PNAS
www.pnas.org

Supplementary Information for

Mechanism of Glucocerebrosidase Activation and Dysfunction in Gaucher Disease Unraveled by Molecular Dynamics and Deep Learning

Raquel Romero[^], Arvind Ramanathan[^], Tony Yuen, Debsindhu Bhowmik, Mehr Mathew, Lubna Bashir Munshi, Seher Javaid, Madison Bloch, Daria Lizneva, Alina Rahimova, Ayesha Khan, Charit Taneja, Se-Min Kim, Li Sun, Maria New[*], Shozeb Haider[*] and Mone Zaidi[*]

[^]Joint first authors
*Corresponding authors: shozeb.haider@ucl.ac.uk, maria.new@mssm.edu and mone.zaidi@mssm.edu

This PDF file includes:

Figs. S1 to S37
Table S1

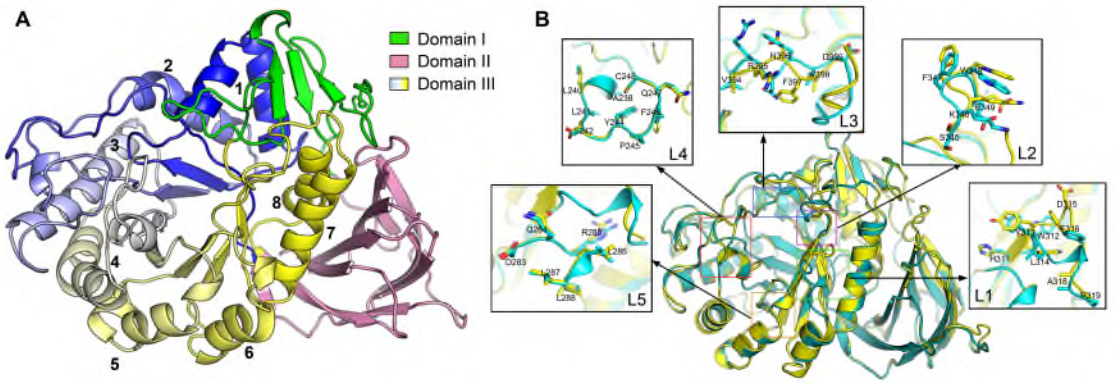


Figure S1: (**A**) Crystal structure of GCase (PDB: 2NSX), showing Domain I (green), Domain II (pink) and Domain III (gradient blue to yellow). The helices of Domain III have been numbered. (**B**) Comparison of the loops at the entrance of the binding site in active (cyan) and inactive (yellow) GCase.

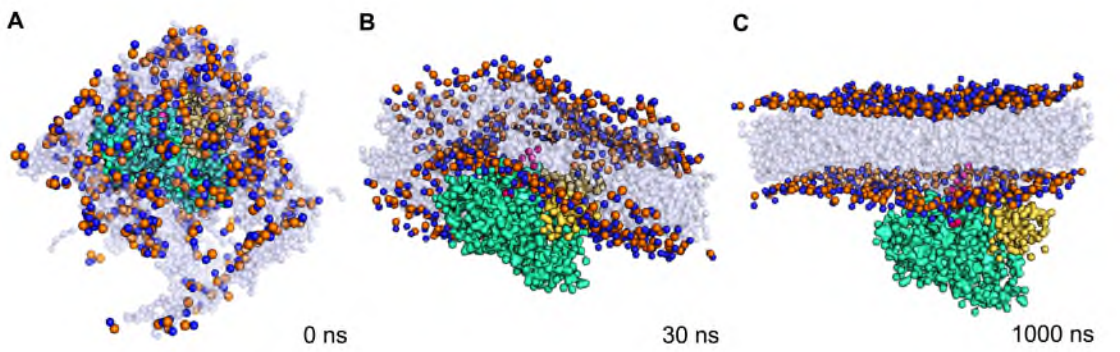


Figure S2: Self-assembly of GCase, SAPC and GluCer. Snapshots taken at (**A**) 0 ns, (**B**) 30 ns and (**C**) 1000 ns from coarse-grained simulation of the complex (3-CG). At 30 ns, the membrane is not completely formed although the bilayer has taken shape, while after ~120 ns, the membrane is completely formed. The protein–substrate complex is well anchored until the end of the simulation (1000 ns). (DPPC lipids, mauve; GCase, green; SAPC, yellow; and GluCer, red)

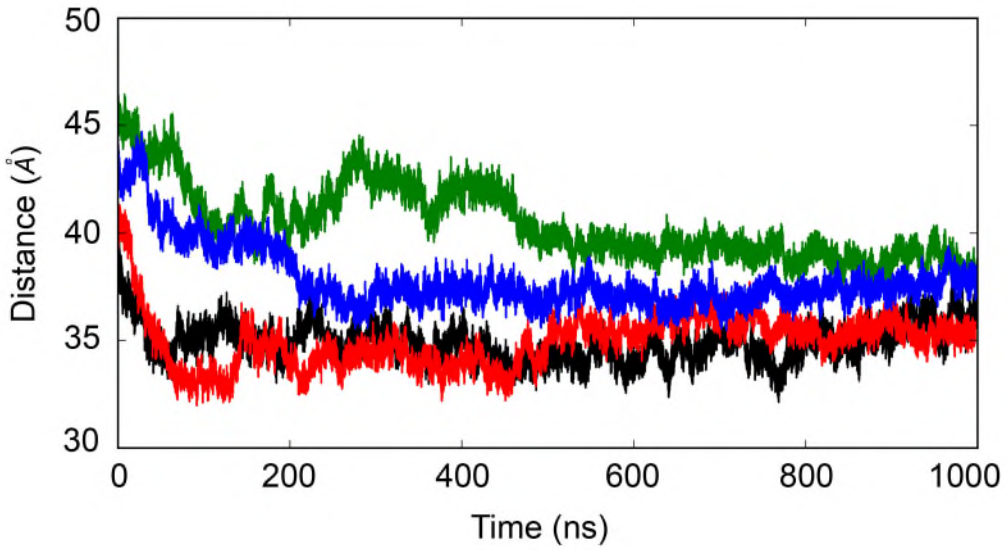


Figure S3: Distance between the center of mass of GCase and the center of mass of the lipid membrane. From 500 ns, proteins and complexes adopt a stable configuration on the membrane. Both complexes exhibit higher distances between centers of mass as the facilitator protein SAPC is positioned between GCase and the membrane (Black – active GCase, No SAPC, simulation 2a; Red – inactive GCase, no SAPC, simulation 2b; Green – active complex, simulation 3a; Blue – inactive complex, simulation 3b).

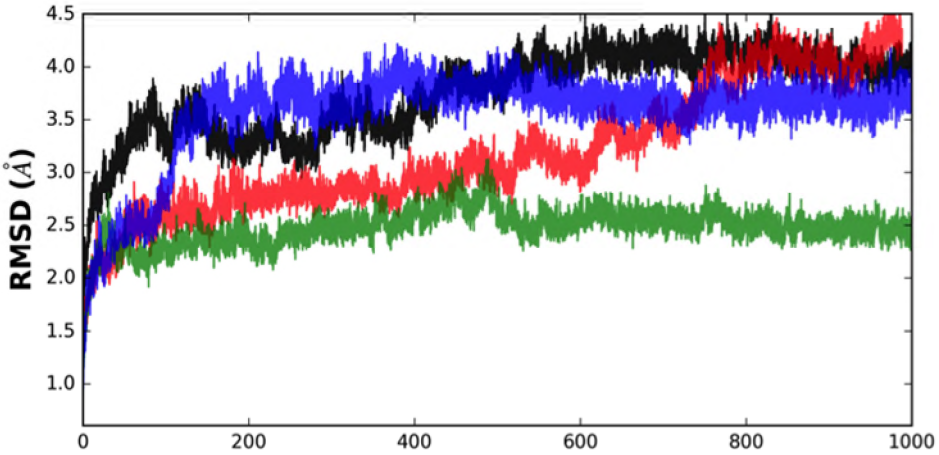


Figure S4: Cα-RMSD values of GCase, plotted as a function of time for simulations 2a, 2b, 3a and 3b. The conformation of GCase is more stable when it is simulated along with its facilitator protein SAPC in simulations 3a and 3b (Black – active GCase, No SAPC, simulation 2a; Red – inactive GCase, no SAPC, simulation 2b; Green – active complex, simulation 3a; Blue – inactive complex, simulation 3b).

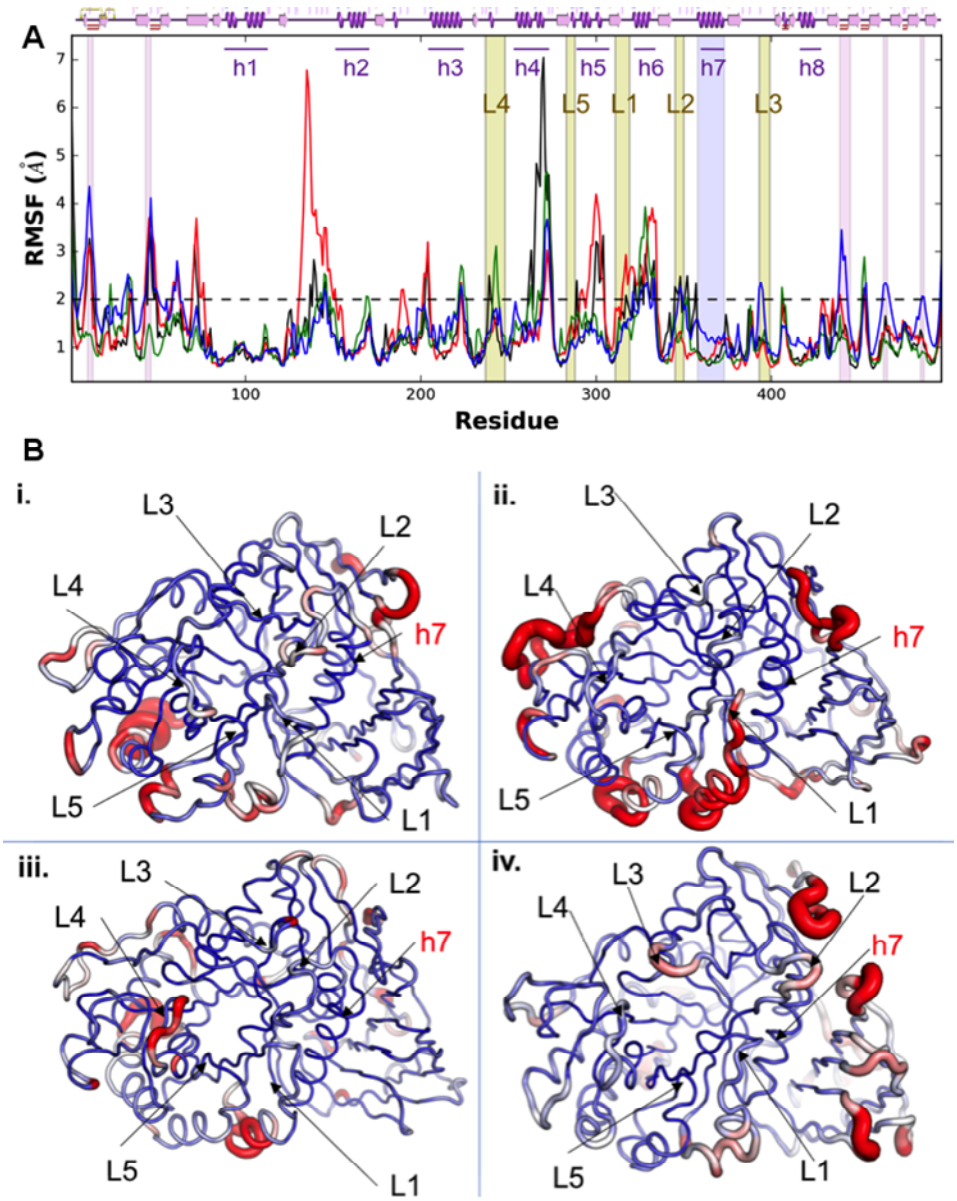


Figure S5: (**A**) Comparison of RMSF (GCase) as a function of each residue in simulations 2a, 2b, 3a and 3b. Loops-1 to -5 at the entrance of the binding site have been highlighted in yellow and tagged with the label L1 to L5, helix-7 has been highlighted in blue and the protein-protein binding site, other than Loops 1 and 2 and helix-7, has been highlighted in magenta. The secondary structure of GCase is annotated on top of the graph and helices (h1-h8) of the TIM barrel have been labeled. (Black – active GCase, No SAPC, simulation 2a; Red – inactive GCase, no SAPC, simulation 2b; Green – active complex, simulation 3a; Blue – inactive complex, simulation 3b) (**B**) RMSF values from the simulation translated on the structure of GCase in simulation (i) 2a (active GCase, no SAPC) (ii) 2b (inactive GCase, no SAPC) (iii) 3a (active complex) and (iv) 3b (inactive complex).

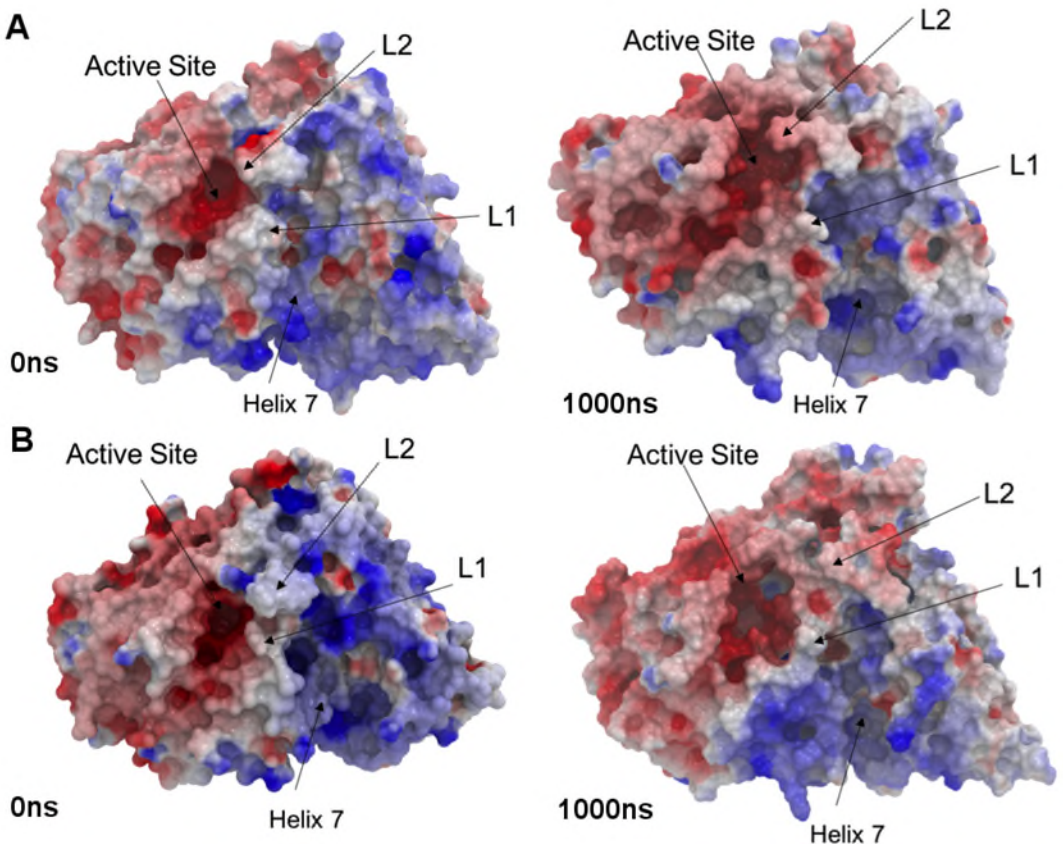


Figure S6: Snapshots of the electrostatic surface in simulation (**A**) 3a (active complex) and (**B**) 3b (inactive complex) at 0ns and 1000ns. Loop-1 does not change conformation in simulation 3a as it does in simulation 3b; Loop-4 progress towards the active site in simulation 2 although it does not occlude the active site significantly. It is worth noting that Loop-1 and Loop-3 change conformation that leads to widening of the active site. The position of helix-7 has also been highlighted to illustrate the electropositive cluster.

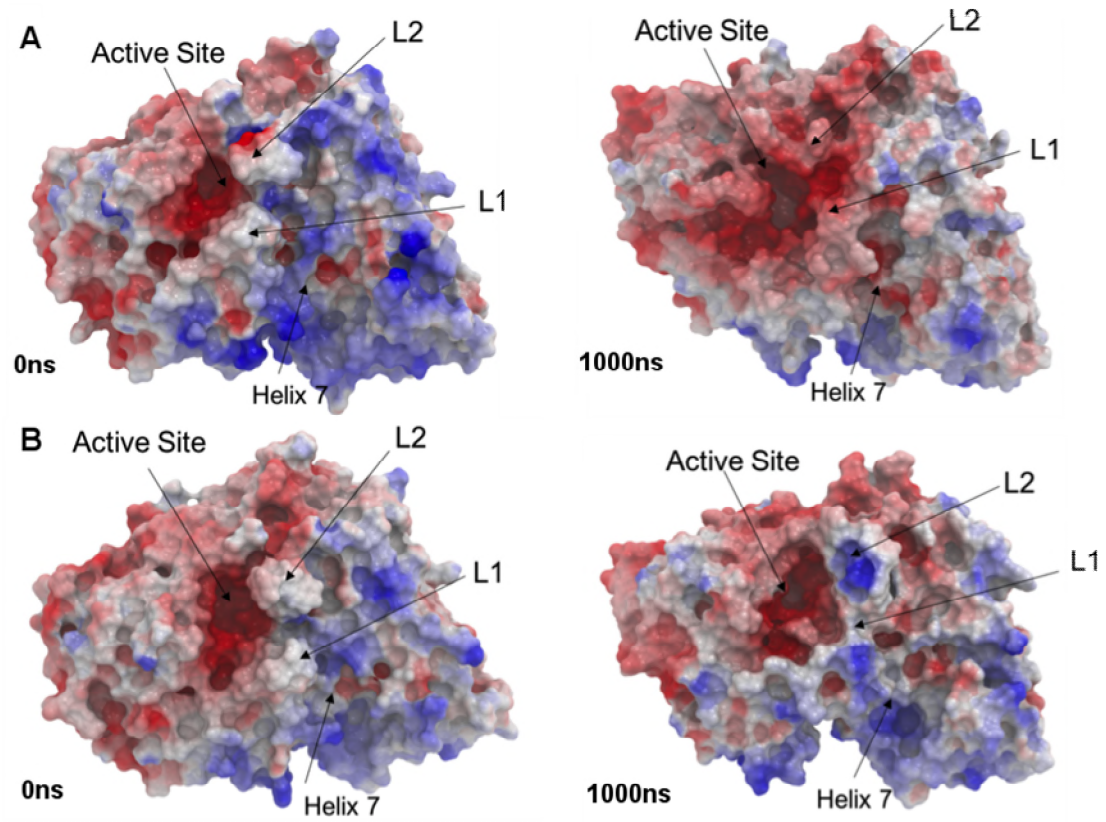


Figure S7: Snapshots of the electrostatic surface in simulation (**A**) 5a ($GCase^{N370S}$ active state) and (**B**) 6a ($GCase^{L444P}$ active state) at 0ns and 1000ns. The position of active site, Loop-1, -2 and helix-7 have been highlighted.

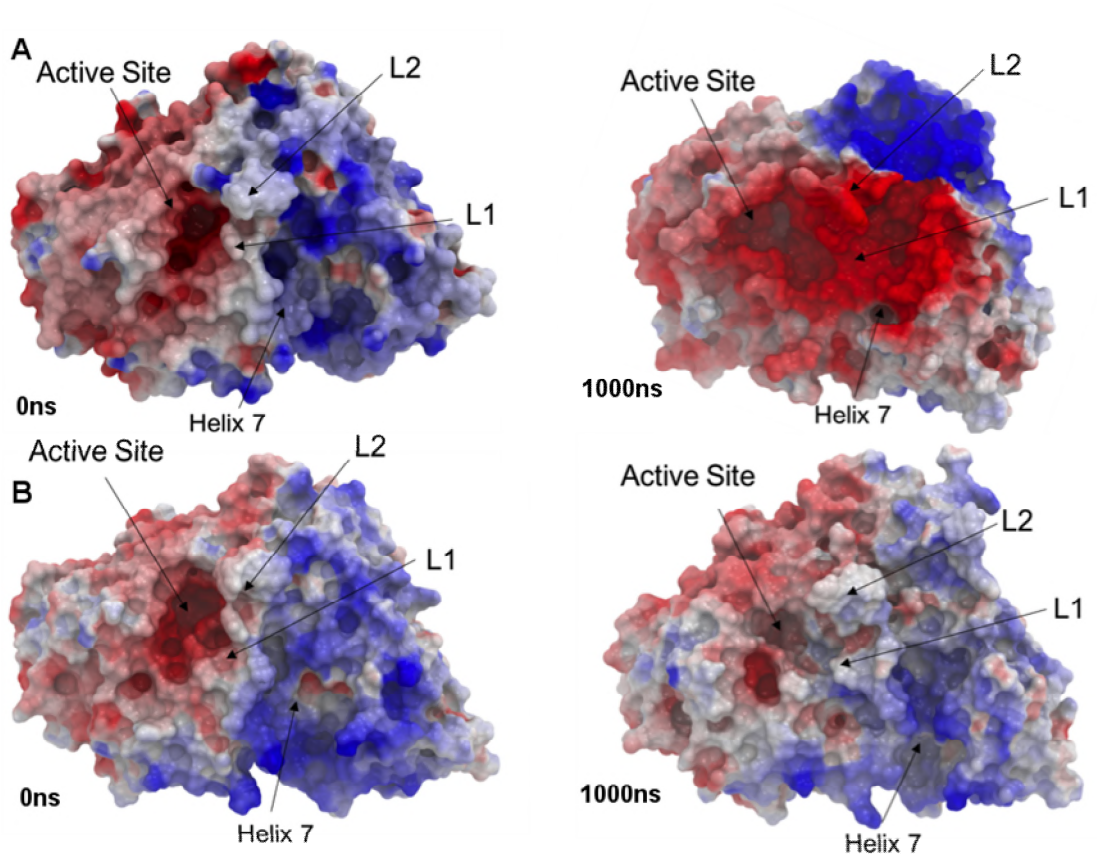


Figure S8: Snapshots of the electrostatic surface in simulation (**A**) 5b ($GCase^{N370S}$ inactive state) and (**B**) 6b ($GCase^{L444P}$ inactive state) at 0ns and 1000ns. The position of active site, Loop-1, -2 and Helix-7 have been highlighted.

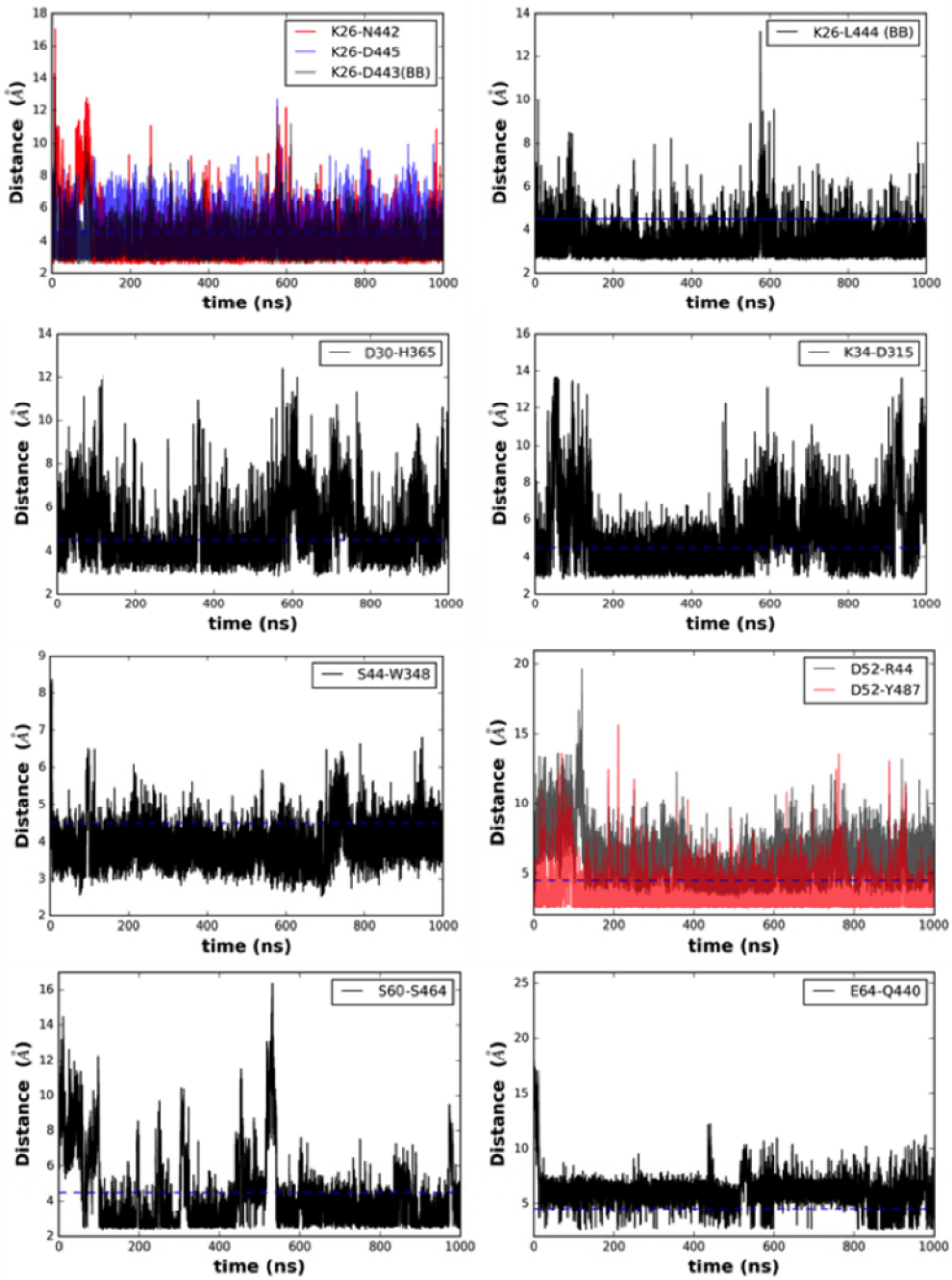


Figure S9: Interactions at the protein-protein interface observed in simulation 3a (active complex). Structural representations of these interactions are illustrated in Figure 4.

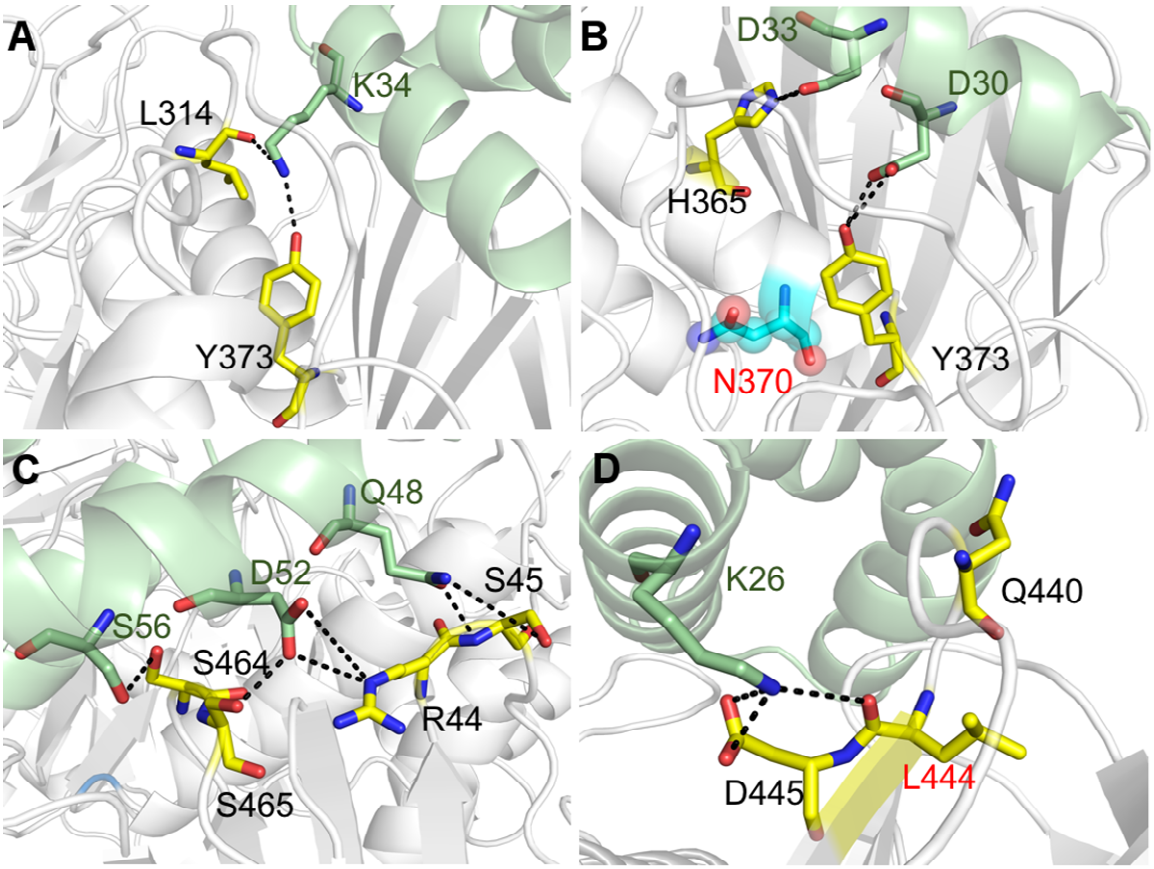


Figure S10: Protein-protein interactions in simulation 3b (inactive complex) at 1000ns, in (**A**) Loop-1, (**B**) helix-7, and (**C** and **D**) Domain II. SAPC is coloured in green and interacting residues in GCase are coloured yellow. Position of residue N370 has been represented with spheres and coloured in cyan. Distances of the interactions, over the course of the simulation, have been illustrated in Supplementary Figure S13.

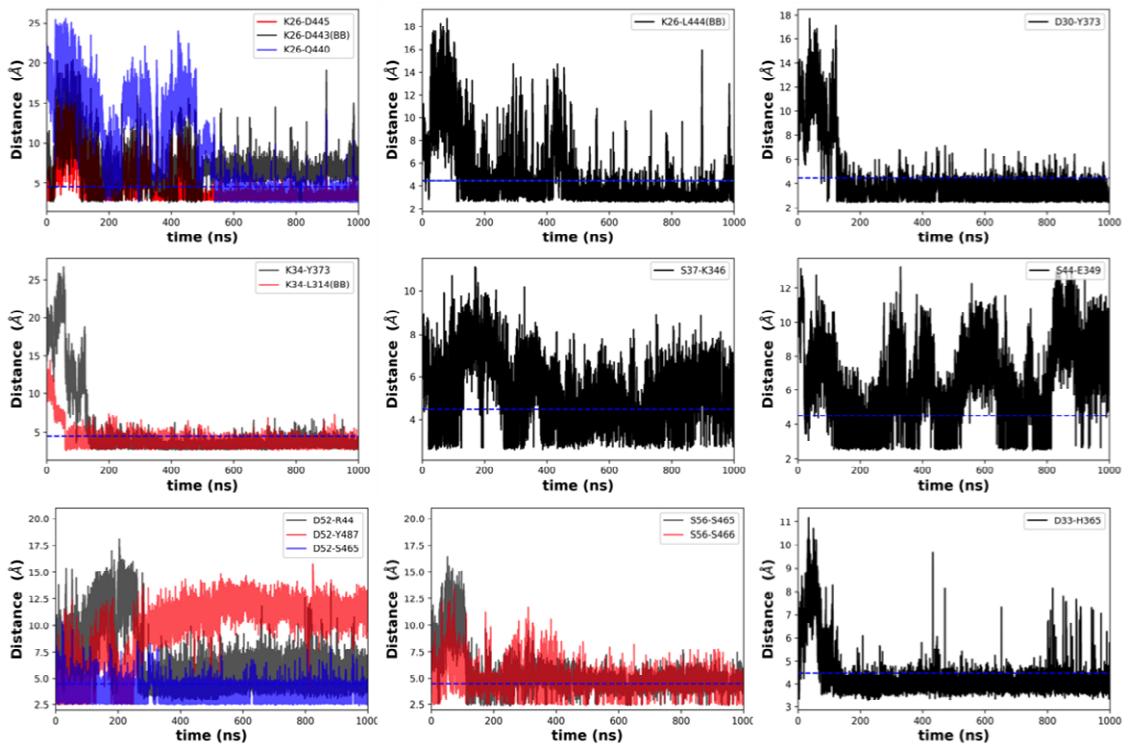


Figure S11: Interactions at the protein-protein interface observed in simulation 3b (inactive complex). Structural representations of these interactions are illustrated in Supplementary Figure S12.

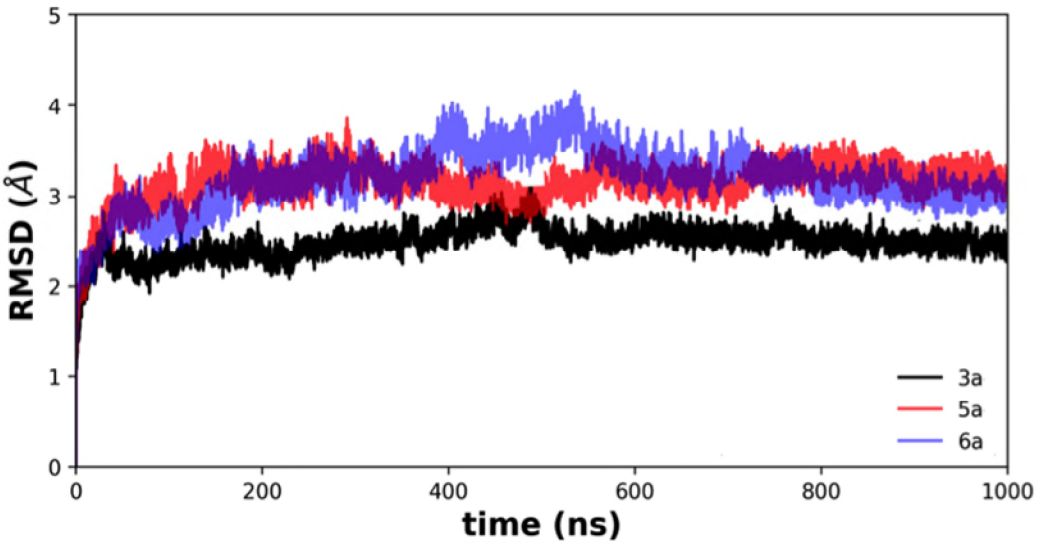


Figure S12: Cα-RMSD values of GCase, plotted as a function of time for simulations 3a (active complex, black), 5a ($GCase^{N370S}$ active state, red) and 6a ($GCase^{L444P}$ active state, blue).

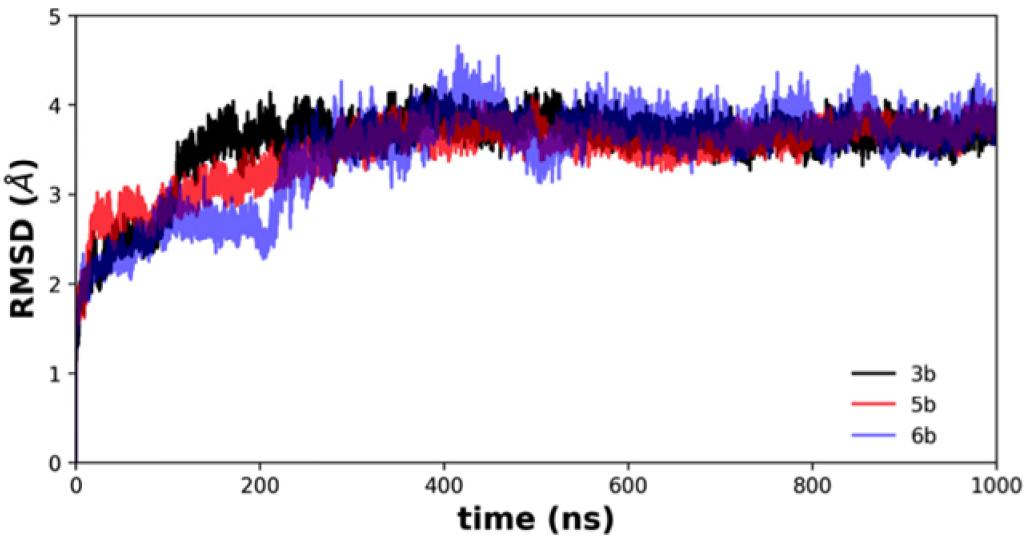


Figure S13: Cα-RMSD values of GCase, plotted as a function of time for simulations 3b (inactive complex, black), 5b ($GCase^{N370S}$ inactive state, red) and 6b ($GCase^{L444P}$ inactive state, blue).

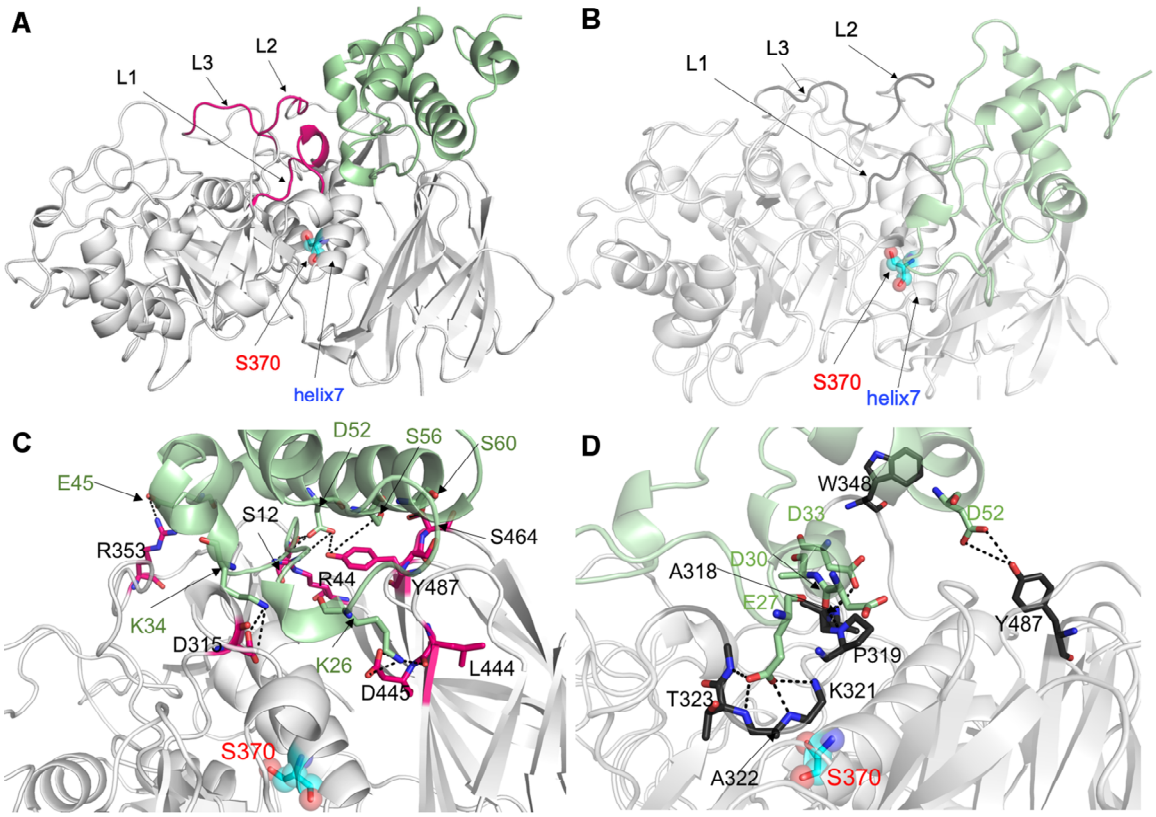


Figure S14: Protein-protein interaction in the GCase^{N370S} complex mutant simulations. **Top row**: General overview of protein-protein interface in (**A**) GCase^{N370S} (active state, simulation 5a) and (**B**) Gcase^{N370S} (inactive state, simulation 5b) complex. GCase is coloured white and SAPC is green. The loops at the entrance of the binding site have been highlighted in pink (simulation 5a) and grey (simulation 5b). The position of mutant residue S370 is coloured cyan. **Bottom row:** Detailed view of the protein-protein interface is illustrated for (**C**) GCase^{N370S} (active state, simulation 5a) and (**D**) Gcase^{N370S} (inactive state, simulation 5b). The corresponding residues have been coloured pink (simulation 5a) and grey (simulation 5b). Comparing with the wild type simulation (3a), some of the interactions are retained in simulation 5a, where as most of the analogous interactions are disrupted in simulation 5b.

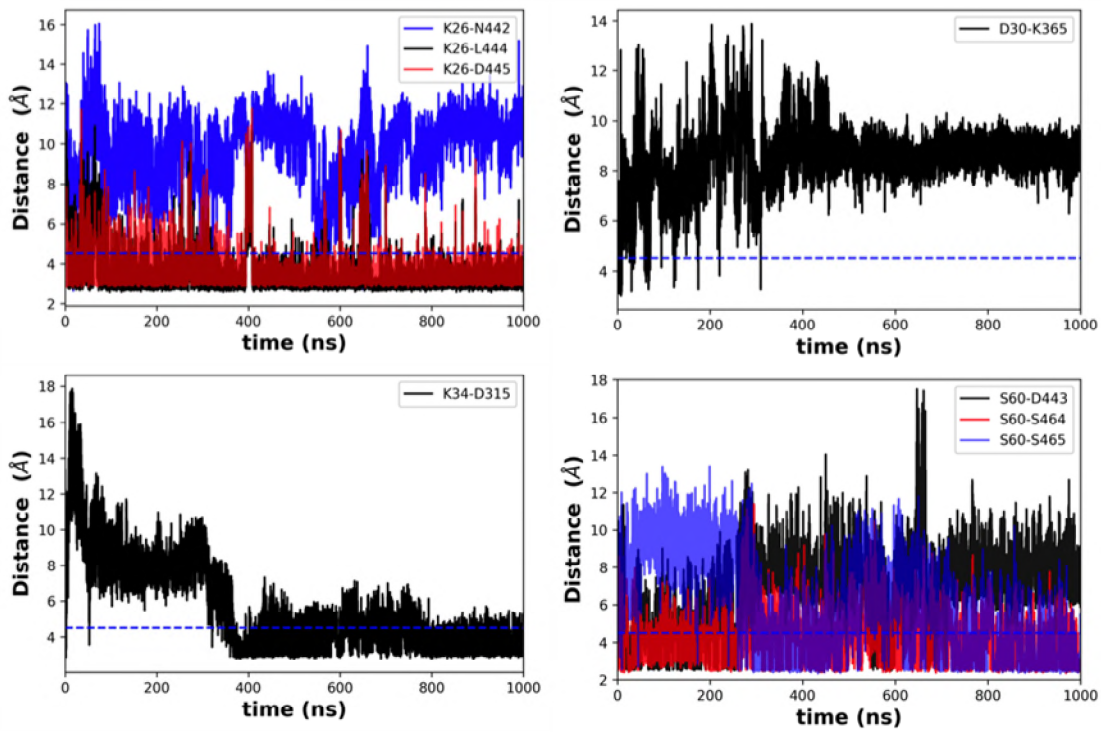


Figure S15: Selected interactions at the protein-protein interface observed in simulation 5a (GCase^{N370S} active complex).

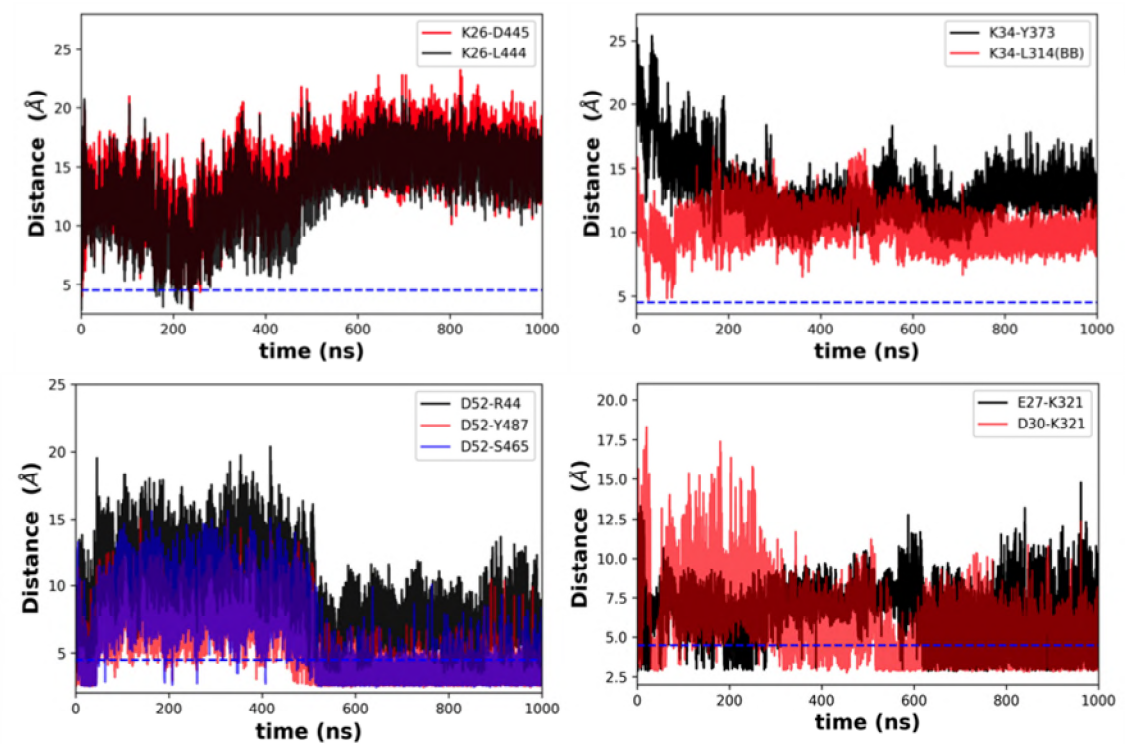
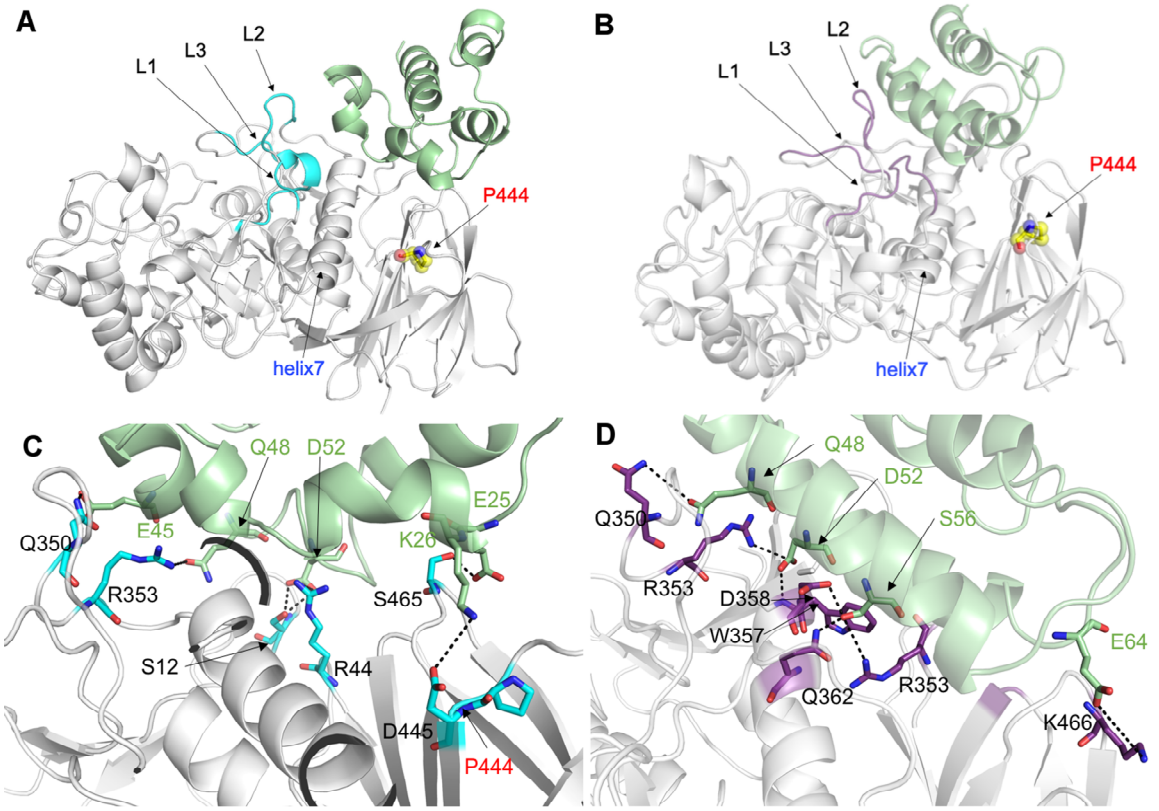


Figure S16: Selected interactions at the protein-protein interface observed in simulation 5b (GCase^{N370S} inactive complex).



<u>Figure S17:</u> Protein-protein interaction in the GCase^{L444P} complex mutant simulations. **Top row**: General overview of protein-protein interface in (**A**) GCase^{L444P} (active state, simulation 6a) and (**B**) Gcase^{L444P} (inactive state, simulation 6b) complex. GCase is coloured white and SAPC is green. The loops at the entrance of the binding site have been highlighted in cyan (simulation 6a) and purple (simulation 6b). The position of mutant residue P444 is coloured yellow. **Bottom row:** Detailed view of the protein-protein interface is illustrated for (**C**) GCase^{L444P} (active state, simulation 6a) and (**D**) Gcase^{L444P} (inactive state, simulation 6b). The corresponding residues have been coloured cyan (simulation 6a) and purple (simulation 6b). Comparing with the wild type simulation (3a), most of the interactions are disrupted in both simulations 6a and 6b. The effect is pronounced in simulation 6b, where all interactions in Loop-1 are lost.

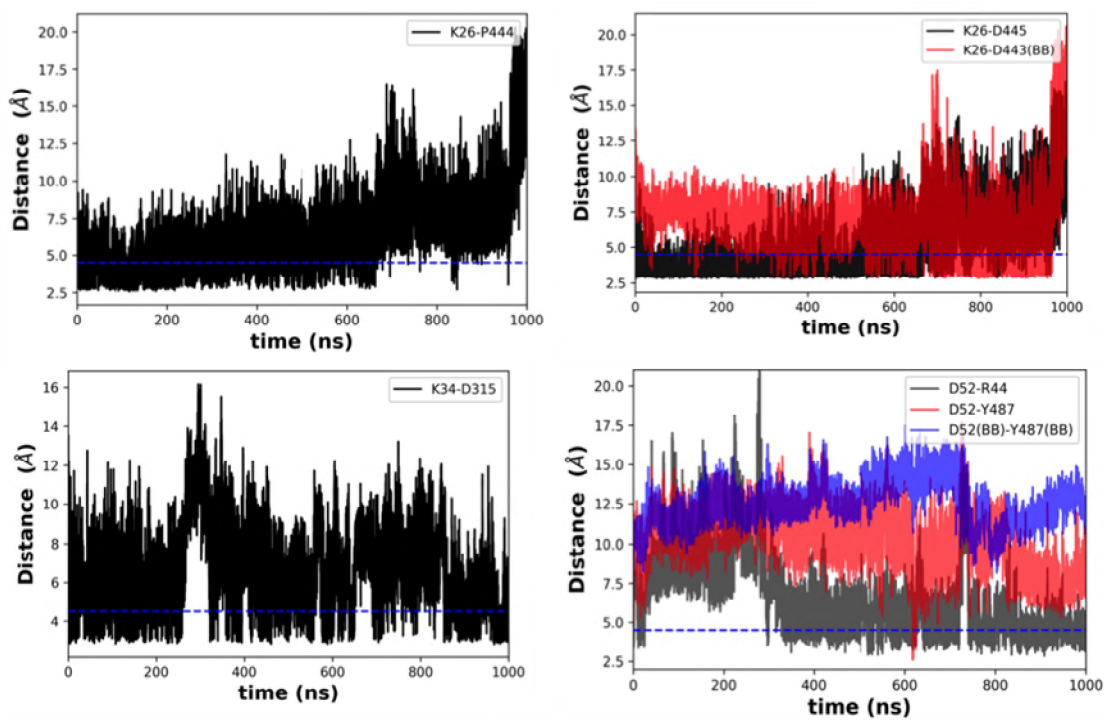


Figure S18: Selected interactions at the protein-protein interface observed in simulation 6a (GCase^{L444P} active complex).

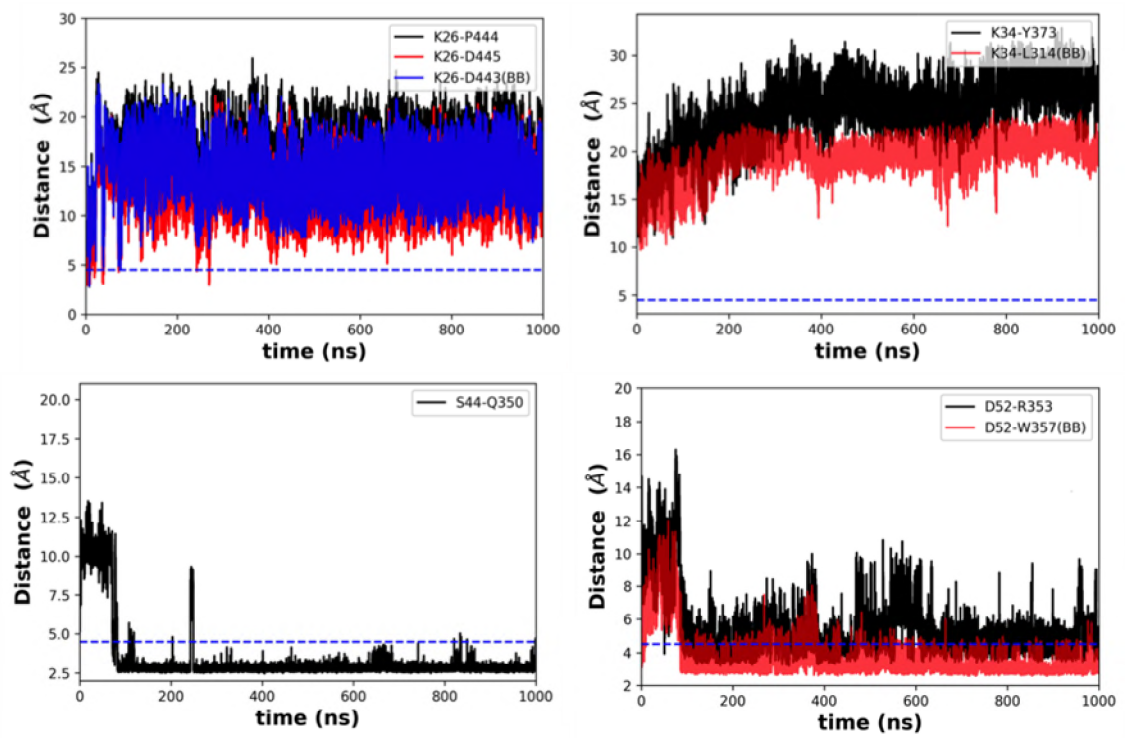


Figure S19: Selected interactions at the protein-protein interface observed in simulation 6b (GCase^{L444P} inactive complex).

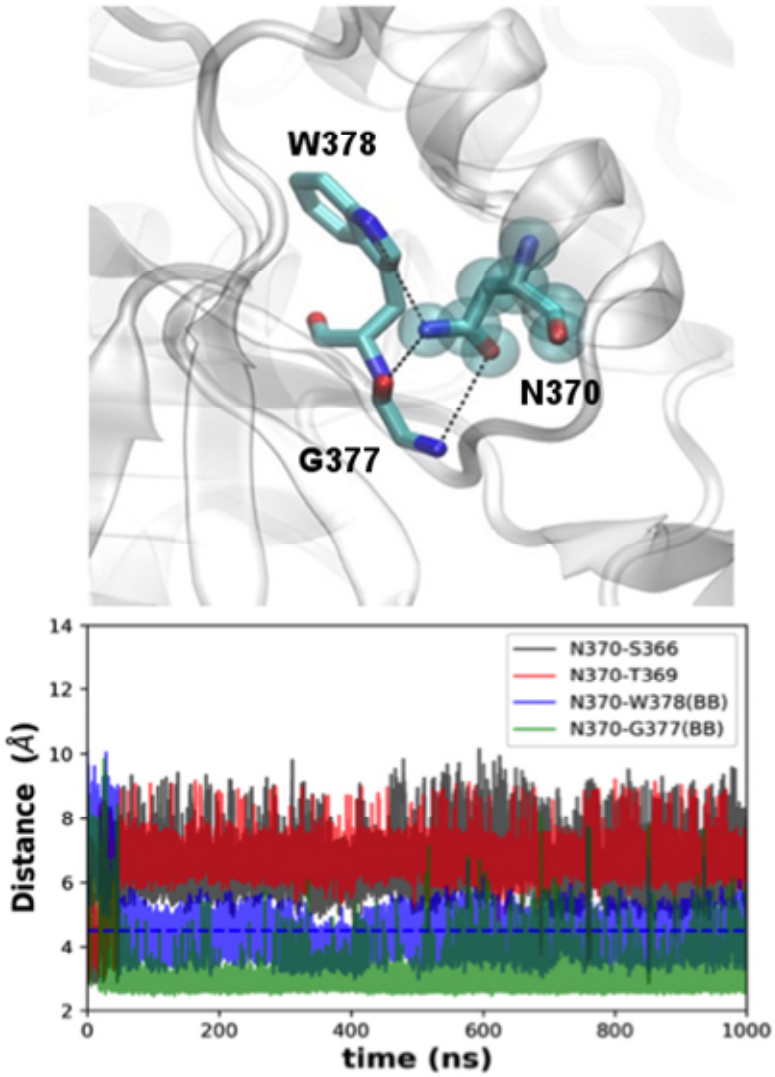


Figure S20: **Interactions of N370 in the GCase active complex (simulation 3a).** At the start of the simulation, N370 interacts with the side chains of residues T369 and S366. During equilibration, the side chain of N370 located in helix-7 of Domain III flips towards β-strand 7 forming stable hydrogen bonds with the backbone atoms of residues W378 and G377, which are maintained throughout the course of the simulation. **Top:** Snapshot of interactions between N370 and W378 (bb) and G377 (bb) in simulation 3a (GCase active complex) at 1000ns. **Bottom:** Distance between residue N370 and S366, T369, G377 and W378 in simulation 3a.

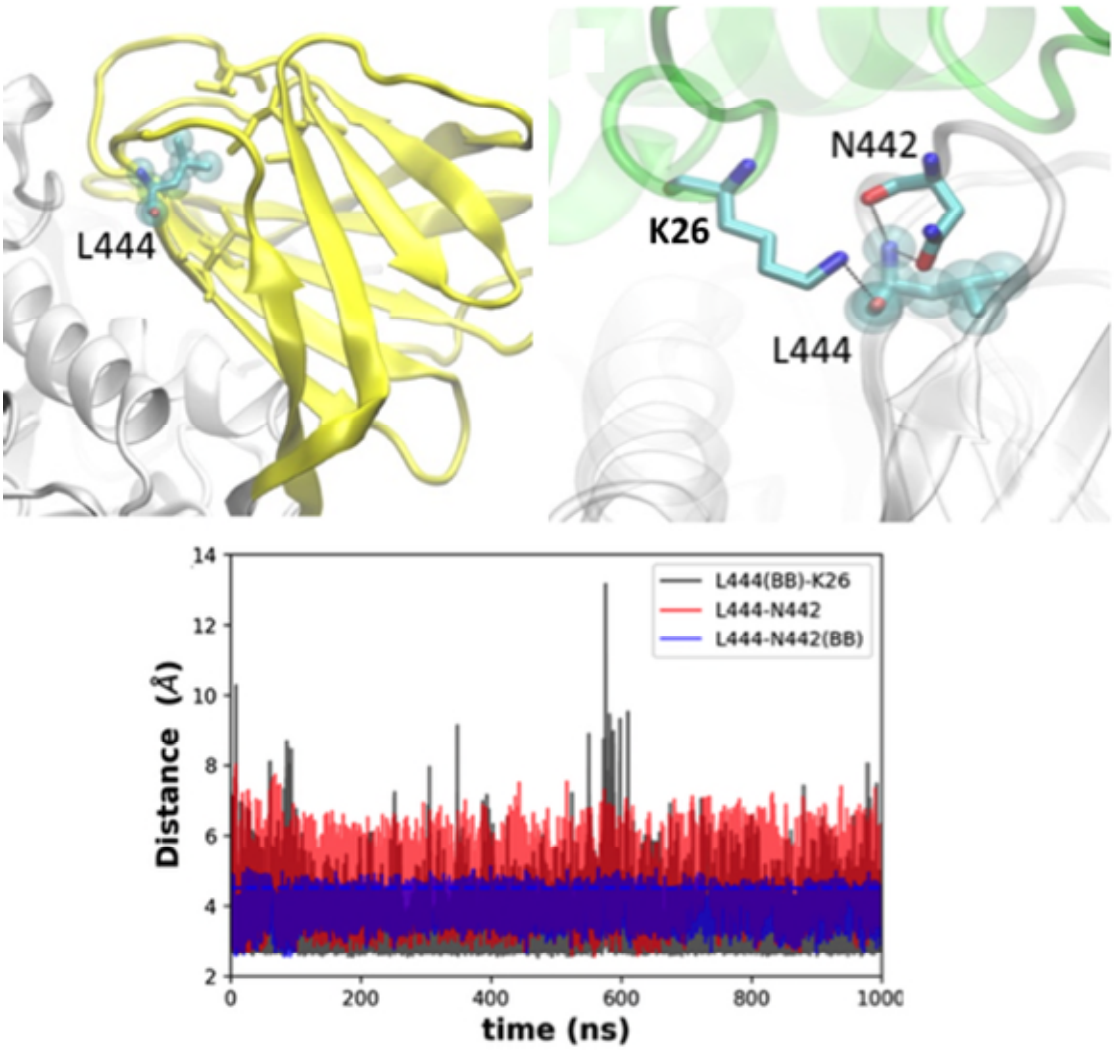


Figure S21: **Interactions of L444 in the GCase active complex (simulation 3a). Top (L):** The side chain of L444 sits in a hydrophobic pocket formed between the two β-sheets of Domain II. Residues A446, V460, V468, L470, I60, L65 also form a part of this pocket. (R) Backbone of residue L444 forms stable hydrogen bonds with the side chain of the residue K26 of SAPC and N442, as well as with the backbone of residue N442. **Bottom:** Distance between residue L444 (bb) and residues K26 of SAPC, N442 and N442 (bb), in simulation 3a (GCase active complex).

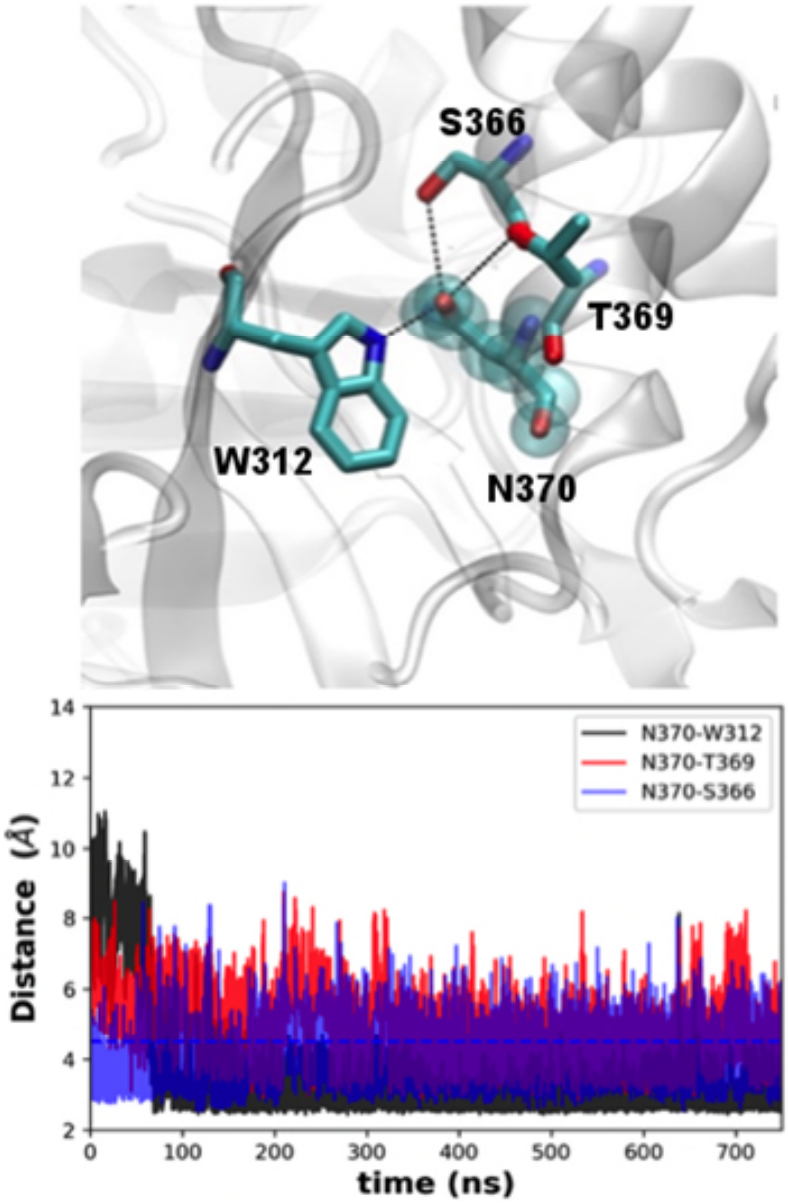


Figure S22: **Interactions of N370 in the GCase inactive complex (simulation 3b).** In the inactive conformation, N370 interacts with residue S366 and T369 in the same helix and forms very stable hydrogen bond. N370 also interacts with the side chain of residue W312 in Loop-1. It should be worth mentioning here that the difference in the conformation of Loop-1 is one of the markers that differentiate the active and inactive states of GCase. Thus an interaction between the side chain of N370 and residues in Loop-1 is a direct link that plays a role in activation. **Top:** Snapshot of interactions between N370 and W312, S366 and T369 in simulation 3b (GCase inactive complex) at 1000ns. **Bottom:** Distance between residue N370 and W312, S366 and T369 in simulation 3b.

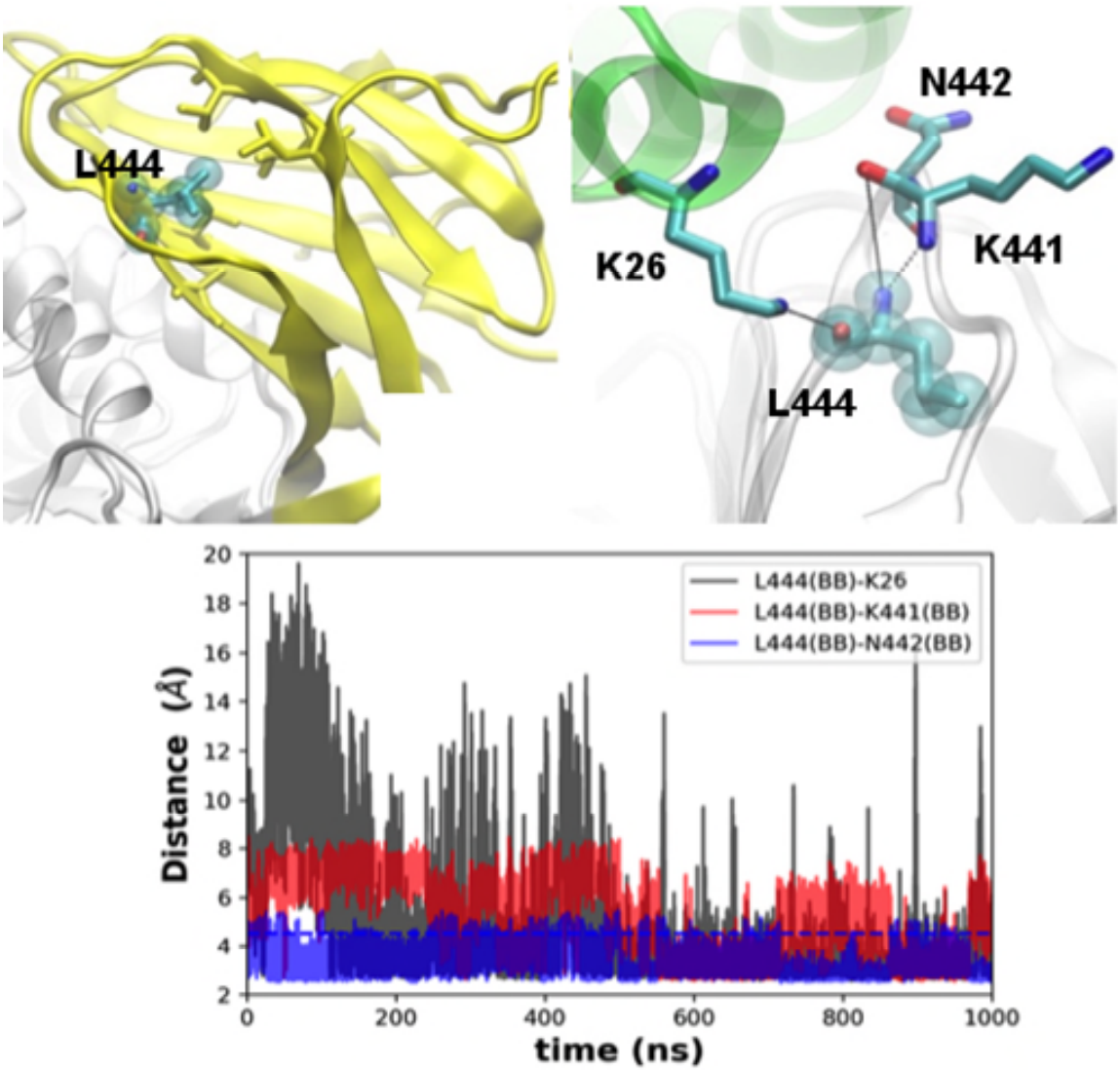


Figure S23: **Interactions of L444 in the GCase inactive complex (simulation 3b). Top (L):** The side chain of L444 sits in a hydrophobic pocket formed between the two β-sheets of Domain II. Residues A446, V460, V468, L470, I60, L65 also form a part of this pocket. (**R**) Backbone of residue L444 forms stable hydrogen bonds with the side chain of the residue K26 of SAPC and with the backbone of residue K441 and N442. **Bottom:** Distance between residue L444 (bb) and residues K26 of SAPC, K441 (bb) and N442 (bb), in simulation 3b (GCase inactive complex).

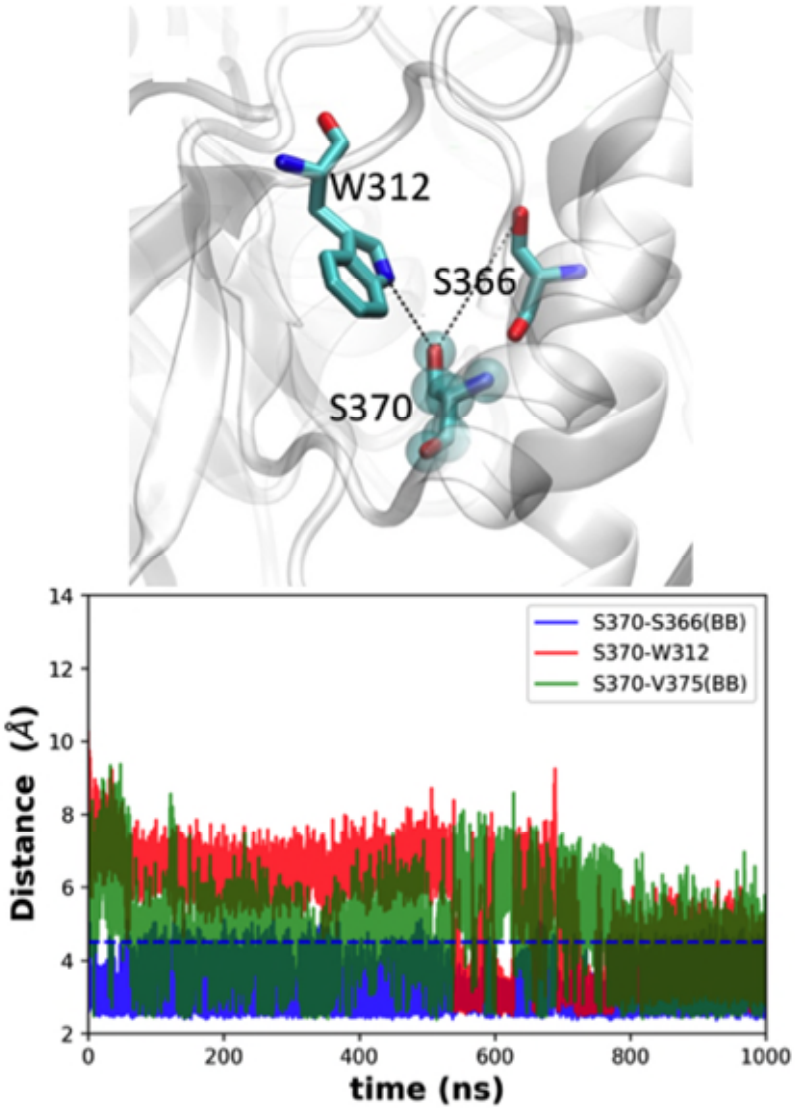


Figure S24: **Interactions of S370 in the mutant N370S active complex (simulation 5a).** In the active conformation, the side chain of S370 interacts with residue S366 in the same helix and forms very stable hydrogen bond. It also interacts with the backbone of V375 in β strand 7 up to 550 ns. Towards the end of the simulation, the mutated S370 interacts with the side chain of W312 in Loop-1. **Top:** Snapshot of interactions between N370 and W312 and S366 in simulation 5a (N370S mutant active complex) at 1000ns. **Bottom:** Distance between residue N370 and W312, S366 and V375 in simulation 5a.

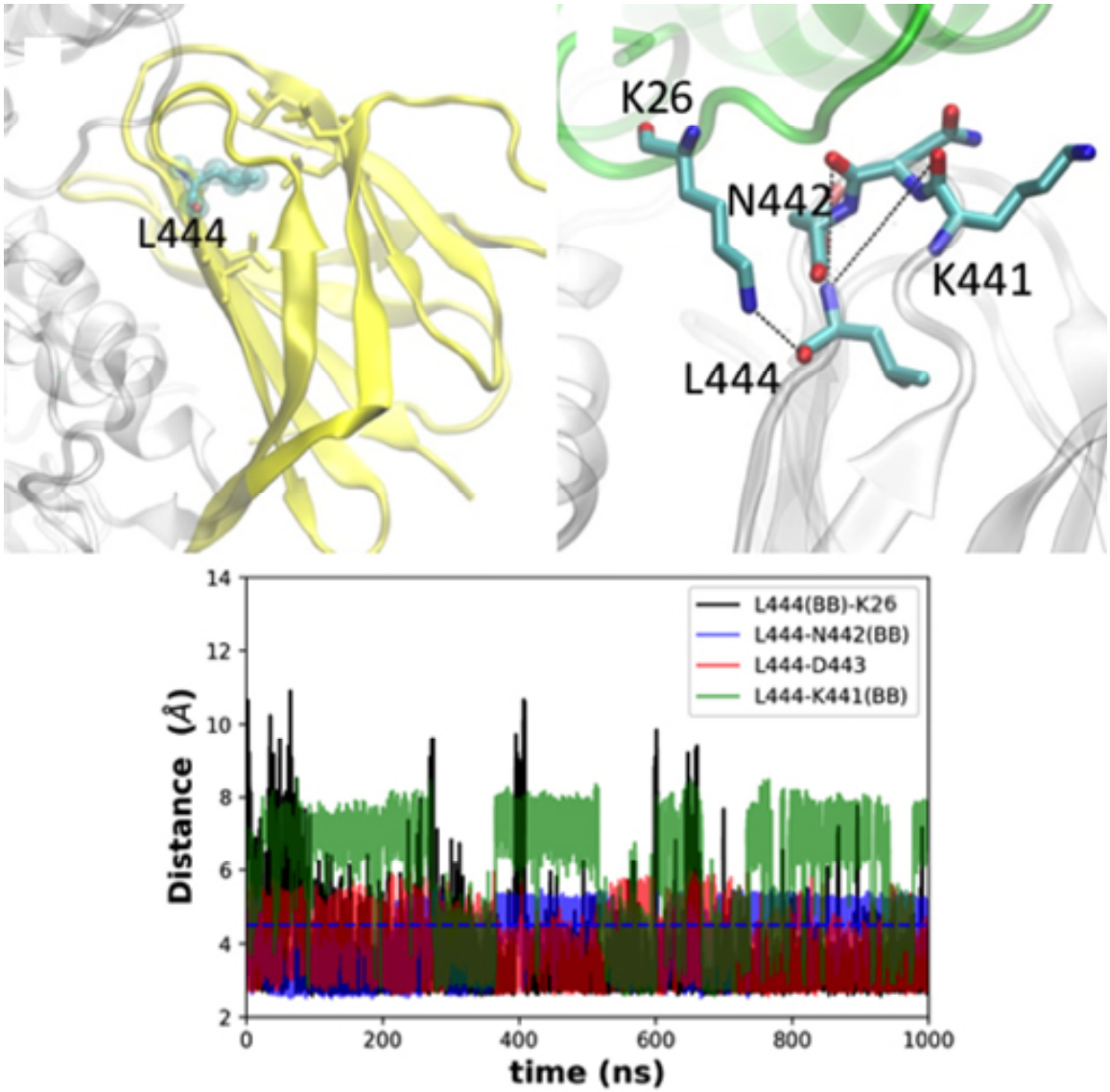
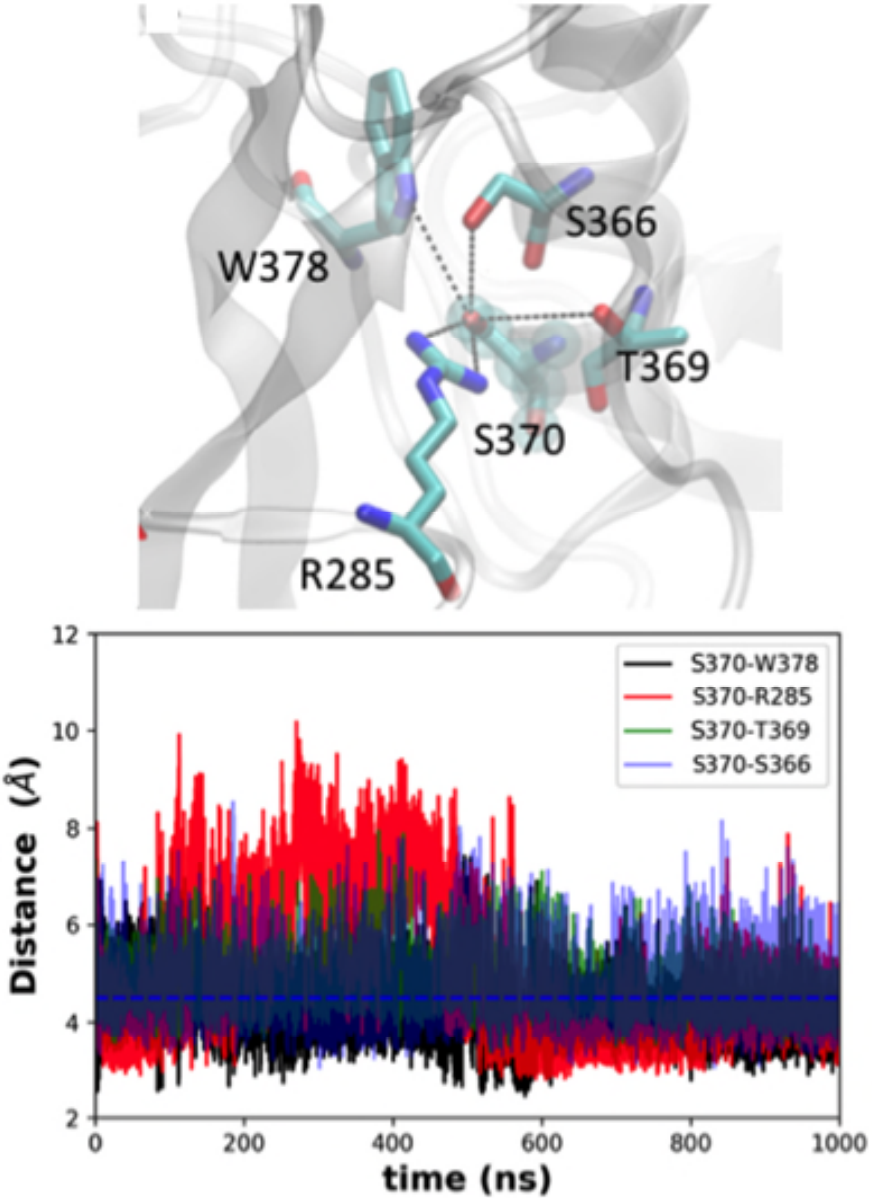


Figure S25: **Interactions of L444 in the N370S active complex (simulation 5a). Top (L):** The side chain of L444 sits in a hydrophobic pocket formed between the two β-sheets of Domain II. Residues A446, V460, V468, L470, I60, L65 form a part of this pocket. (**R**) Backbone of residue L444 forms stable hydrogen bonds with the side chain of the residue K26 of SAPC and with the backbone of residue N442 and the side chain of D443. **Bottom:** Distance between residue L444 (bb) and residues K26 (sc) of SAPC, N442 (bb), D443 (sc) and K441 (bb), in simulation 5a (N370S mutant active complex).



<u>Figure S26</u>: Interactions of S370 in the mutant N370S inactive complex (simulation 5b). In the inactive conformation, the side chain of S370 interacts with residue S366 in the same helix. It also interacts with the side chain of residues T369 and W378 in β strand 7. From ~ 500 ns, the mutated hydroxyl group in S370 interacts with the guanidinium side chain of R285 in helix-5, and forms a stable interaction until the end of the simulation. **Top:** Snapshot of interactions between S370 and side chains of R285, S366, T369 and W378 in simulation 5b (N370S mutant inactive complex) at 1000ns. **Bottom:** Distance between residue S370 and in R285, S366, T369 and W378 in simulation 5b.

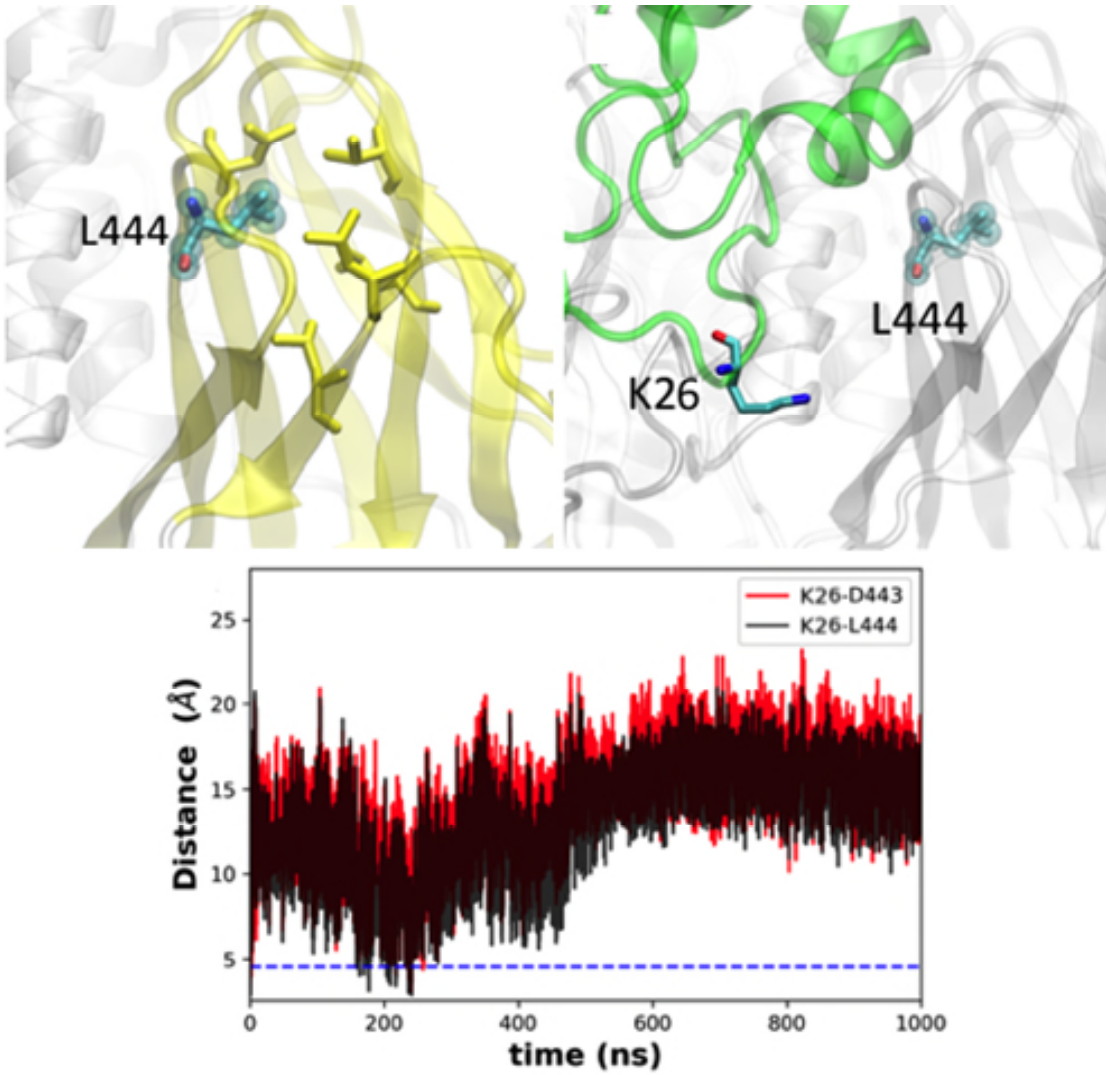


Figure S27: **Interactions of L444 in the N370S inactive complex (simulation 5b). Top (L):** The side chain of L444 sits in a hydrophobic pocket formed between the two β-sheets of Domain II. Residues A446, V460, V468, L470, I60, L65 form a part of this pocket. (**R**) In the inactive conformation, L444 does not form interactions with the side chain of K26 in SAPC. This interaction is disrupted along with the other interactions that the side chain of K26 makes with the surrounding residues. **Bottom:** Distance between residues L444 (bb), D443 (sc) and residues K26 (sc) of SAPC, in simulation 5b (N370S mutant inactive complex).

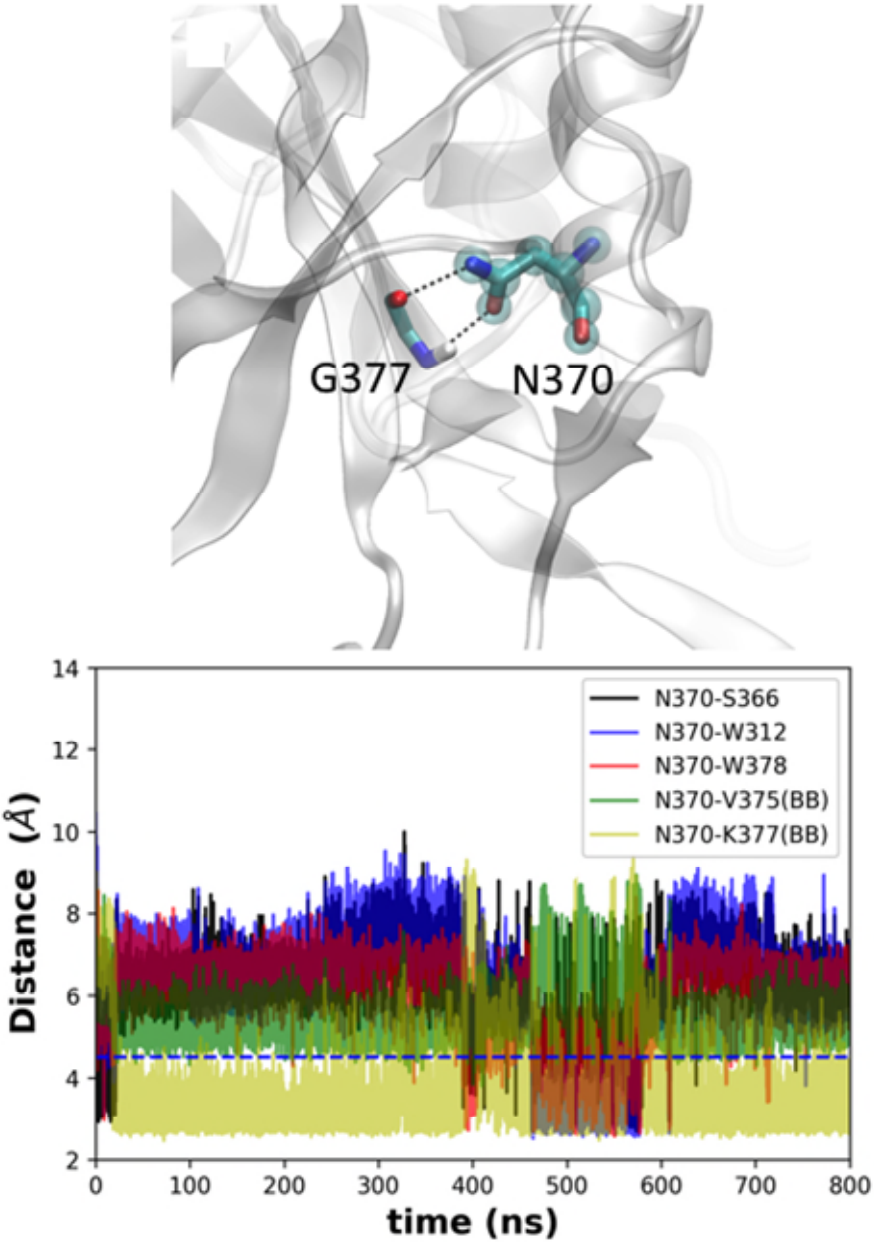


Figure S28: **Interactions of N370 in the mutant L444P active complex (simulation 6a).** In the active conformation, the side chain of N370 interacts with the backbone of residues G377 and V375 in the β strand 7 via hydrogen bonds. At ~500 ns, the side chain of N370 flip and establish hydrogen bonds with residues W312, S366 and W378. However, these interactions are transitional and last only for ~100 ns. **Top:** Snapshot of interaction between N370 and G377 in simulation 6a (L444P mutant active complex) at 1000ns. **Bottom:** Distance between residue N370 and S366, W312, W378, V375 (bb) and K377 (bb) in simulation 6a.

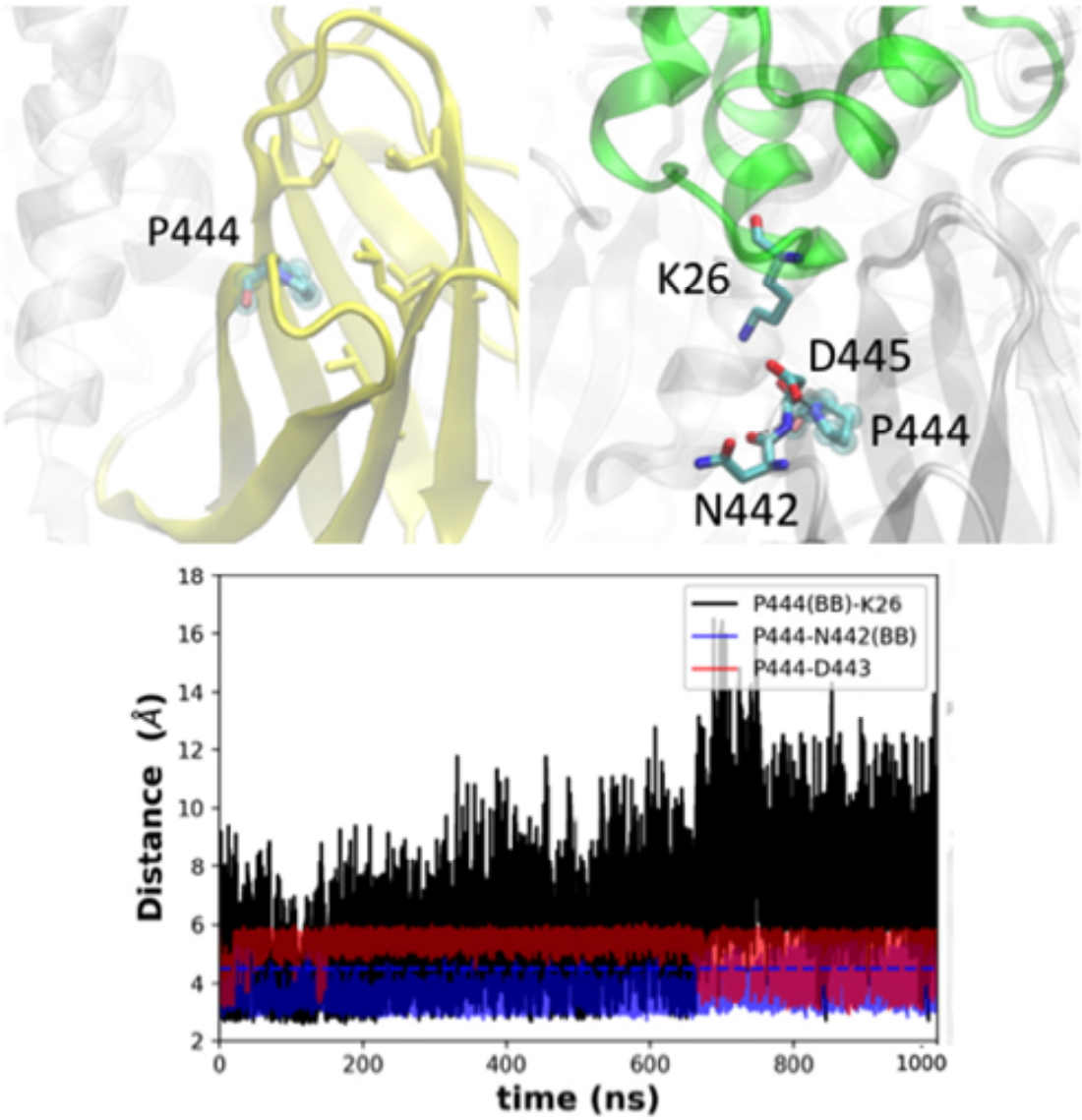


Figure S29: **Interactions of P444 in the mutant L444P active complex (simulation 6a). Top (L):** P444 sits inside hydrophobic cluster in the middle of the two β-sheets of Domain II. (**R**) Snapshot of interactions between P444 (bb) and K26 (sc) of SAPC, N442 (bb) and 443 at 1000 ns. The interactions between P444 and K26 of SAPC, which is present in the rest of the simulations, is disrupted in simulation 6a from ~600 ns. **Bottom:** Distance between residue P444 (bb) and residues K26 (sc) of SAPC, N442 (bb) and D443 (sc), in simulation 6a (L444P mutant active complex).

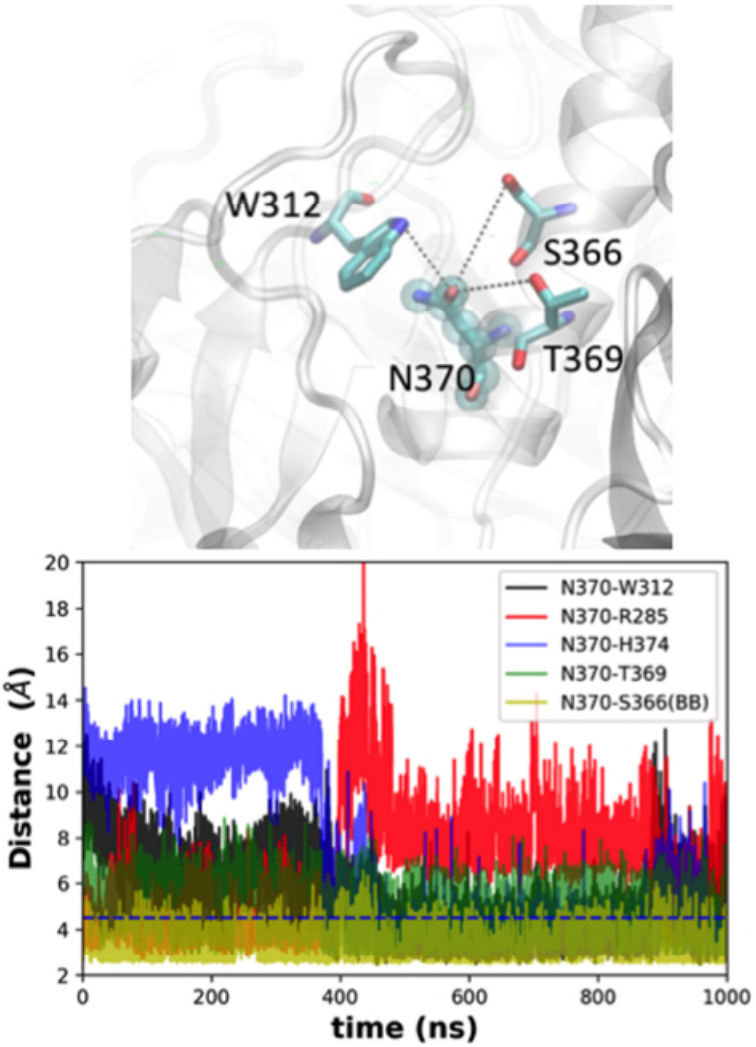


Figure S30: **Interactions of N370 in the mutant L444P inactive complex (simulation 6b).** In the inactive conformation, the side chain of N370 interacts with S366 and T369 in the same helix. N370 forms interactions with R285 until ~400ns, following which it gets disrupted. From 400 ns onwards, additional interactions are formed between W312 in Loop-1 and H374, positioned in the loop connecting helix-7 and β strand 7. **Top:** Snapshot of interaction between N370 and S366, T369 and W312 in simulation 6b (L444P mutant inactive complex) at 1000ns. **Bottom:** Distance between residue N370 and R285, W312, S366, T369 and H374 in simulation 6b.

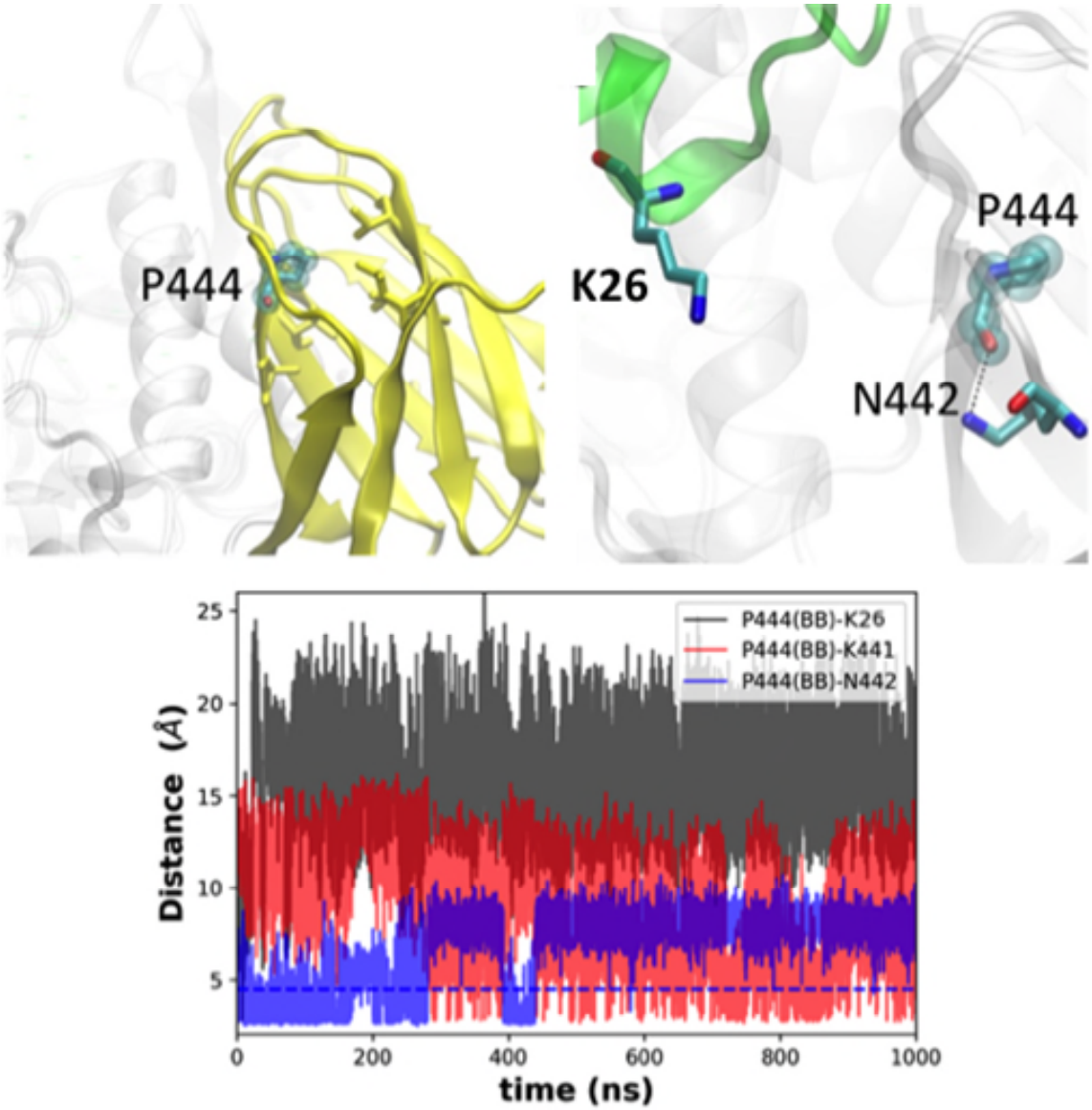


Figure S31: **Interactions of P444 in the mutant L444P inactive complex (simulation 6b). Top (L):** L444 sits in a hydrophobic cluster in the middle of the two β-sheets of Domain II. Mutation to P444 disrupts the hydrophobic pocket in simulation 6b. (**R**) Snapshot of interactions between P444 (bb) and K26 (sc) of SAPC and N442 at 1000 ns. The interaction between P444 and K26 of SAPC, which is present in the rest of the simulations, is disrupted in simulation 6b. The backbone of P444 forms direct interactions with N442 in the first half of the simulation, after which it makes weak interaction with K441. **Bottom:** Distance between residue P444 (bb) and residues K26 (sc) of SAPC, K441 and N442, in simulation 6b (L444P mutant inactive complex). Interaction with K26 of SAPC does not occur in simulation 6b.

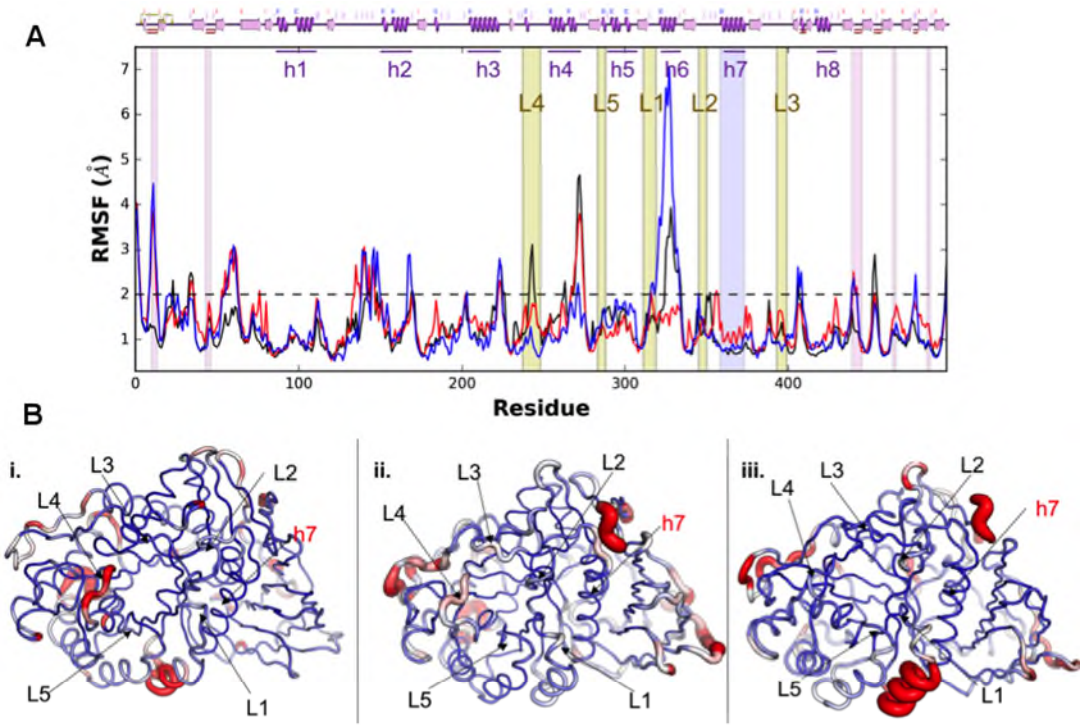


Figure S32: (**A**) Comparison of RMSF (GCase) as a function of each residue in simulations 3a, 5a and 6a. Loops-1 to -5 at the entrance of the binding site have been highlighted in yellow and tagged with the label L1 to L5, helix-7 has been highlighted in blue and the protein-protein binding site, other than Loops 1 and 2 and helix-7, has been highlighted in magenta. The secondary structure of GCase is annotated on top of the graph and helices (h1-h8) of the TIM barrel have been labeled. (Black – active complex, simulation 3a; Red – $GCase^{N370S}$ active state, simulation 5a; Blue – $GCase^{L444P}$ active state, simulation 6a) (**B**) RMSF values from the simulation translated on the structure of GCase in simulation (i) 3a (active complex), (ii) 5a ($GCase^{N370S}$ active state) and (iii) 6a ($GCase^{L444P}$ active state).

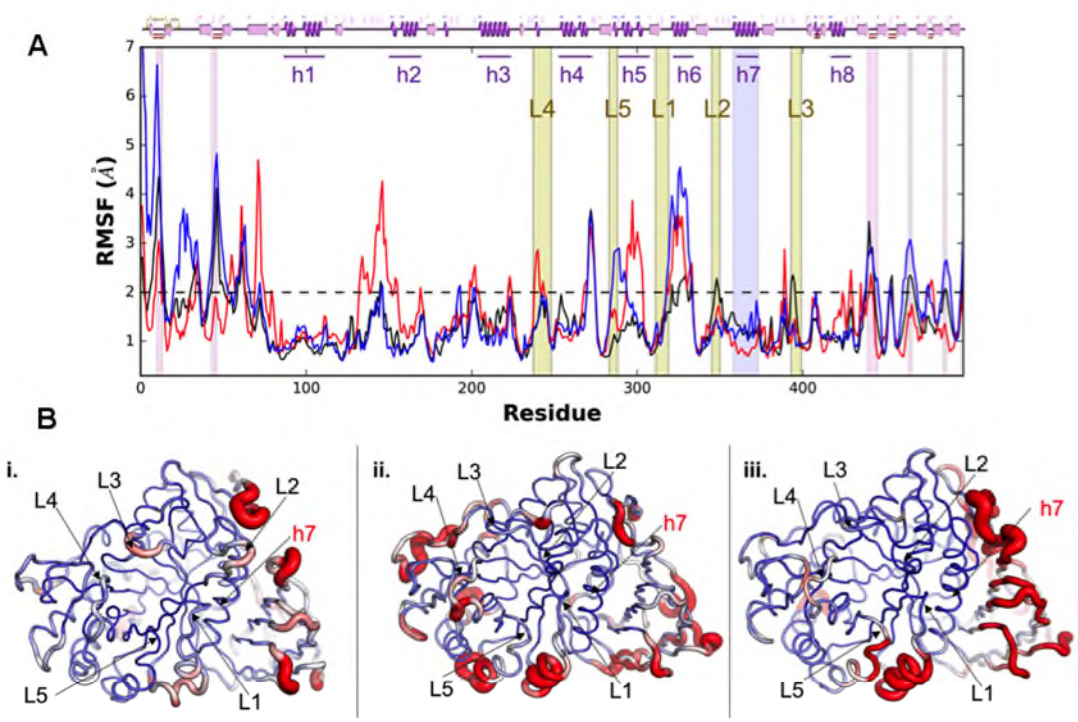


Figure S33: (**A**) Comparison of RMSF (GCase) as a function of each residue in simulations 3b, 5b and 6b. Loops-1 to -5 at the entrance of the binding site have been highlighted in yellow and tagged with the label L1 to L5, helix-7 has been highlighted in blue and the protein-protein binding site, other than Loops 1 and 2 and helix-7, has been highlighted in magenta. The secondary structure of GCase is annotated on top of the graph and helices (h1-h8) of the TIM barrel have been labeled. (Black – inactive complex, simulation 3b; Red – $GCase^{N370S}$ inactive state, simulation 5b; Blue – $GCase^{L444P}$ inactive state, simulation 6b) (**B**) RMSF values from the simulation translated on the structure of GCase in simulation (i) 3b (inactive complex), (ii) 5b ($GCase^{N370S}$ inactive state) and (iii) 6b ($GCase^{L444P}$ inactive state)

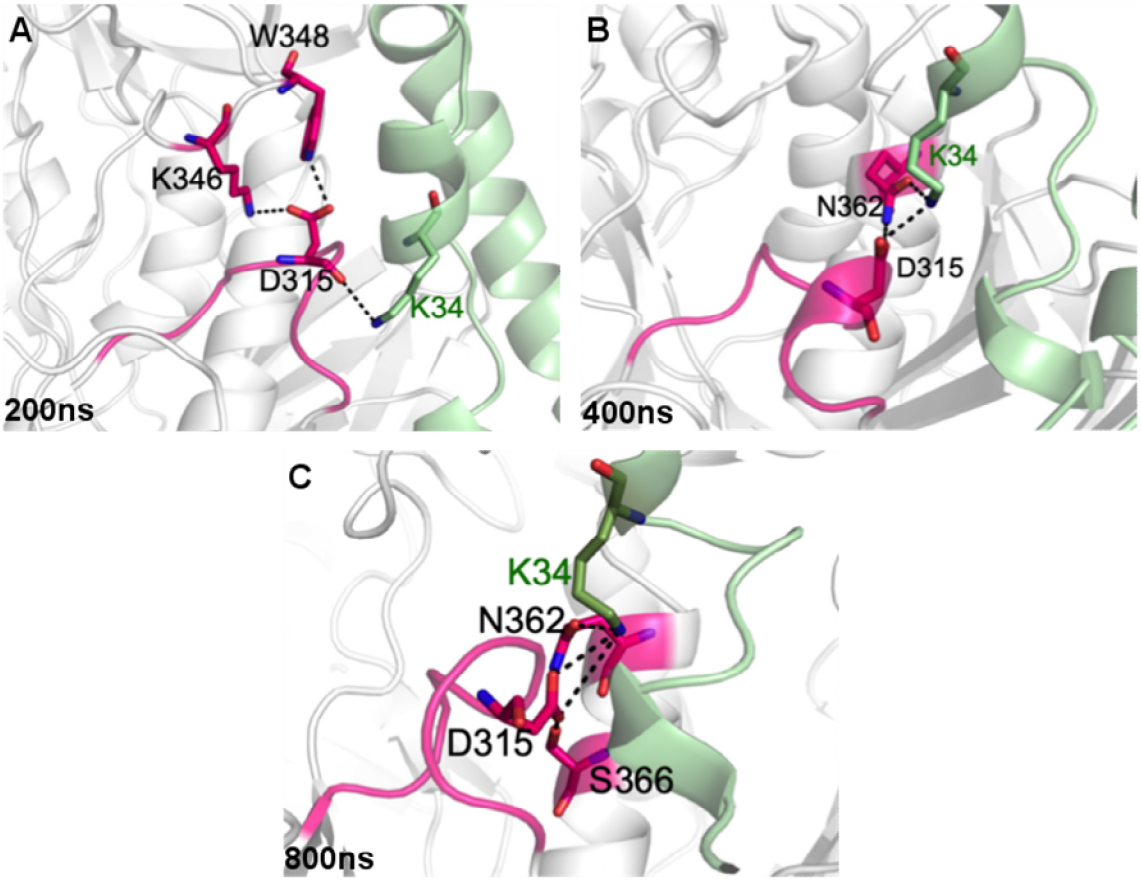


Figure S34: **Loop-1 dynamics in L444P active complex (simulation 6a)** at (**A**) 200ns, (**B**) 400ns and (**C**) 800ns. At the beginning of the simulation, D315 is extended towards Loop-2 (a characteristic of Loop-1 of the inactive conformation). As the simulation progresses, D315 begins to make interactions with K34 of SAPC, which forces Loop-1 to orient towards helix-7, a characteristic of the active conformation. GCase is colored white with the interacting residues in pink and SAPC is colored green.

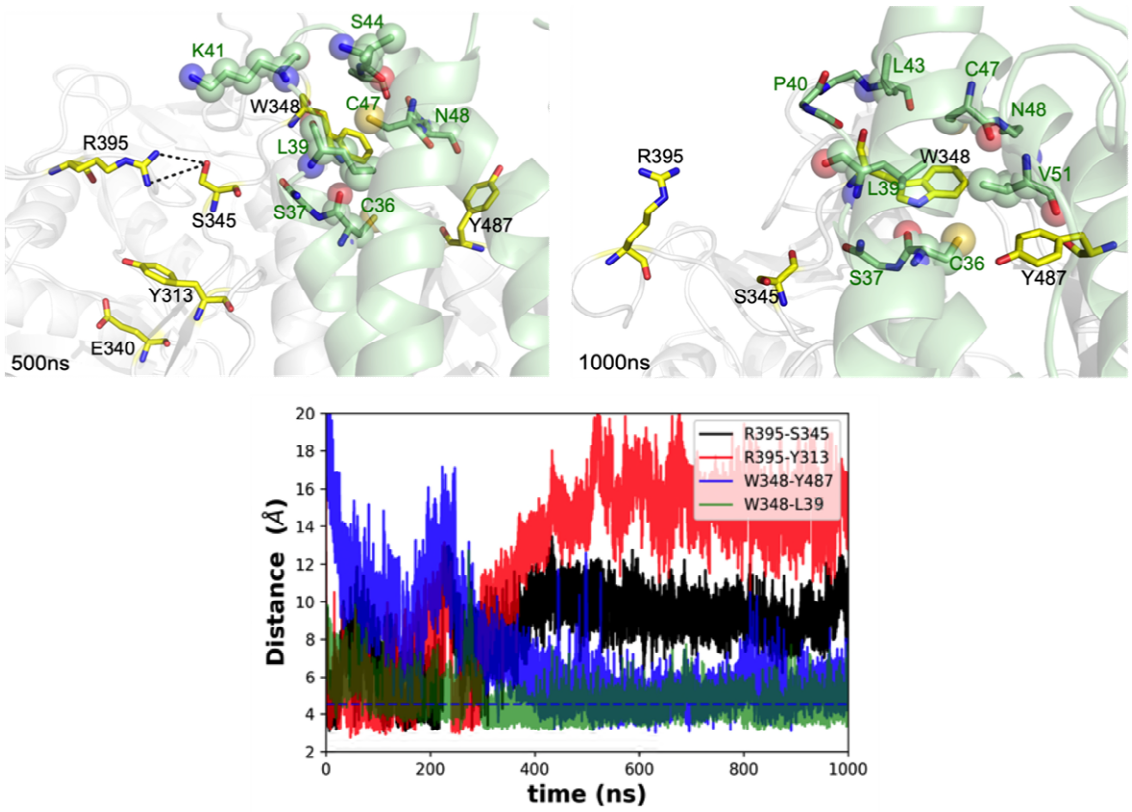


Figure S35: **Stabilization of W348 inside the hydrophobic pocket formed by SAPC**. This coincides with the opening of Loop-2 in simulation 3b (GCase inactive complex). **Top**: Interactions of residues W348 and R395 at (**L**) 500ns and (**R**) 1000ns. GCase has been illustrated in white, with interacting residues in yellow and SAPC in green. **Bottom:** Distance between pair of residues R395-S345, R395-Y313, W348-Y487, W348-L39 in simulation 3b (GCase inactive complex).

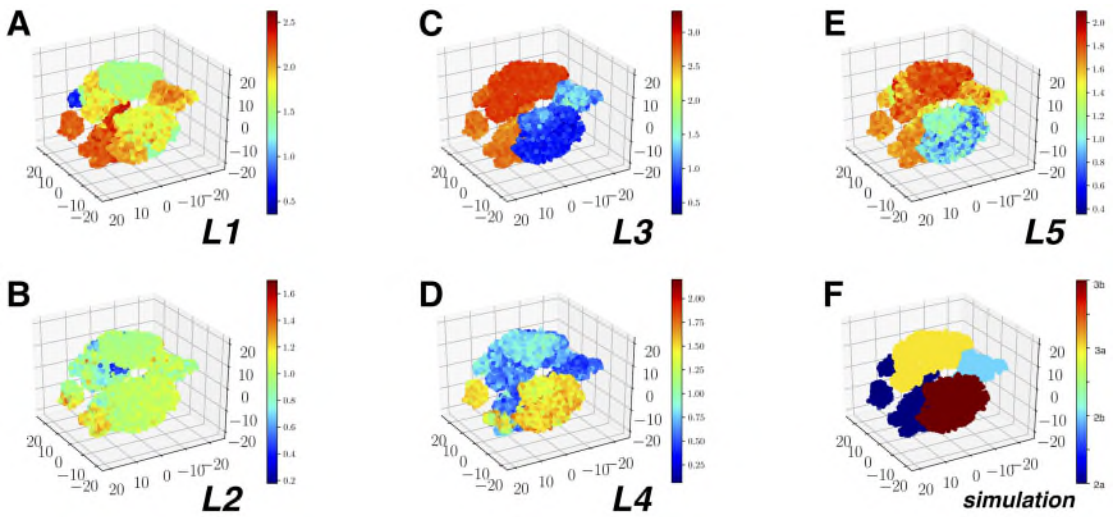


Figure S36: **Deep clustering of WT AT-MD simulations (2a, 2b, 3a and 3b from Table 1) of GCase reveals conformational states sampled based on loop conformations (L1-L5). (A-E)** Each conformation in the 3D t-SNE space (see Methods section) is painted using the root-mean squared deviations (RMSD) of the Cα atoms from L1-L5 respectively. Notably, distinct conformational states can be identified with respect to the RMSD variations for each of the five loops and the simulations themselves **(F)**.

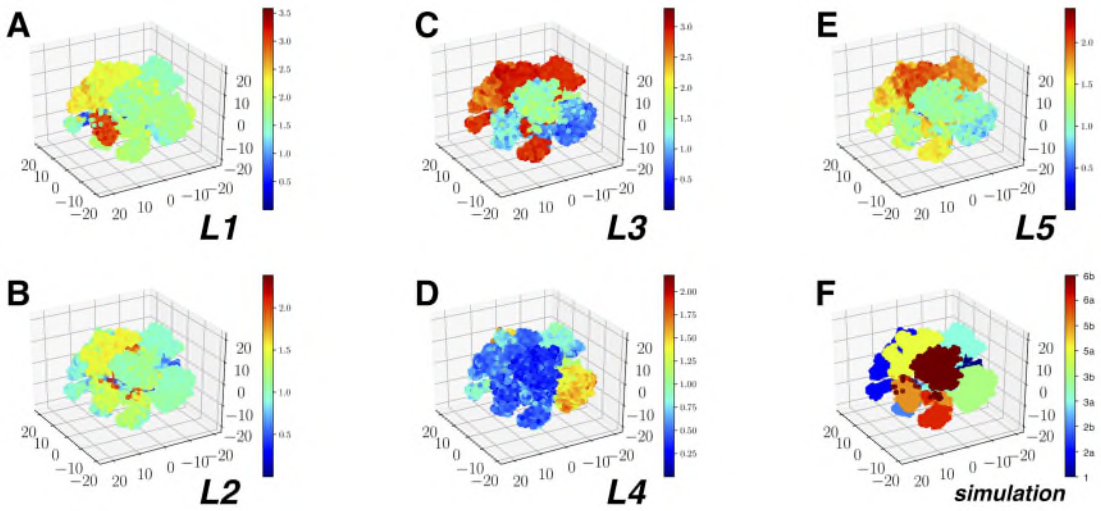


Figure S37: **Projecting conformations from WT- and mutant- AT-MD simulations using the deep clustering approach GCase reveals conformational states sampled based on loop conformations (L1-L5). (A-E)** Each conformation in the 3D t-SNE space (see Methods section) is painted using the root-mean squared deviations (RMSD) of the Cα atoms from L1-L5 respectively. Notably, distinct conformational states can be identified with respect to the RMSD variations for each of the five loops and the simulations themselves **(F)**.

Table S1: Summary of the interacting residues in the selected models and relevant electrostatic interactions for selected poses of GCase and closed and open conformation of SAPC.

MODEL	GCase Residues	SAPC Residues	HB and Ion Pairs
Closed SAPC	F316-P319, K321, K346, W348, R353, D358, Q362, H365, T369, Y373, K441, D443, R463-S465, Y487	K26, E27, L29, A31, K34, M35, K38, S42, E45, E46, E49, T53, Y54, I58, V71	K26-T369, D30-H365, K34-D315, Q48-D358, E25-K441
Open SAPC	D315-K321, K346-E349, R353, D358, Q362, H365, Y373, K441, D443-D445, R463-S465, Y487	K26, E27, L29, A31, K34, M35, K38, S42, E45, E46, E49, T53, Y54, I58, V71	Q48-W348, K26-Q362, D30-H365, E25-K441, S59-D443, N22-D445, S57- K466