nonsynonymous single-nucleotide polymorphisms in the G6PC2 gene affect protein expression, enzyme activity, and fasting blood glucose. *J Biol Chem* **298**, 101534 (2022).
52. M. Hattori, R. E. Hibbs, E. Gouaux, A fluorescence-detection size-exclusion chromatography-based thermostability assay for membrane protein precrystallization screening. *Structure* **20**, 1293-1299 (2012).
53. R. Bhattacharya, P. W. Rose, S. K. Burley, A. Prlic, Impact of genetic variation on three dimensional structure and function of proteins. *PLoS One* **12**, e0171355 (2017).
54. M. Gao, H. Zhou, J. Skolnick, Insights into Disease-Associated Mutations in the Human Proteome through Protein Structural Analysis. *Structure* **23**, 1362-1369 (2015).
55. P. Yue, Z. Li, J. Moult, Loss of protein structure stability as a major causative factor in monogenic disease. *J Mol Biol* **353**, 459-473 (2005).
56. S. Ittisoponpisan *et al.*, Can Predicted Protein 3D Structures Provide Reliable Insights into whether Missense Variants Are Disease Associated? *J Mol Biol* **431**, 2197-2212 (2019).
57. A. Valanciute *et al.*, Accurate protein stability predictions from homology models. *Comput Struct Biotechnol J* **21**, 66-73 (2023).
58. M. Akdel *et al.*, A structural biology community assessment of AlphaFold2 applications. *Nat Struct Mol Biol* **29**, 1056-1067 (2022).
59. G. R. Buel, K. J. Walters, Can AlphaFold2 predict the impact of missense mutations on structure? *Nat Struct Mol Biol* **29**, 1-2 (2022).
60. J. M. McBride, K. Polev, V. Reinharz, B. A. Grzybowski, T. Tlusty, AlphaFold2 can predict single-mutation effects on structure and phenotype. *bioRxiv* (2023).
61. M. A. Pak *et al.*, Using AlphaFold to predict the impact of single mutations on protein stability and function. *bioRxiv* (2021).

62. D. Del Alamo, D. Sala, H. S. McHaourab, J. Meiler, Sampling alternative conformational states of transporters and receptors with AlphaFold2. *Elife* **11** (2022).
63. H. K. Wayment-Steele, S. Ovchinnikov, L. Colwell, D. Kern, Prediction of multiple conformational states by combining sequence clustering with AlphaFold2. *bioRxiv* (2022).
64. R. A. Stein, H. S. McHaourab, SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2. *PLoS Comput Biol* **18**, e1010483 (2022).
65. E. Furuya, K. Hotta, K. Tagawa, Anomer specificity of glucose-6-phosphatase and glucokinase. *Biochem Biophys Res Commun* **141**, 931-936 (1986).
66. E. F. Pettersen *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).
67. K. D. Tsirigos, C. Peters, N. Shu, L. Kall, A. Elofsson, The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res* **43**, W401-407 (2015).
68. Schrodinger, LLC (2015) The PyMOL Molecular Graphics System, Version 1.8.
69. A. Stein, D. M. Fowler, R. Hartmann-Petersen, K. Lindorff-Larsen, Biophysical and Mechanistic Models for Disease-Causing Protein Variants. *Trends Biochem Sci* **44**, 575-588 (2019).
70. J. Schindelin *et al.*, Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682 (2012).
71. M. J. Landrum *et al.*, ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* **46**, D1062-D1067 (2018).

Table 1. Rosetta ΔΔG of known G6PC1 variants
Nine variants are found within the active site. Q14R and E319K are variants found at the protein surface and distal from the active site.

Variant	Clinical Significance*	ΔΔG (kcal/mol)[#]
Q14R	Unknown	+1.0
D38V	Pathogenic	+3.7
K76N	Pathogenic	+2.8
R83H	Pathogenic	+11.7
E110K	Likely Pathogenic	+7.0
P113L	Pathogenic	+161.0
G118S	Pathogenic	-0.7
H119L	Pathogenic	-4.0
R170Q	Pathogenic	-5.5
T255I	Pathogenic	+71.5
E319K	Unknown	+0.8

*ClinVar (71), gnomAD (45) and HGMD (22) genomic databases
[#]Calculated with the bilayer-relaxed Apo model

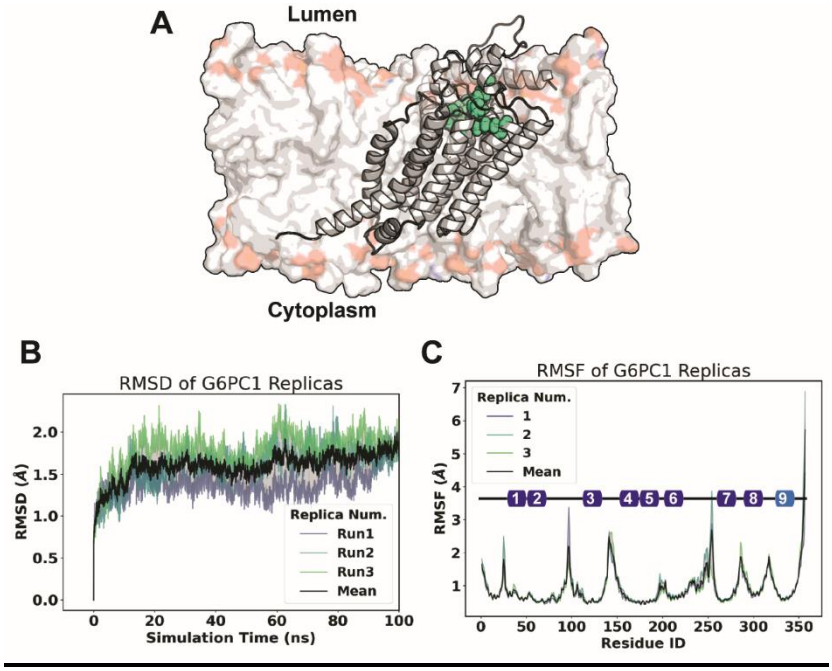

Figure 1. Modeling the AF2 G6PC1 structure in a simulated bilayer. (A) Cartoon rendering of G6PC1 immersed into a representative ER bilayer. The composition is reported in Methods. The active site is demarcated by residues contributing to the phosphatidic acid phosphatase sequence motif (green). (B) Time series of RMSD of G6PC1 over the course of equilibrium simulations. (C) RMSF showcases regions of higher disorder that map to termini and loops. The colored boxes illustrate the topology of the nine transmembrane helices along the primary sequence. The color represents the pLDDT score of the AF2 model for these segments, with the darker blue indicating higher pLDDT that reflects higher accuracy of the model.

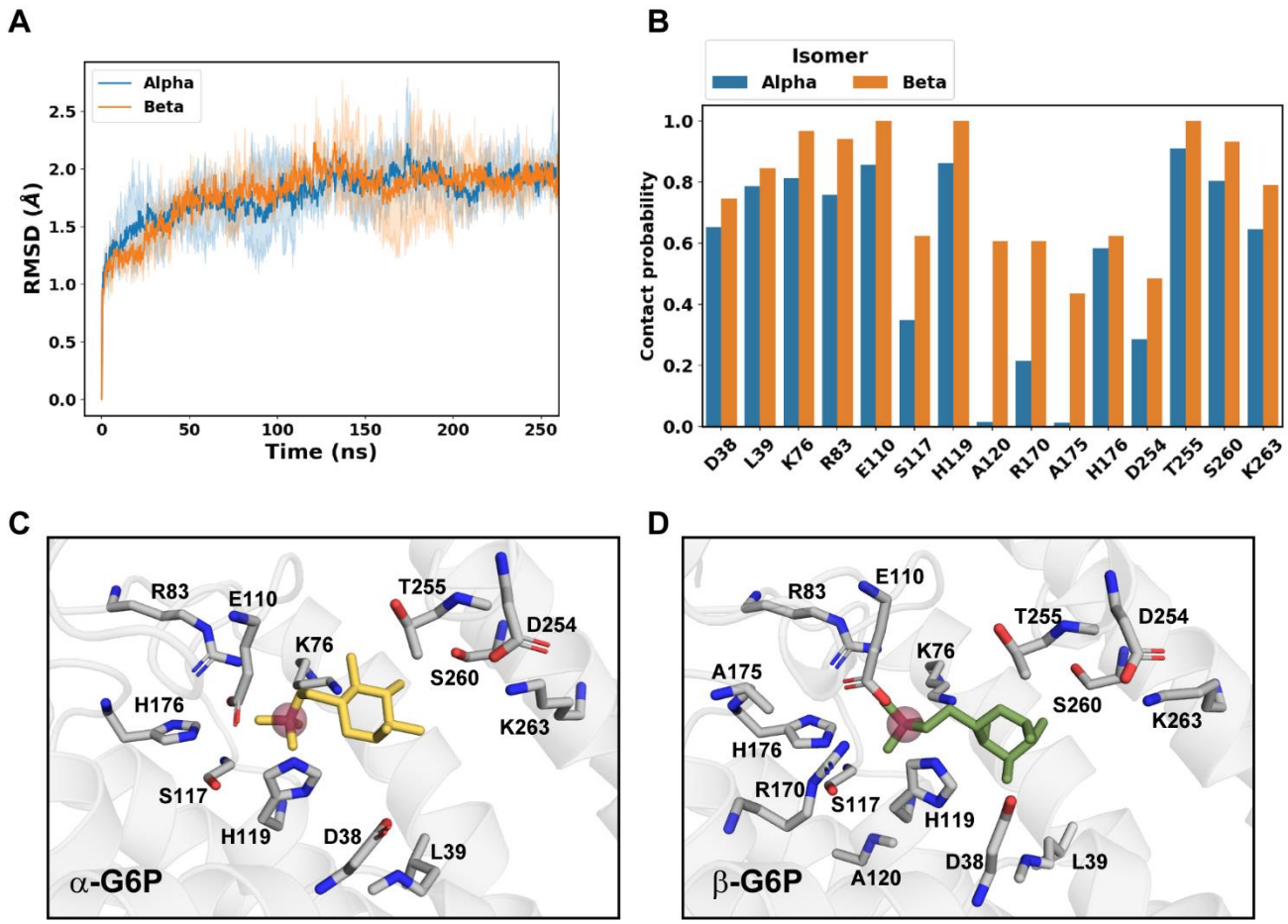

Figure 2. Identification of G6P binding residues and stability of the ligand-bound model. (A) RMSD of substrate-bound simulations for each anomer. Structural convergence is observed over 250ns of simulation. The dark traces are the average RMSD over multiple runs (n=3). (B) Per residue contact probability, defined as the proportion of time each residue spends within 4 Å of G6P, demonstrates some anomer dependence. Only residues for which either anomer has a probability greater than 0.5 are shown. (C-D) Representative binding modes of α-G6P (C) or β-G6P (D) to G6PC1. The pose chosen was based on the shortest contact distance to side chains of the phosphatase sequence motif.

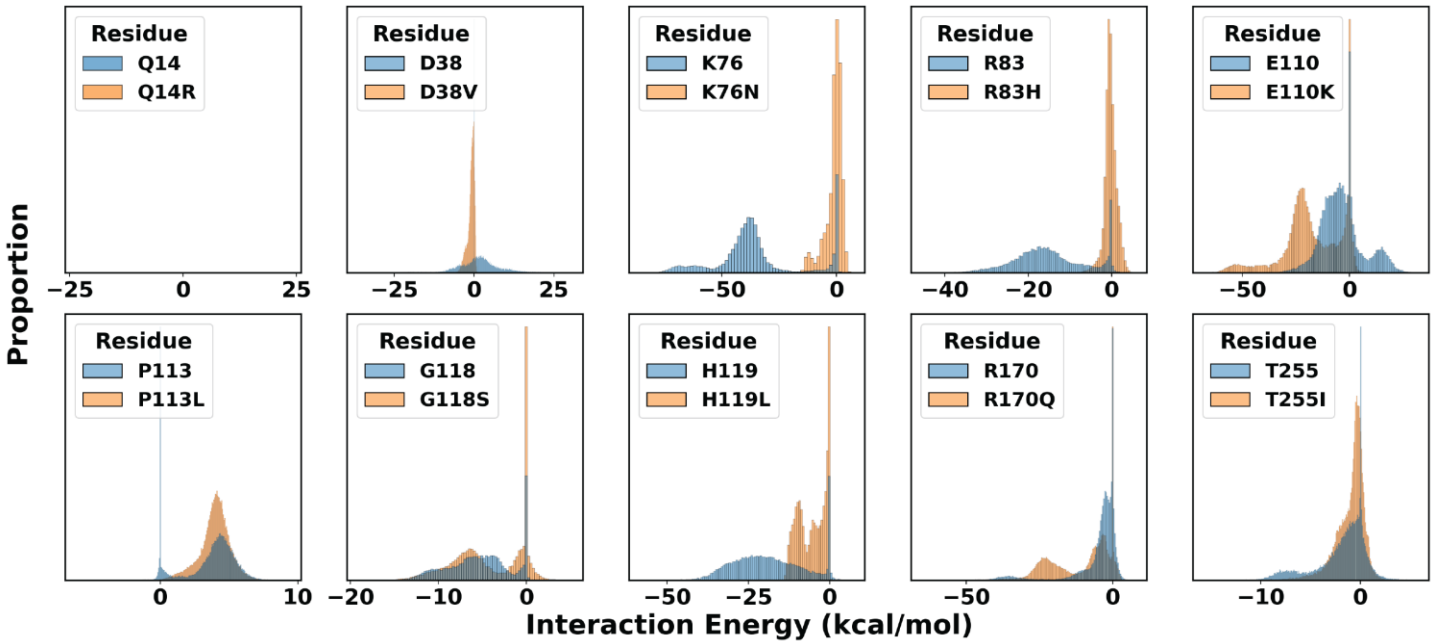

Figure 3. Perturbations of non-bonded interaction energies with G6P in GSD type1a variants. Interaction energy of each residue with G6P illustrates changes in the enthalpy of binding. Data are pooled from both G6P anomers. Energies shown are the summation of van der Waals and electrostatics. Q14R, a negative control, sits outside the active site and does not interact with G6P.

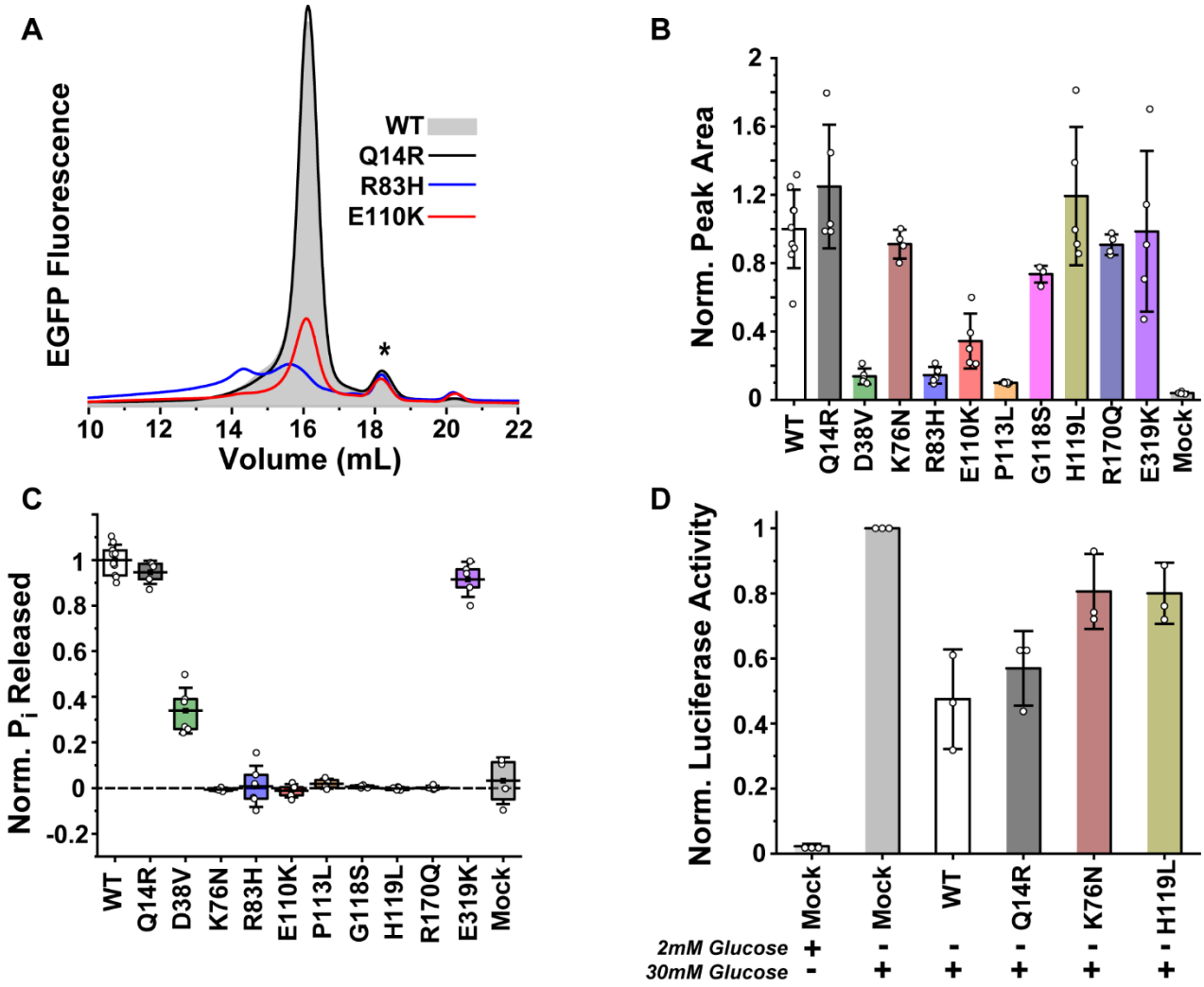

Figure 4. Expression and activity profiles of GSD type1a variants from transfection of adherent cells. (A) Representative FSEC traces illustrate the range of chromatographic behaviors, including reduced expression (E110K) or the presence of misfolded species (R83H). The (*) identifies a minor population of cleaved EGFP. (B) Expression levels, quantified by the area under the curve in FSEC traces, is dependent on the variant. The bar graph captures the average ± standard deviation of multiple independent experiments. (C) The corresponding variant catalytic capacity normalized to its enzyme concentration obtained from (B) is plotted relative to the WT. (D) *In situ* assay of G6PC1 variant activity stimulated by glucose in adherent 832/13 rat islet-derived cells confirms compromised activity of the K76N and H119L variants relative to functional constructs. The metrics in B-D were analyzed by one-way ANOVA and Tukey tests (SI Appendix, Table S1).

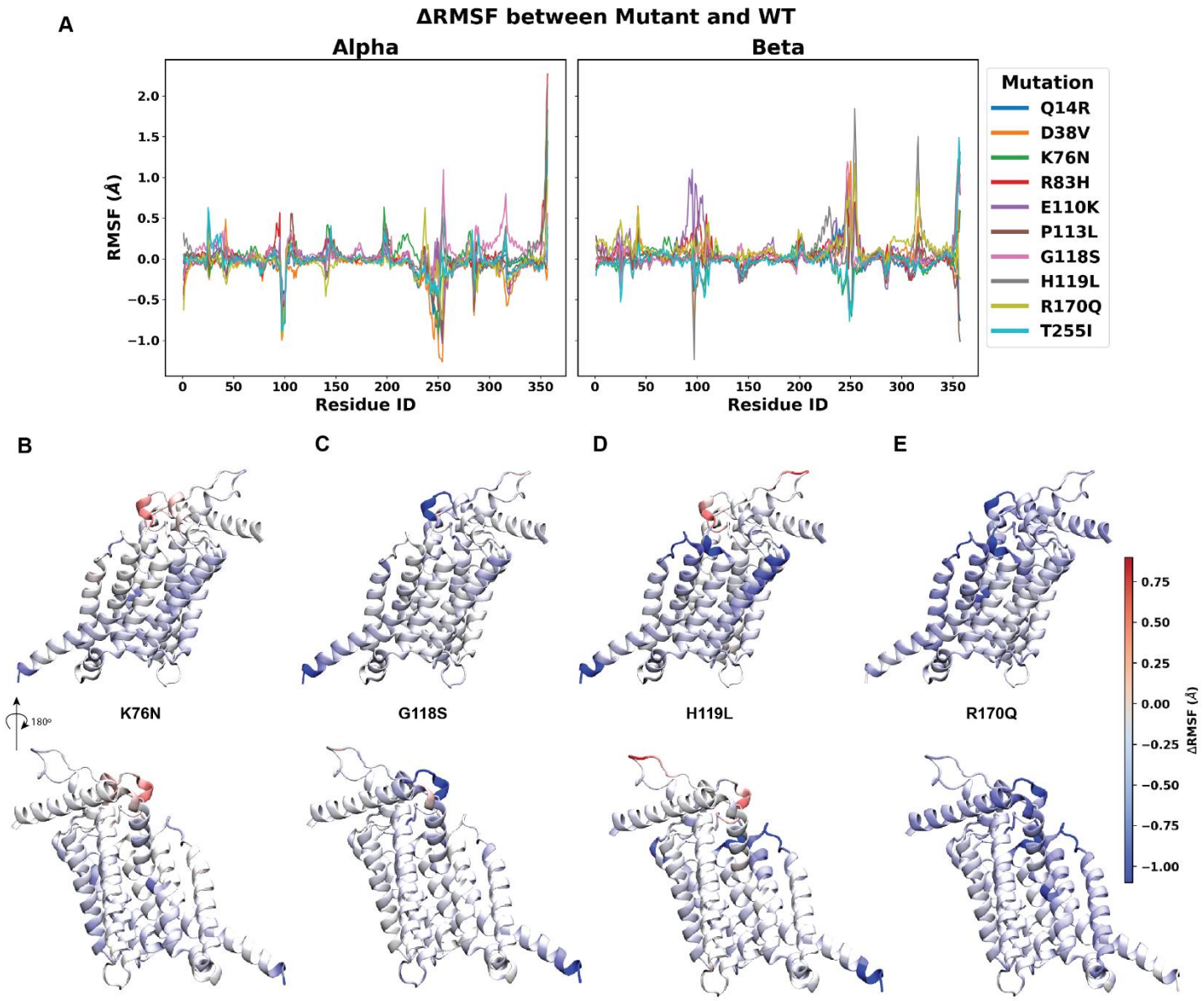

Figure 5. Pattern of altered backbone fluctuations in GSD type1a variants. (A) The per residue ΔRMSF of the protein backbone as a result of mutation indicates regions of changed flexibility relative to WT. ΔRMSF was computed by subtracting RMSF of the WT over the simulation time course from the individual variants and parsed according to G6P anomer. Regions of positive ΔRMSF indicate an increase in flexibility while regions of negative ΔRMSF indicate a decrease in flexibility. (B-E) ΔRMSF values acquired in the presence of β-G6P were mapped onto the bilayer-relaxed structural model, highlighting the key regions of differential flexibility in these variants.

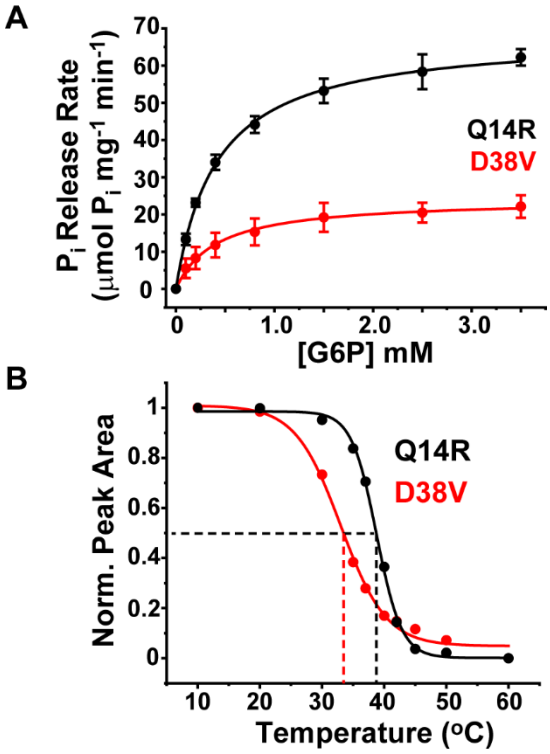

Figure 6. Saturation kinetics of D38V is correlated with reduced thermostability. (A) Michaelis-Menten kinetics of G6P hydrolysis indicates that the specific activity of GSD type1a variant D38V (V_{max} = 24.5 ± 1.9 μmol mg^{-1} min^{-1}; K_m = 0.472 ± 0.246 mM) is reduced threefold relative to the WT-like Q14R variant (V_{max} = 68.4 ± 3.0 μmol mg^{-1} min^{-1}; K_m = 0.413 ± 0.017 mM). The data points, which represent the average Pi released at the indicated G6P concentrations from four measurements, were fit (solid lines) with a single binding site model weighted by the standard deviation. (B). The left shift in the FSEC-based melting curve indicates reduced thermostability of D38V relative to Q14R.

SI Appendix

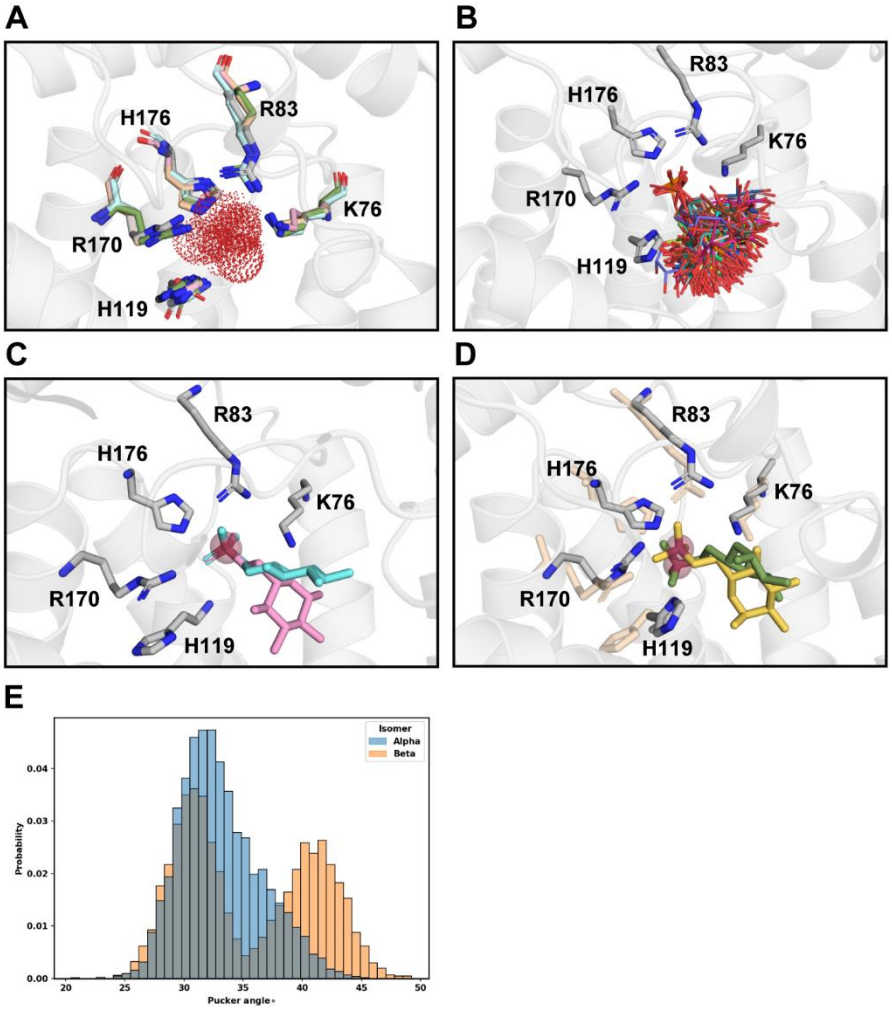

Figure S1. Docking of G6P anomers to the predicted G6PC1 active site. (A) Superposition of the active site from PAP2 crystal structures (PDB ID 6EBU, 6FMX, 5JKI and 1IDQ) with the AF2 G6PC1 model (Uniprot P35576) indicating the conserved side chain projections and atomic positions for surrogate phosphate ions. (B) Preliminary docking of 95 G6P structures from both anomers to the consensus position of the quaternary ions in (A). (C) Representative α- (pink) and β-G6P (cyan) models docked to the bilayer-relaxed G6PC1 model. (D) Comparison of G6P anomer positioning and side chain rotamers at t = 0 of the MD simulations. Here, the α- and β-G6P models are yellow and green, respectively. The position of rotamers in the Apo model (tan) are as shown in (C) prior to initiating the equilibrium simulations. (E) Distribution of sugar pucker angles over the course of simulation in WT protein. The bimodal nature of β-G6P correlates with ring flipping observed in simulation.

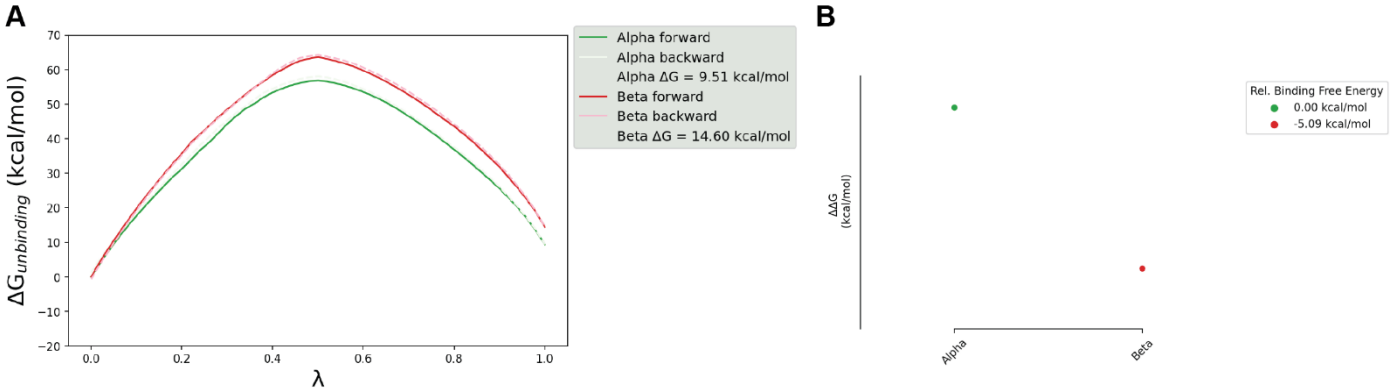

Figure S2. Free energy perturbation calculations of G6P anomer binding. (A) Forward and backward paths for binding of each anomer during a 50-window FEP simulation set. Convergence is observed for both anomers as well as a clean overlay of the forward and backward paths. (B) Relative binding free energy of the anomers. G6PC1 has a -5.09 kcal/mol preference for β-G6P over the α isomer.

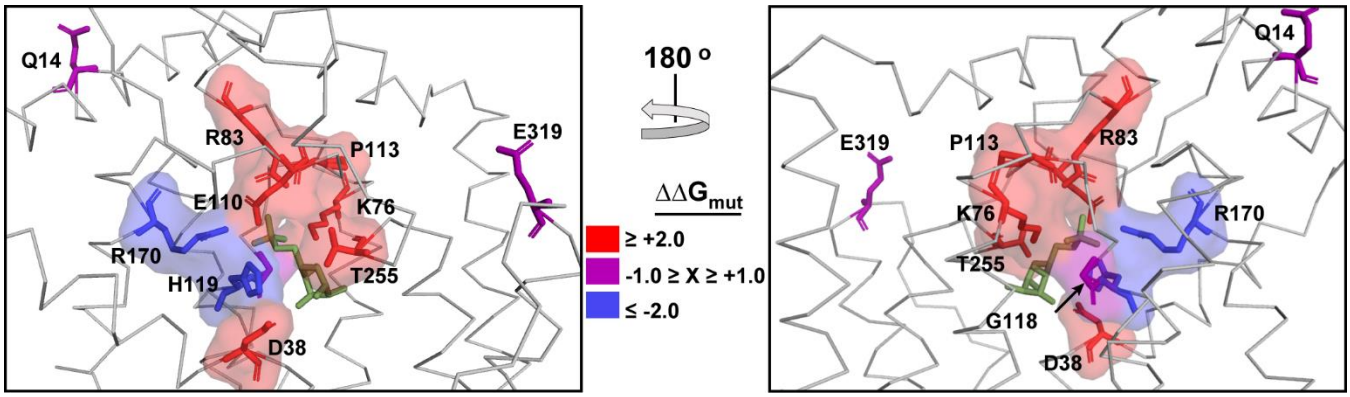

Figure S3. Mapping of Rosetta ΔΔG onto the AF2 G6PC1 model. Relative location of amino acids associated with GSD type 1a in the active site. Variants predicted to be destabilizing are colored red whereas stabilizing variants are colored blue (Table 1). The position of two variants of unknown significance that reside outside the active site, Q14 and E319, are shown as stick representations.

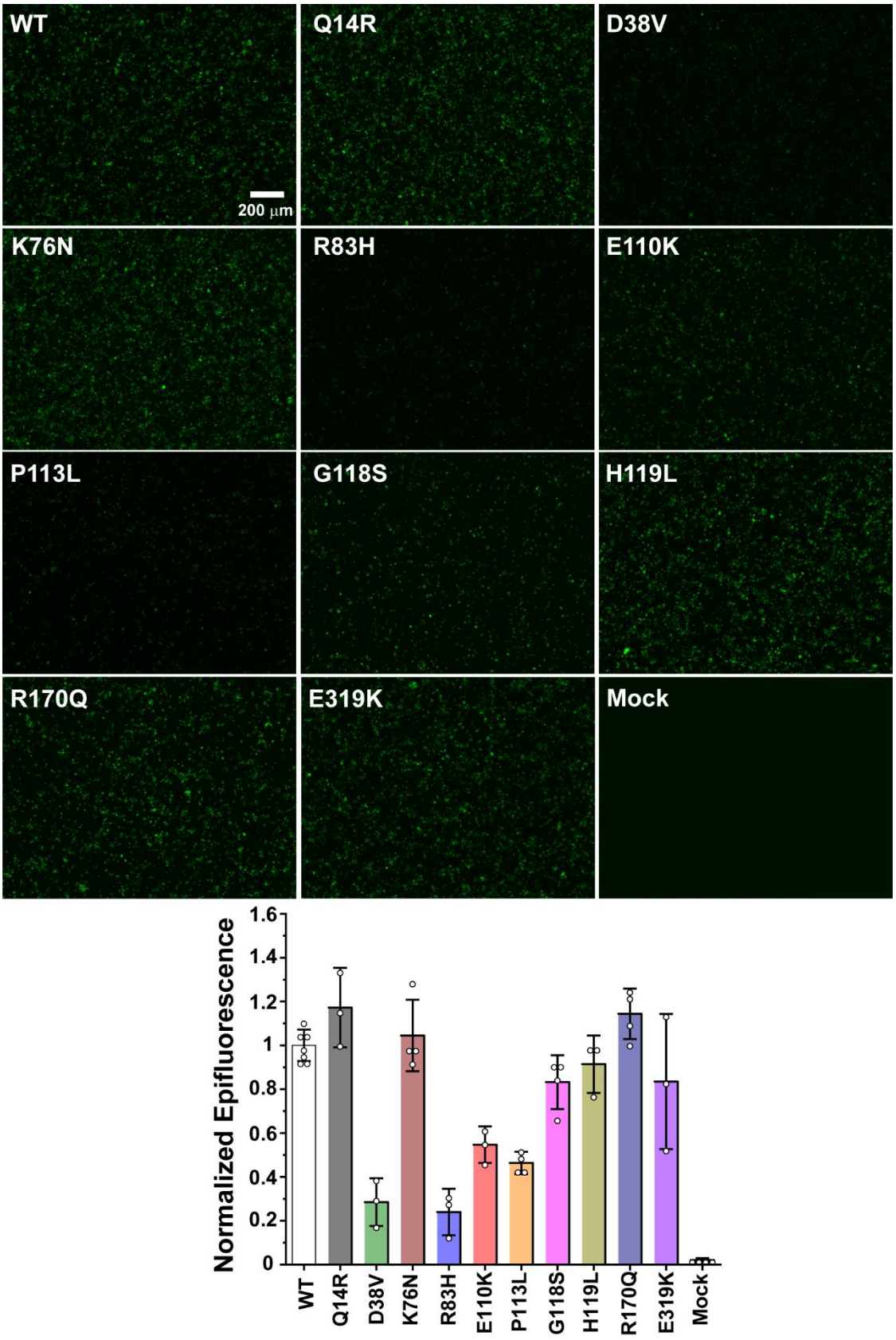

Figure S4. Epifluorescence of adherent HEK293SG cells following transfection of WT and variant G6PC1. Representative images are shown at the indicated scale and the fluorescence quantified in ImageJ software. The pattern of expression is comparable to that observed from FSEC analysis of detergent solubilized enzyme. The quantified fluorescence was analyzed by one-way ANOVA and Tukey tests (SI Appendix, Table S1).

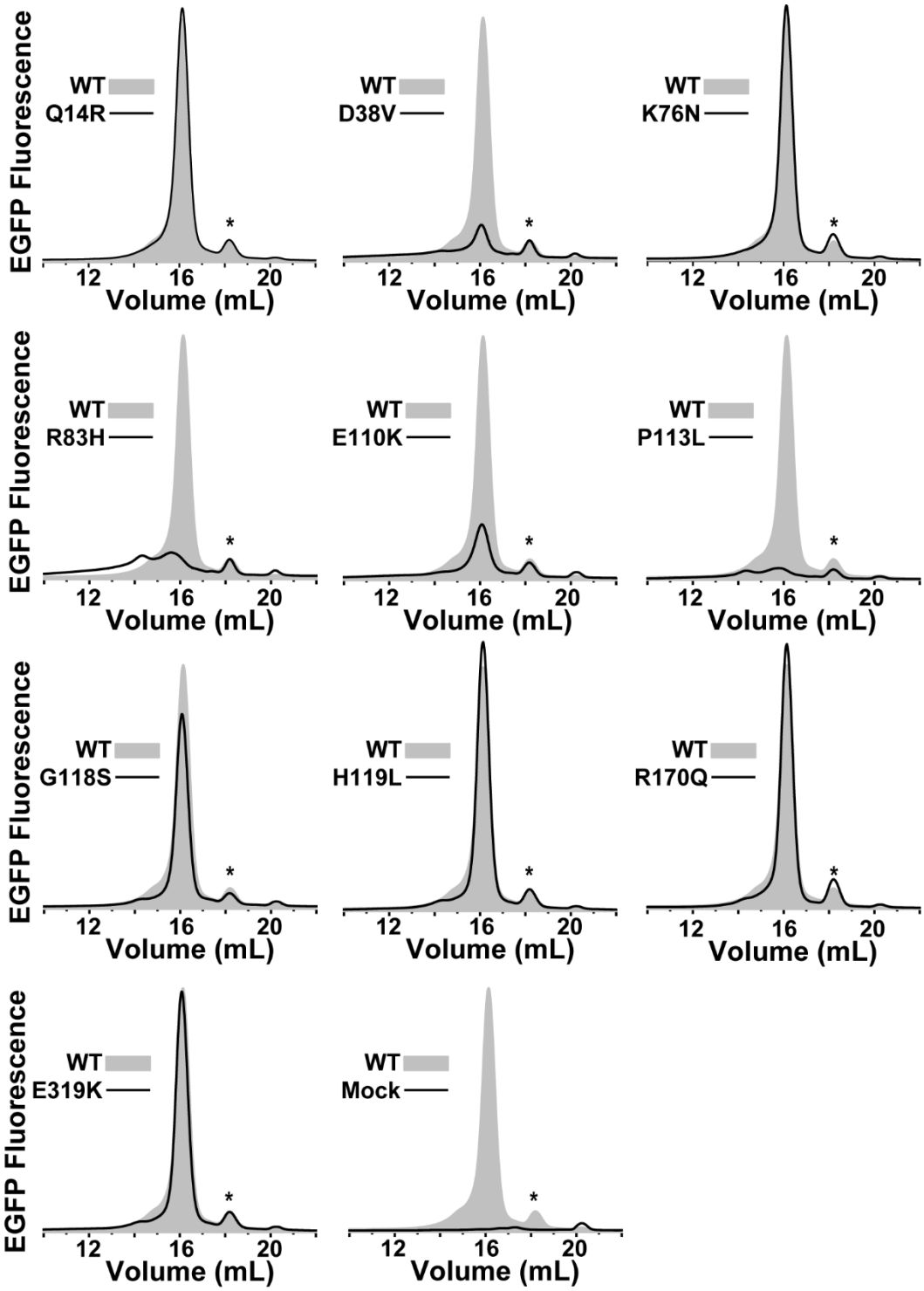

Figure S5. FSEC traces of GSD type1a variants relative to WT. Representative chromatographic profiles of detergent solubilized G6PC1 illustrate the impact of disease linked variants to enzyme expression and folding. The (*) identifies a minor population of cleaved EGFP in the samples.

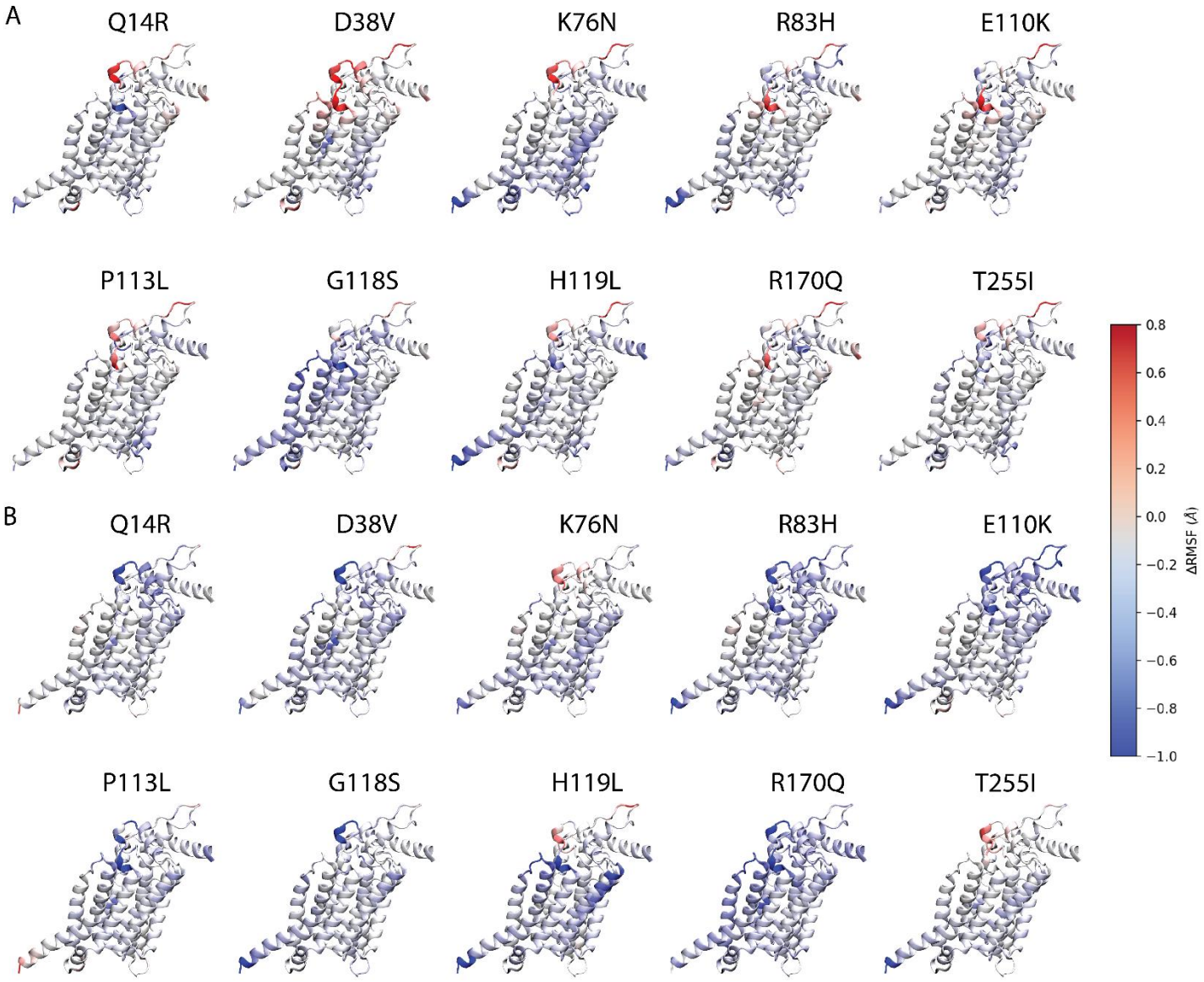

Figure S6. ΔRMSF of GSD type1a variants. Mapping of changes in backbone flexibility induced by each variant and parsed according to G6P anomer. (A) α-G6P and (B) β-G6P. Shades of blue indicate reduced dynamics over the course of the MD simulations.

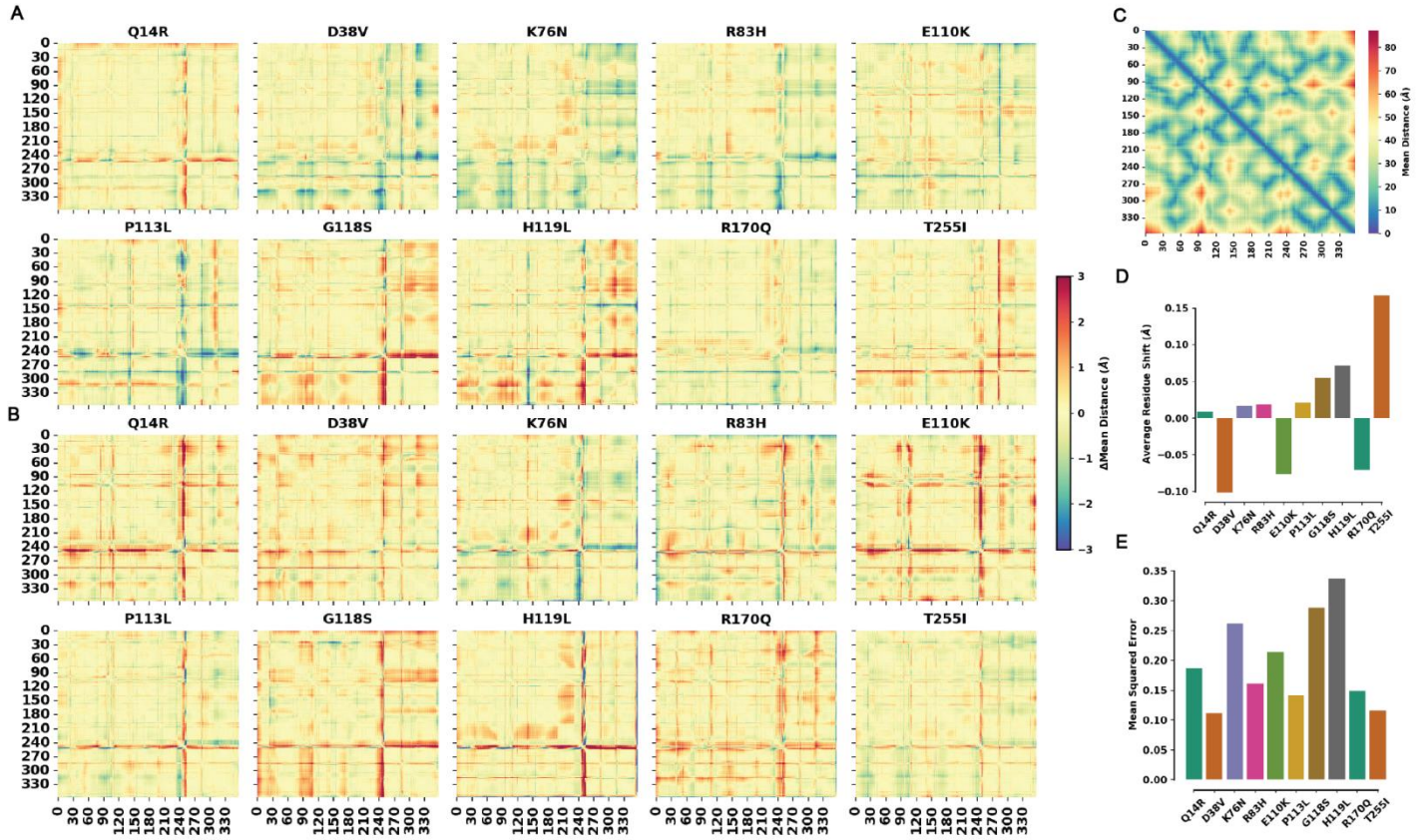

Figure S7. Identification of structural distortions induced by GSD type 1a variants. Δ mean pairwise residue-residue distance matrices for each mutant with bound (A) α-G6P and (B) β-G6P. Matrices obtained by subtracting the mean pairwise residue-residue distance matrix in WT from values for each mutant. Positive values indicate a given pair of residues are farther apart in the variant, while negative values indicate they get closer together. (C) Mean pairwise residue-residue distance matrix for WT G6PC1. (D) Mean pairwise distance changes over all residues for a given mutant and the associated (E) mean squared error (MSE) identify changes mainly due to backbone dynamics (higher MSE, lower mean distance) or to induced structural distortions (lower MSE, higher mean distance).