Supporting Information

N-Glycosylation and Gaucher Disease Mutation Allosterically Alter Active Site Dynamics of Acid-β-Glucosidase

Michael Gregory Souffrant[1], Xin-Qiu Yao[1], Mohamed Momin[1], Donald Hamelberg*[1,2,]

[1]Department of Chemistry, [2]Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, Georgia 30302-3965, USA.

*Corresponding Author Tel.: (404) 413-5564; E-mail: dhamelberg@gsu.edu

KEYWORDS: molecular dynamics, allosteric regulation, principal component analysis, contact, community.

SI Text

Active site access is blocked due to the dynamics of loops going from WT to L444P and to a lesser extent from WT to N370S

Overall, L444P displays a more substantial deviation from WT than N370S. Subsequent evaluations of the dynamics of acid-β-glucosidase suggest that the active site is more dynamic and less ordered in L444P than in WT indicating denaturation or an enzyme less fit for catalysis. The L444P mutant variant has a distance distribution with maxima at similar positions to wild type (**Figure S7B**, *red*). However, upon the L444P mutation (**Figure S7E, H**), the enzyme experiences different wells of conformational states at broader distances between E235 and E340 relative to WT along PC1 and PC2. This suggests the sampling of more conformational states regarding the catalytic residues in L444P, which could structurally hinder GlcCerase from hydrolyzing GlcCer at an efficient rate or represent an early sign of denaturation. The distal mutations could alter the distance between the δ carbons of E340 and E235, sampling a much broader range of catalytically incompetent conformational states of the L444P mutant than WT.

Projections of the principal components of the PCA suggest loop D (residues 344 – 357) and loop B (residues 242 – 252), located in the TIM barrel catalytic domain (domain III) of GlcCerase, to form a gate at the entrance of the active site. Access to the active site is more restricted in L444P due to the drastic deviation of loop D as shown in **Figure S5B**. Thus, whether the L444P variant of GlcCerase reaches the peripheral membrane of the cellular lysosome without denaturing, access to the catalytic site by the substrate is occluded to a larger extent than N370S. This suggest a higher buildup of GlcCer as a severe expression of the Gaucher disease in L444P, even if the structure is stabilized. Furthermore, loop F (**Figure S5B, C**) in domain II could be

critical in initiating the communication flow to different regions of GlcCerase upon enzymatic mutations and ligand binding. Such findings suggest that information from the altered conformations of loop F is passed in a concerted fashion between amino acid residues to gating loops D and B.

Overall, residue-residue contact difference analysis going from WT to L444P shows a more drastic change in contacts formed and broken in comparison to that from WT to N370S, in which case, the amount of contacts formed in the active site with L444P is suggested to limit catalysis to a greater extent than N370S (**Figure S11B, D**). Comparing the difference community networks of both mutations (N370S and L444P) with respect to wild type GlcCerase (**Figure S11C, E**) suggests a concerted network between communities upon distal perturbations of the enzyme that is more drastic in WT to L444P as oppose to WT to N370S. Based on dCNA results, both L444P and N370S mutations are derived from the red community in the network modularity**.** Looking at such establishments, the distal influences of the regulatory domain (domain II) are more evident in L444P in comparison to N370S (**Figure S11C, E**). For instance, going from WT to L444P, a significant loss in interactions with a net dp_c of -10.3 between domain I (**Figure S11C**, *blue vertex*) and portion of the catalytic domain (**Figure S11C**, *grey vertex*) is shown, whereas slightly more formed contacts with a net dp_c of 0.3 are observed from WT to N370S. Changes between surrounding communities and the green community that makes up loop D (**Figure S11E**, *green*) are less drastic in N370S than those in L444P. In general, the comparison between WT and N370S reveals a distinctive community linkage pattern not seen from WT to L444P, suggesting the outcome of a different network of communication from the mild mutation causing a subtle change in catalysis and enzymatic integrity. The difference community networks of both mutations with

respect to wild type GlcCerase suggests a concerted network between communities upon distal perturbations of the enzyme (i.e., allosteric effect). Associations between communities from both mutations eventually lead towards a favorable non-covalent interaction between the green and orange communities, which hinders substrate entry to the active site and is more manifested in L444P with a net dp_c of 0.3 compared to N370S with a net dp_c of 0.2. Such a small difference in net dp_c is critical, because it represents an accumulation of links at the residue-level between these two communities that is more evident in L444P than in N370S. A more coarse-grained view of the allosteric communication network using the dCNA method shows that the more severe L444P mutant, if stable, is more disruptive than the milder N370S mutant and leads to an even more restricted access to the active site, resulting in a higher buildup of substrate and extreme manifestation of Gaucher disease.

In general, L444P shows a lower probability of open state relative to WT and N370S at small cut-off bottleneck radii, indicative of its tendency to sample narrower tunneling cavities. As the cut-off value increases, however, N370S and L444P exhibit similar open-state probabilities, with L444P having a slightly higher open-state probability at cut-offs of ~1.6 Å or larger. However, this small difference might be caused by the fluctuation in dynamics over time and might not indicate a significant separation of L444P and N370S in the large cut-off range (**Figure S13C**).

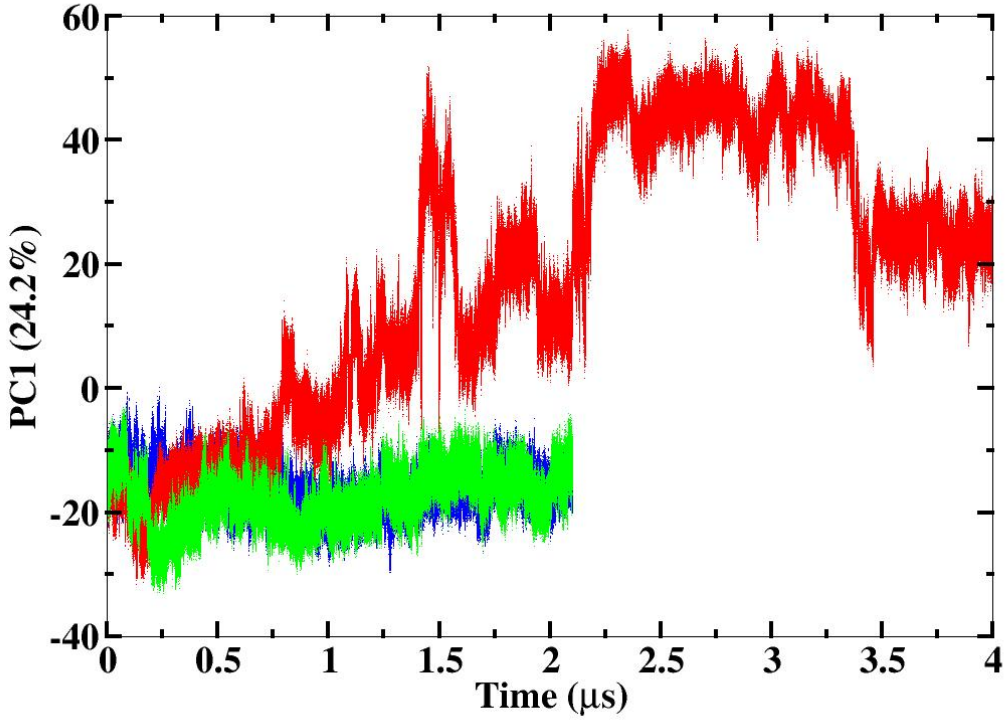

Figure S1. Conformational dynamics of WT (blue), N370S (green), and L444P (red) along PC1. The PCA is performed using entire trajectories of the three simulations.

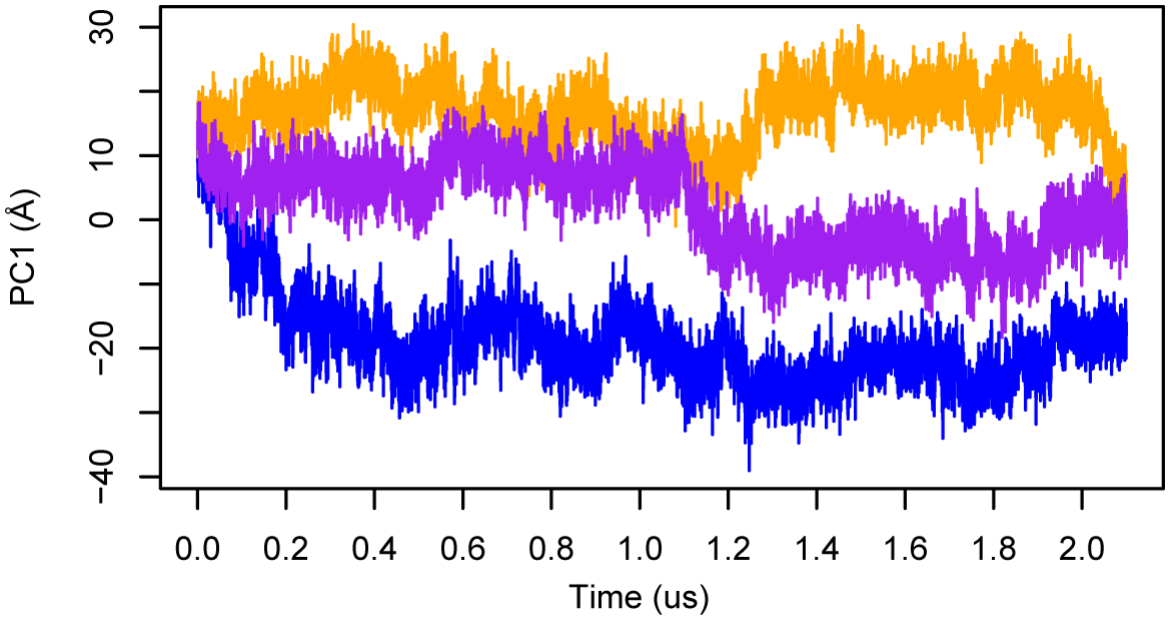

Figure S2. Conformational dynamics of WT (blue), NAG (orange), and 2NAGs (purple) along PC1. The PCA is performed using entire trajectories of the three simulations.

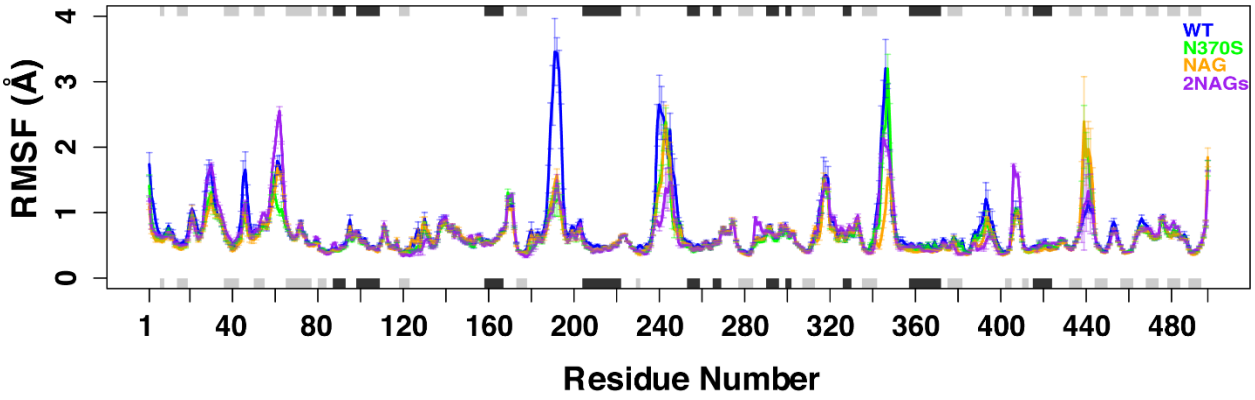

Figure S3. Error estimate of root mean square fluctuation (RMSF). Each trajectory is split into three chunks of equal lengths. Standard errors of RMSF across the chunks are calculated and displayed as error bars.

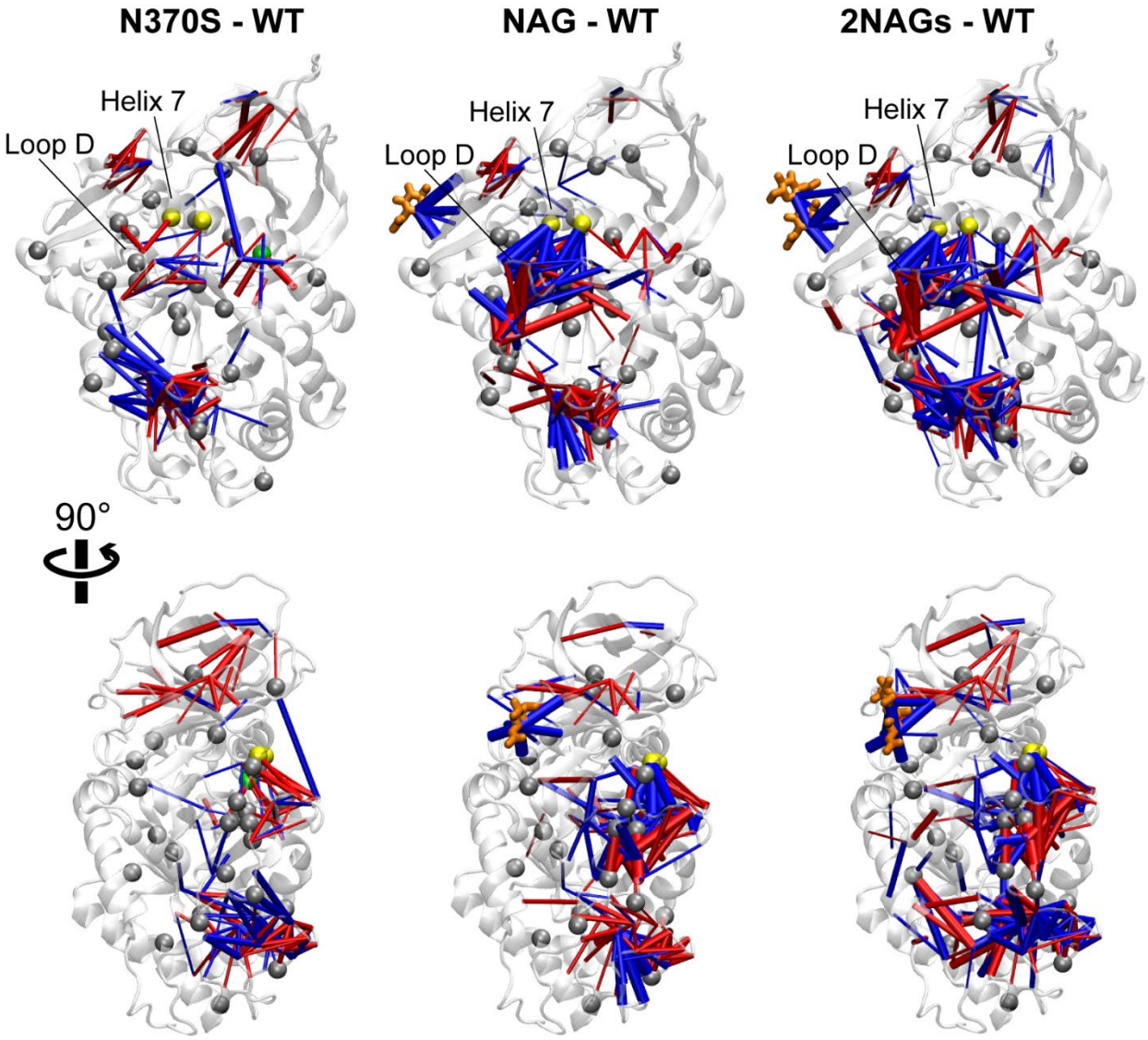

Figure S4. Inter-community residue-residue contact analysis. Probability differences (dp_c) of contacts connecting two different communities going from WT to N370S, NAG, and 2NAGs are shown. Blue (red) cylinders indicate contacts that are more often formed (broken). Cylinder radii are proportional to the absolute value of dp_c. Grey beads, mutation sites associated with severe Gaucher disease found in patients. Green bead, the N370S mutation site. Yellow beads, residues predicted to be critical for the allosteric communication between distal regions and the active site. NAGs are shown as sticks in orange.

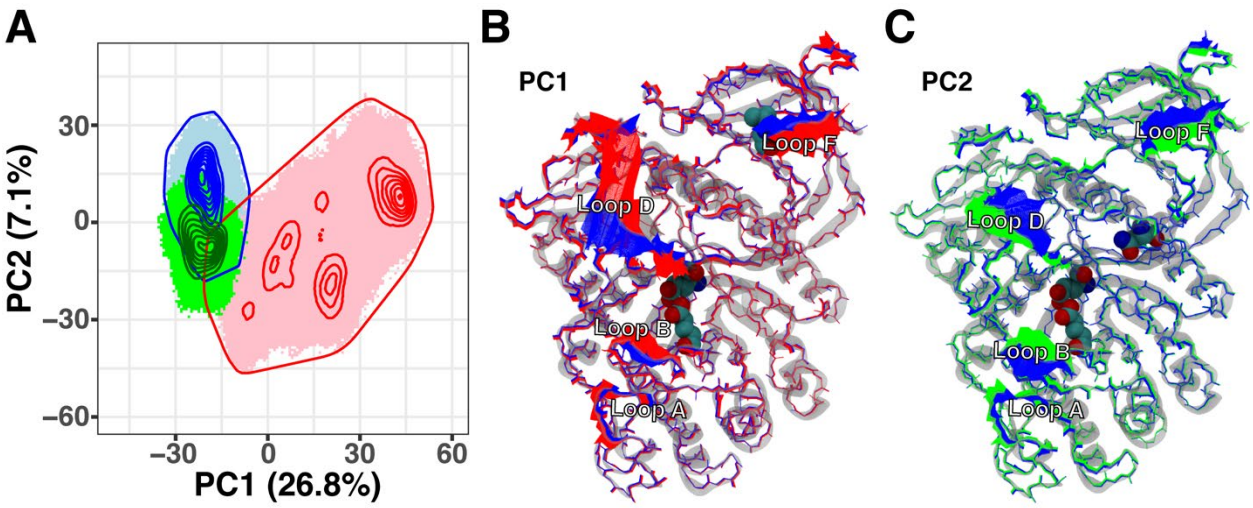

Figure S5. PCA results of backbone conformational dynamics between WT, L444P and N370S. (A) PCA plot of the top two projections of backbone coordinates of GlcCerase from the simulations of WT (blue), L444P (red), and N370S (green). The percentage of the total atomic displacements/variance described by each PC is indicated in the axis labels. The contour lines represent the probability density distributions of the conformational states sampled throughout the trajectories. The outlines of L444P (red) and N370S (green) are also shown to distinguish the subspace spanned by all three simulations. Collective motions represented by the first (B) and second (C) PC are displayed on top of the enzyme, where colors indicate ranges of PC projections of WT (blue), L444P (red), and N370S (green).

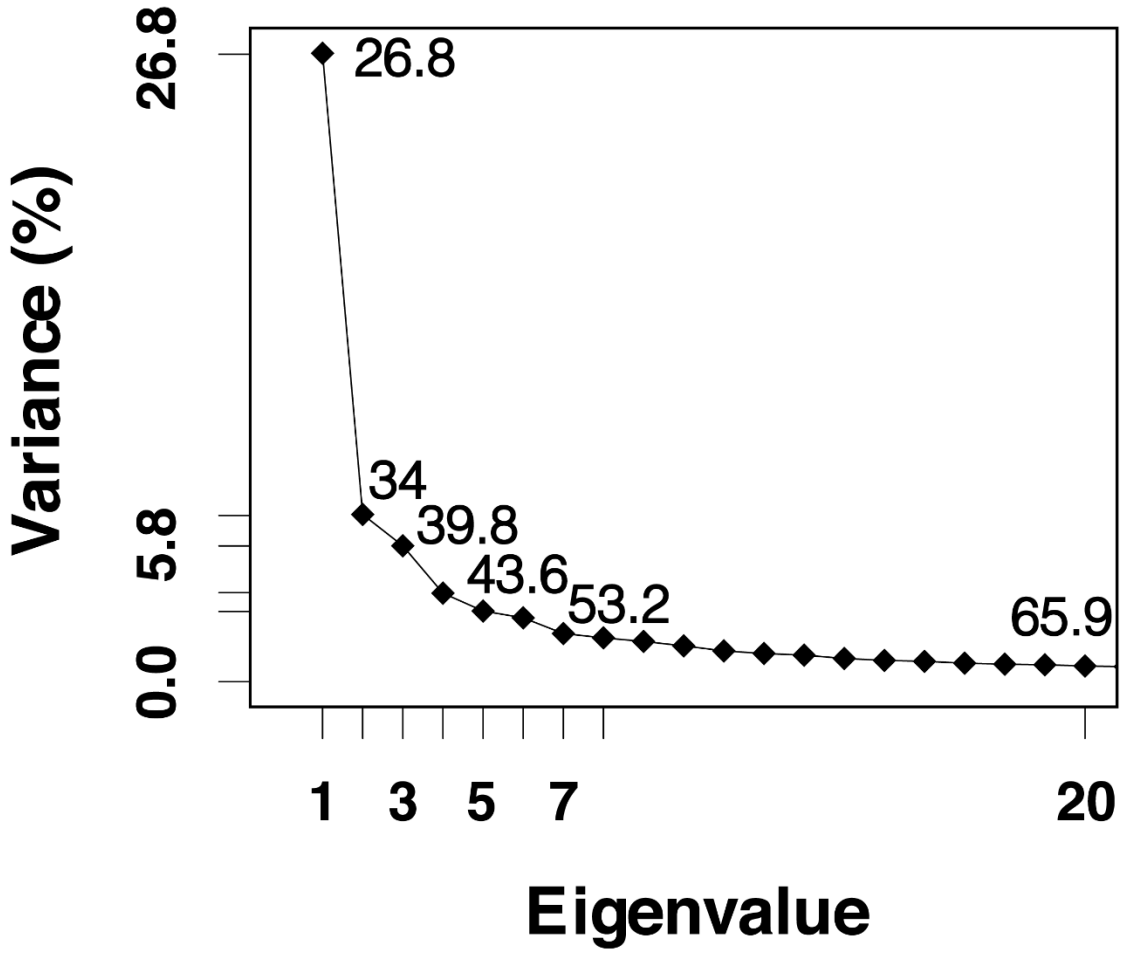

Figure S6. The scree plot for the PCA of backbone conformational dynamics between WT, L444P and N370S.

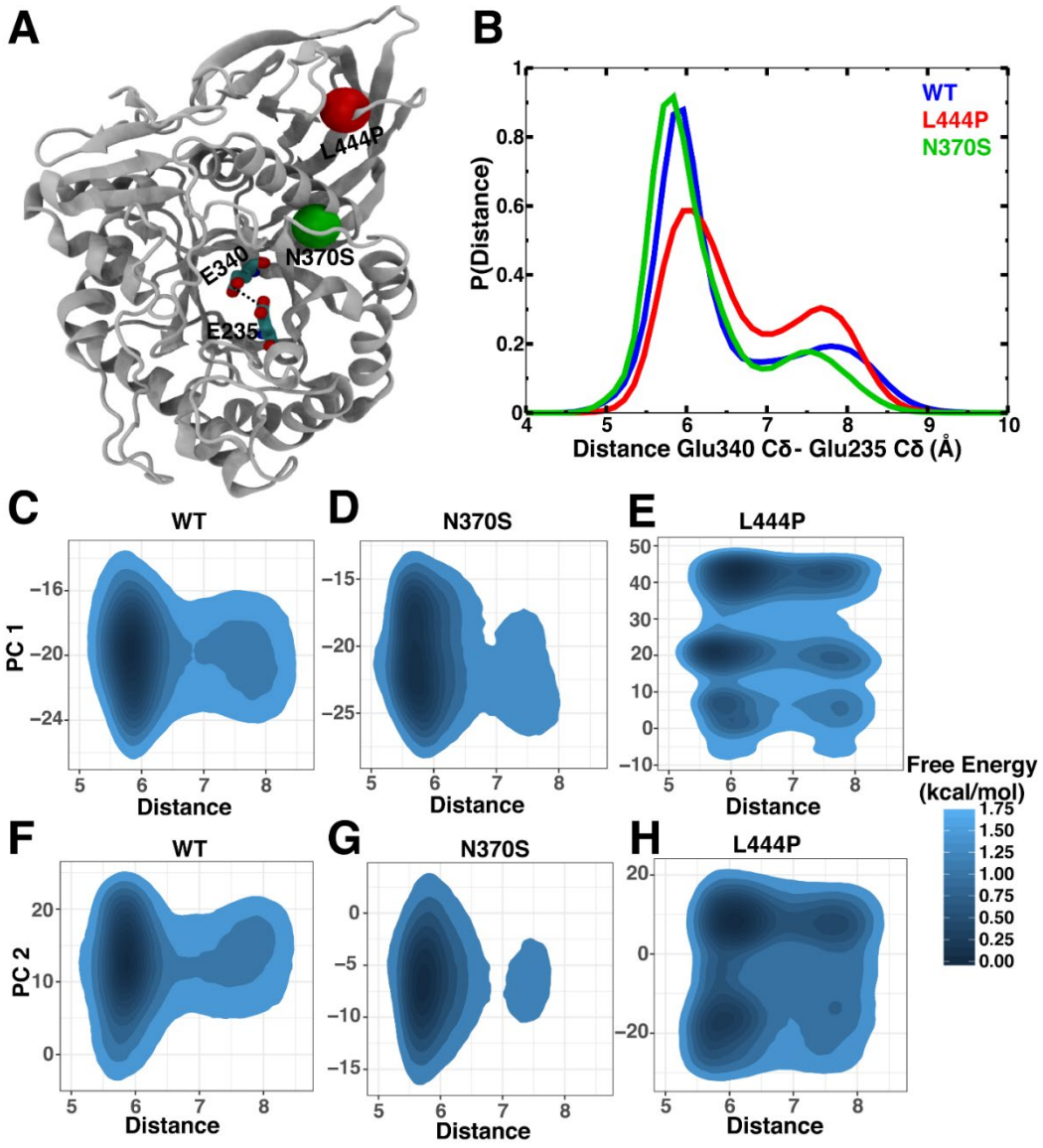

Figure S7. Distance distributions between catalytic residues and free energy analyses. (A) The L444P and N370S mutation sites of the enzyme are represented as red and green beads respectively. The catalytic residues are shown as licorice. (B) The probability distribution of the distance between the δ carbons of the catalytic residues (E340 and E235) shown in (A), for WT (blue), L444P (red), and N370S (green). (C)-(H) represent the free energy landscapes calculated from the top two principal components and distances between catalytic residues of WT, N370S and L444P.

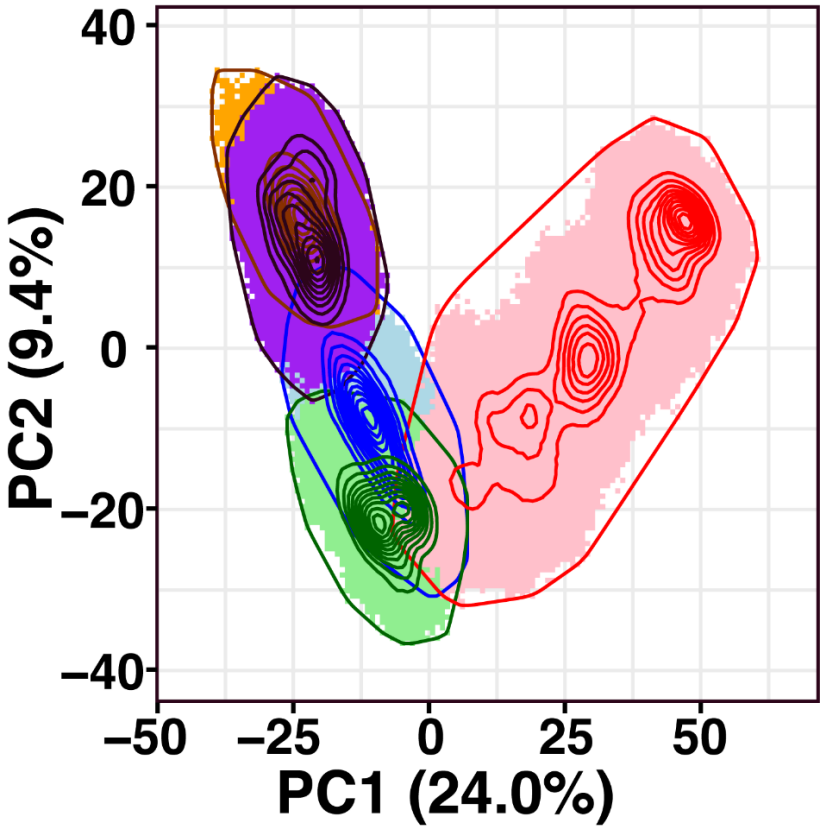

Figure S8. PCA results of L444P (red), WT (blue), N370S (green), NAG (orange) and 2NAGs (purple).

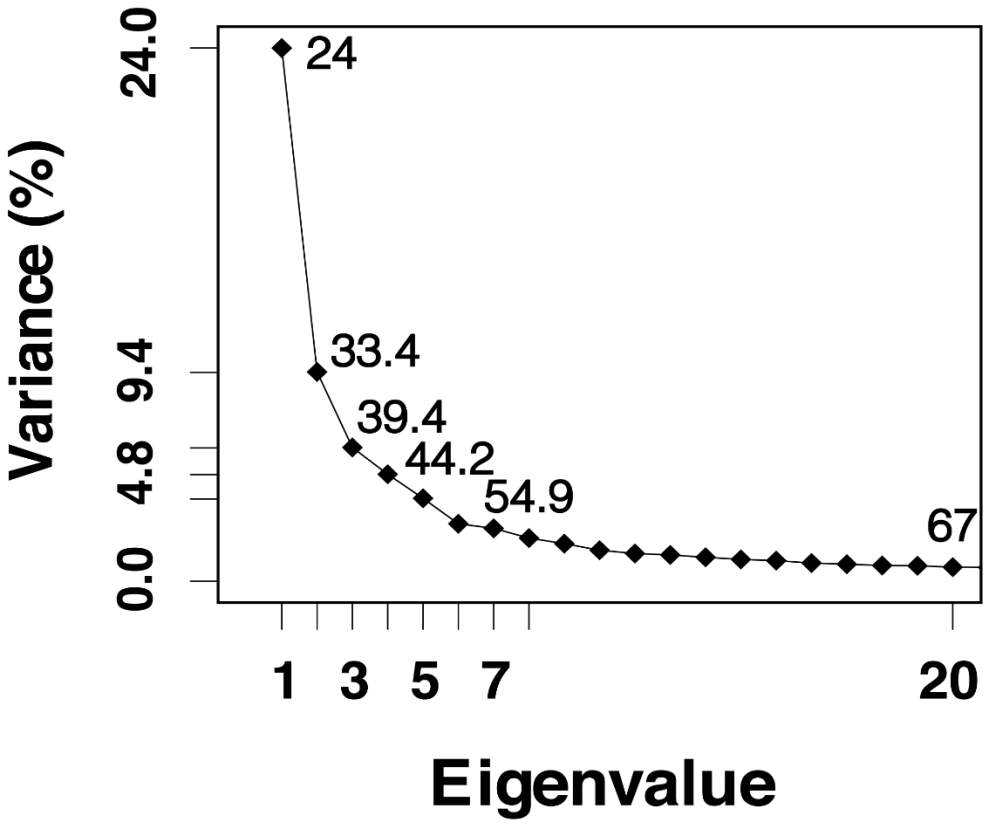

Figure S9. The scree plot for the PCA of L444P, WT, N370S, NAG, and 2NAGs.

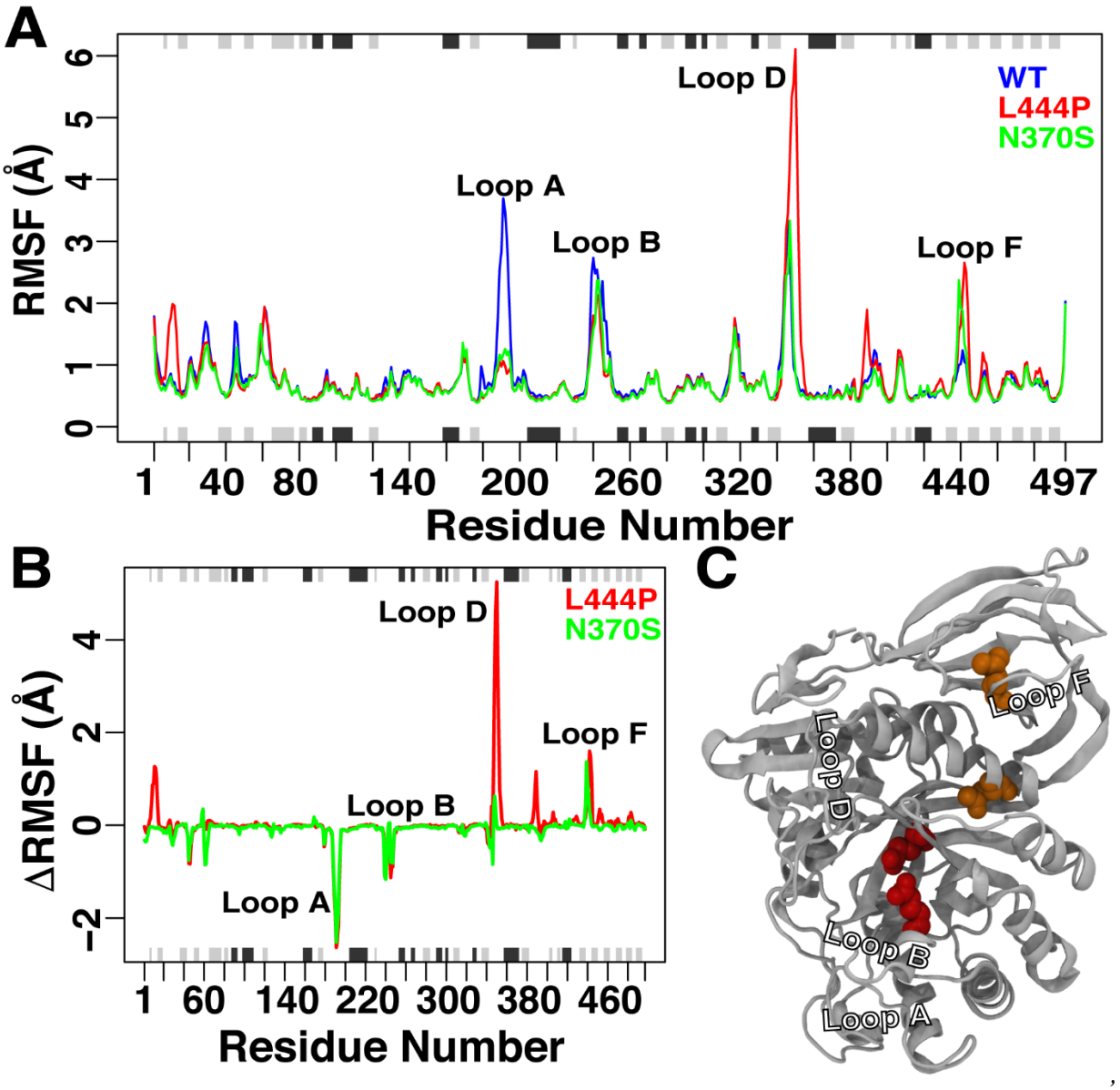

Figure S10. RMSF reveals distinct backbone conformational dynamics between WT and mutants. (A) Average RMSF analyses of the backbone atoms of WT (blue), L444P (red), and N370S (green). (B) ΔRMSF of WT to L444P (red) and WT to N370S (green). The α-helices and β-strands of the GlcCerase structure are represented by black and grey bars respectively on the top and bottom of plots (A) and (B). (C) Structural representation of GlcCerase with labeled loops A-D. Catalytic (E235 and E340) and mutated (L444 and N370) residues are represented as van der Waals' spheres in red and orange respectively.

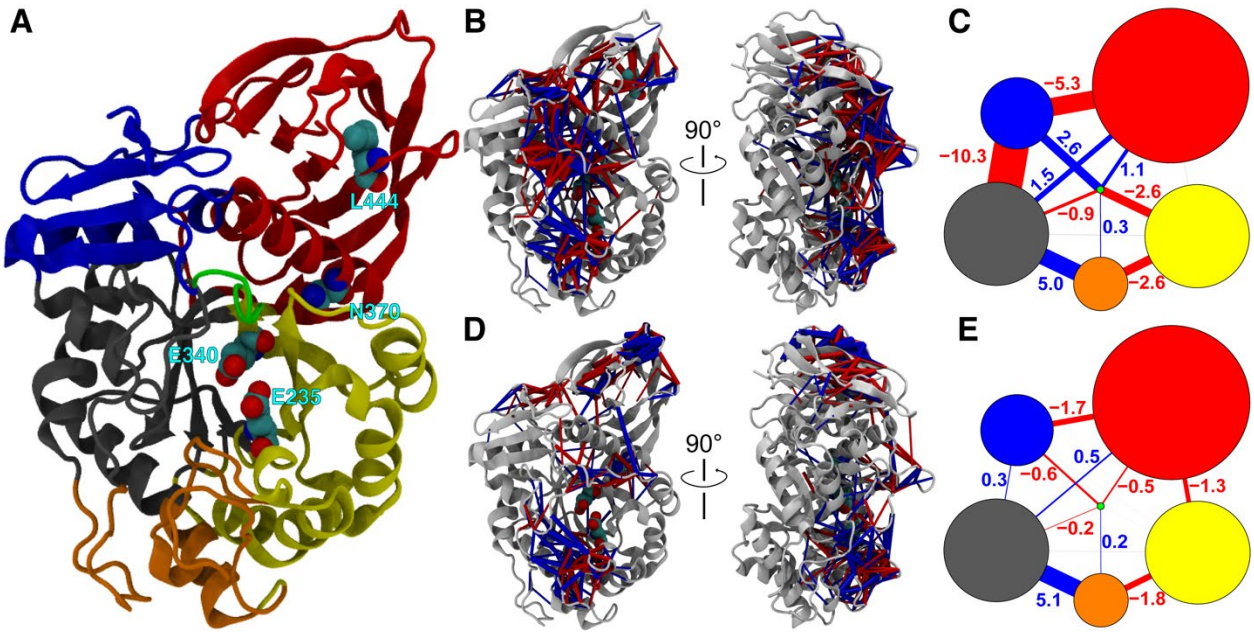

Figure S11. Catalytic regulation of acid-β-glucosidase revealed using dCNA. (A) Six consensus communities are identified and mapped to the structure of GlcCerase with different colors. (B, D) Differences of residue-residue contact probabilities (dp_c) going from WT to L444P (B) and going from WT to N370S (D). Blue cylinders indicate contacts that are more often formed ($dp_c \geq 0.1$) and red cylinders contacts less often formed ($dp_c \leq -0.1$). In both (B) and (D), cylinder radii are proportional to the absolute value of dp_c. Community networks in (C) and (E) correspond to the residue-residue contact analysis in (B) and (D), respectively. Colors of vertices are the same as colors of communities in (A). The radii of community vertices in (C) and (E) correlate to the number of residues included in the communities. The red and blue lines in (C) and (E) represent the net residue-residue contact probability differences between communities, annotated as the quantitative values shown.

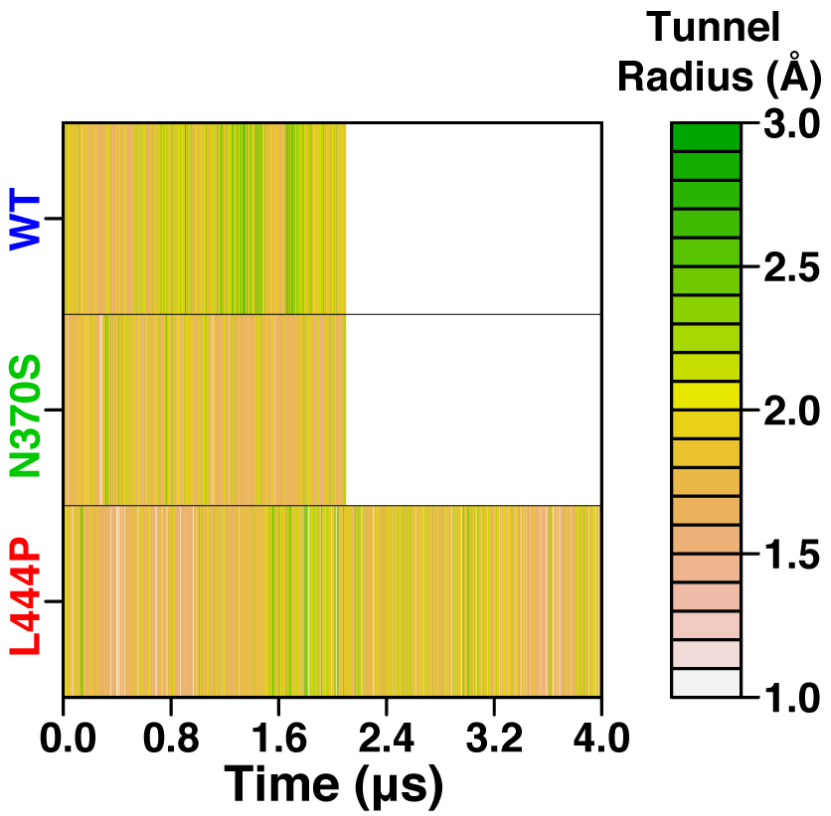

Figure S12. Tunnel bottleneck radii generated over time. Bottleneck tunnel radii of WT, L444P and N370S trajectories are shown from narrow (white) to wide (green) measurements. Lowest cost pathways are established at bottleneck tunnel radii ≥ 1.0 Å.

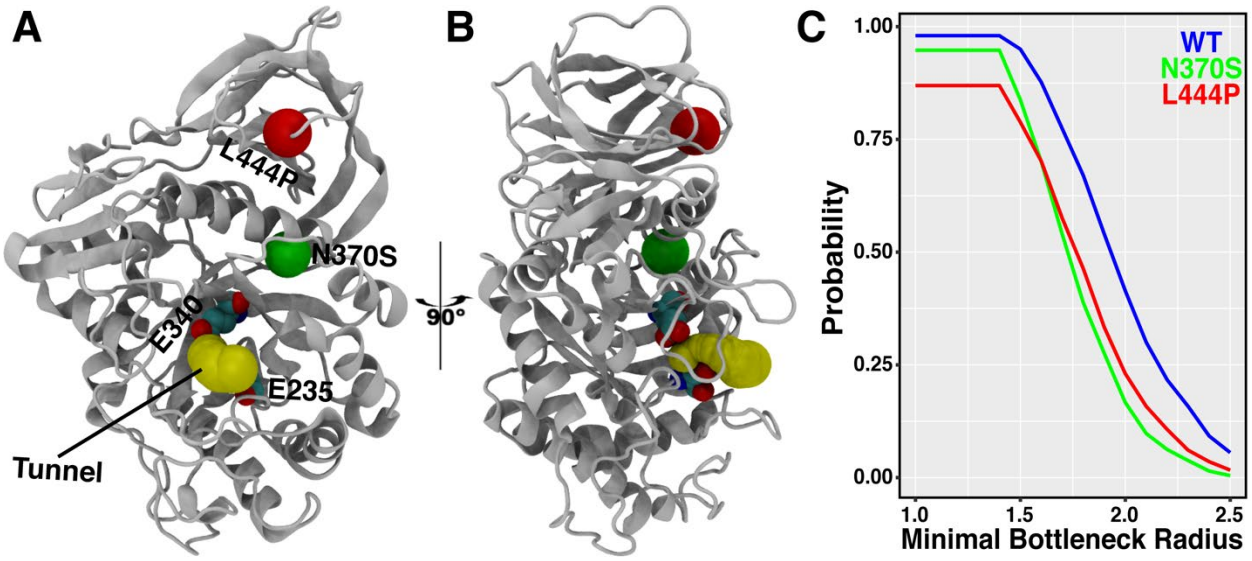

Figure S13. Dynamics of the substrate-entrance tunnel identified in acid-β-glucosidase. (A) A snapshot example of the substrate-entrance tunnel (yellow) generated by CAVER with the origin located between catalytic residues E340 and E235. The L444P and N370S mutation sites are shown as red and green beads, respectively. The 90° clockwise rotation of (A) is shown in (B). (C) The probability of open-state bottleneck radii at different cut-off measurements. Bottleneck radii results are shown as blue, red and green for WT, L444P and N370S, respectively.