Electronic Supplementary Material (ESI) for Chemical Communications.

Integrating Molecular Probes and Molecular Dynamics to Reveal Binding Modes of GLUT5 Activatory and Inhibitory Ligands

Jon Ainsley[a, &], Shobhit S. Chaturvedi[b, &], Tatyana G. Karabencheva-Christova[b, *], Marina Tanasova[b, *] and Christo Z. Christov[b, *]

[a.] Department of Applied Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, United Kingdom.

[b.] Department of Chemistry, Michigan Technological University, Houghton, Michigan 49931, United States.

[&.] J.A. and S.S.C. have equal contributions to this article.

[*] TGKC, MT and CZC are corresponding authors

SUPPLEMENTARY INFORMATION

Table of Contents

1. Supplementary Figures

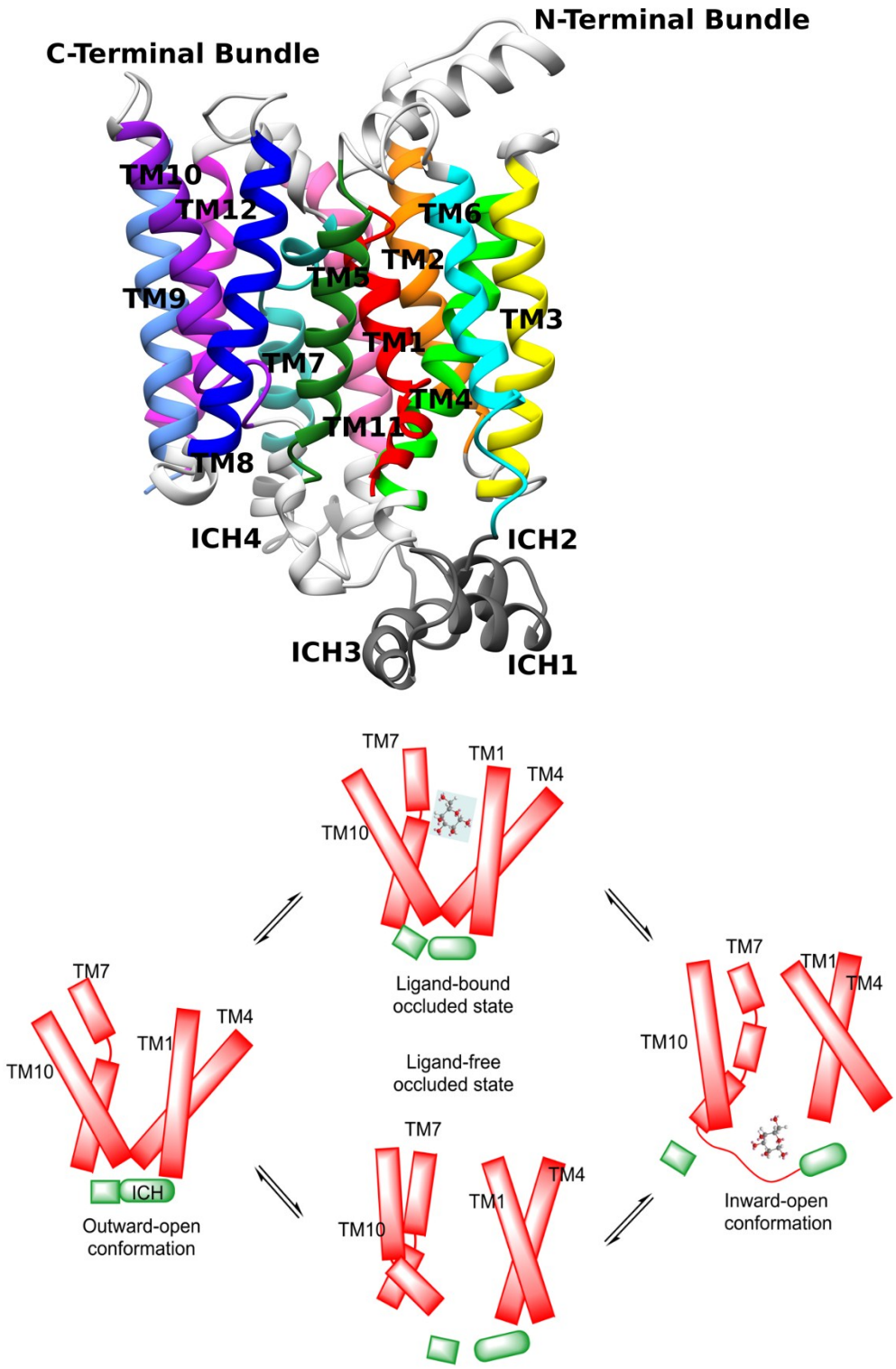

Fig. S1. Topology of GLUT5 and the mechanism explaining the passage of sugar through the protein by a conformational transition between outward-open and inward-open conformations upon substrate binding.

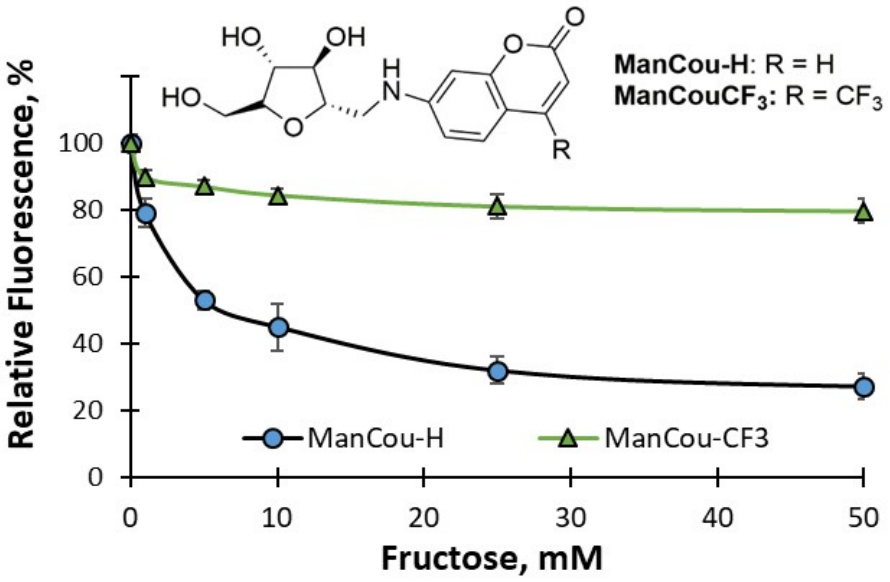

Fig. S2. Competition for the uptake through GLUT5 between fructose and GLUT5-specific ManCou probes. Data obtained with MCF7 breast cancer cells with 20 μM ManCou probes and varied concentrations of fructose in 96-wellplate format (ex. 405 nm, em. 465 nm, 10 min incubation at 37 °C) according to standard procedure described in section 2.1.

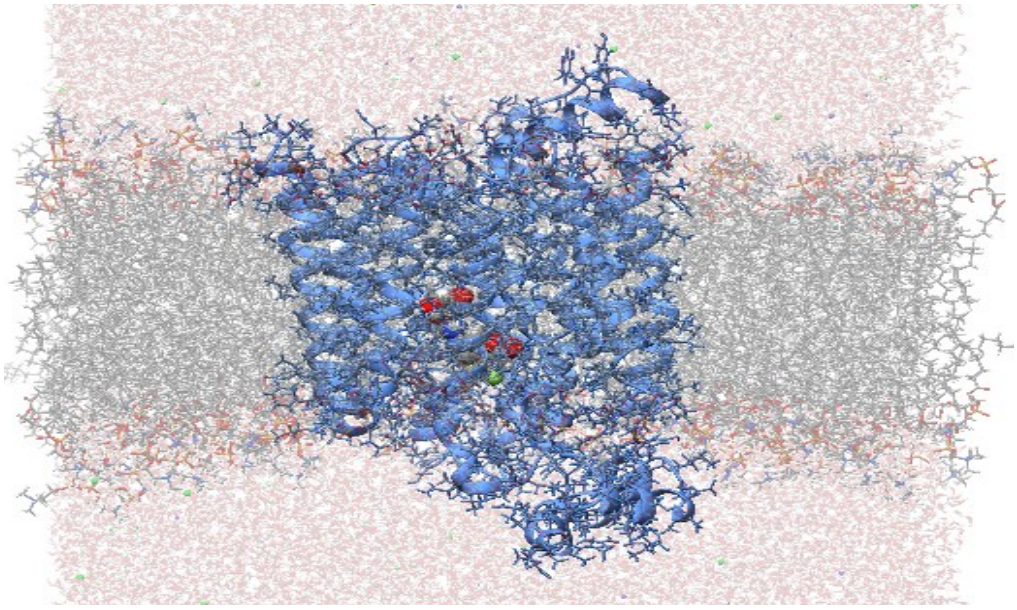

Fig. S3. Simulation system. ManCou-CF_3 bound to GLUT5 in lipid membrane surrounded by water molecules (solvent) and ions.

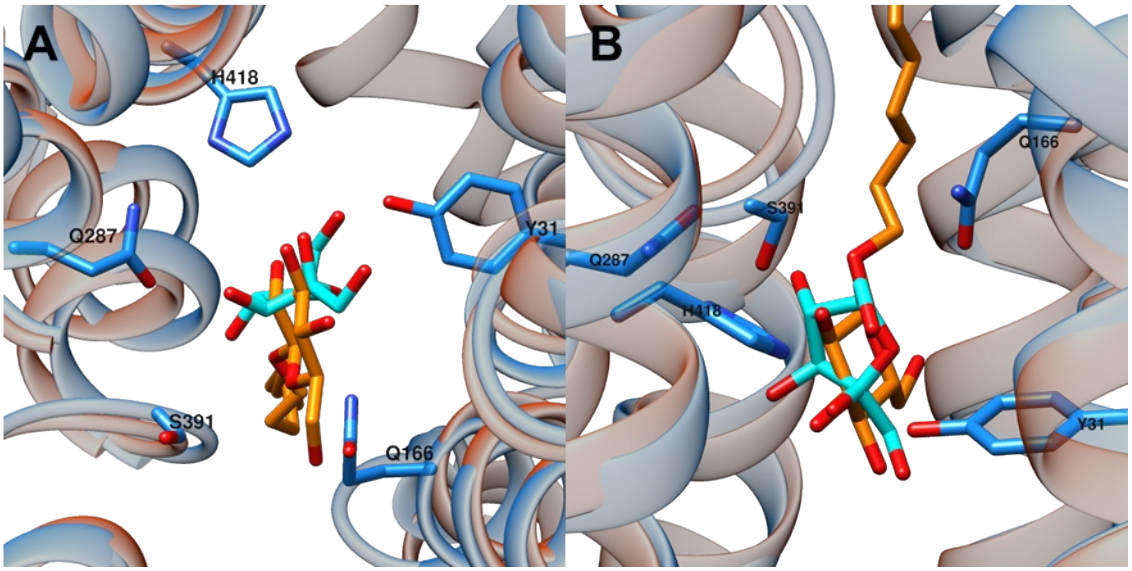

Fig. S4. The overlaid structures of GLUT1 (PDBID: 4PYP) is rendered with transparent red ribbons, its ligand n-nonyl-β-D-glucopyranoside is rendered with tubes and orange colored carbon atoms, the protein structure of GLUT5 (PDBID: 4YB9) is rendered with transparent blue ribbons, the docked ligand fructose is rendered with tubes and cyan colored carbon atoms. The residues of GLUT5 observed to make hydrogen bonding interactions with the docked fructose are labelled. A) The view from the extracellular side and B) Side view.

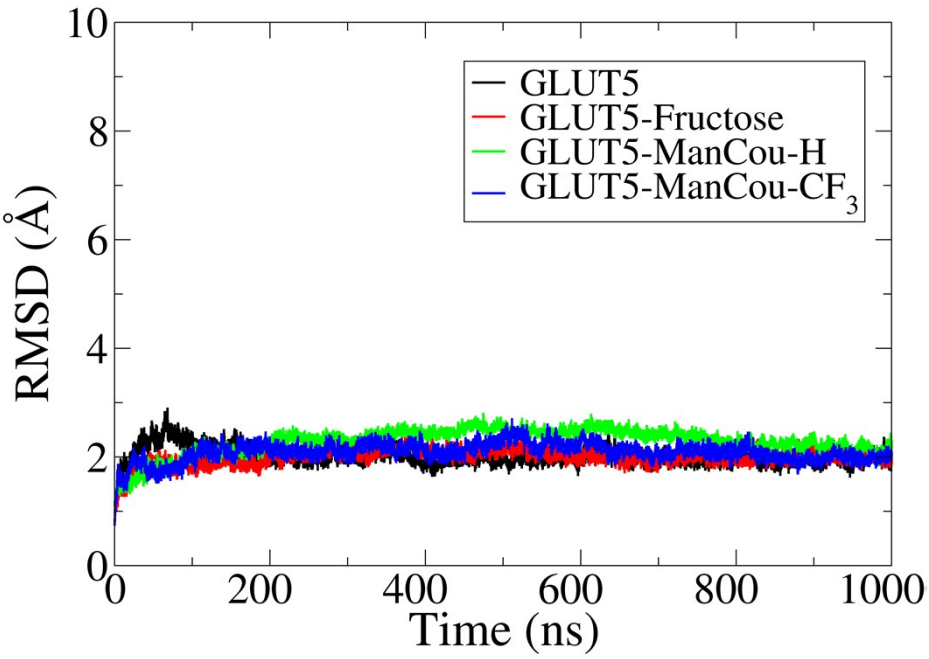

Fig. S5. RMSDs of (i) free Glut-5 (ii) GLUT5-fructofuranose complex (iii) GLUT5-ManCou-H complex (iii) GLUT5-ManCou-CF_3 complex. The plots indicate about stable simulations.

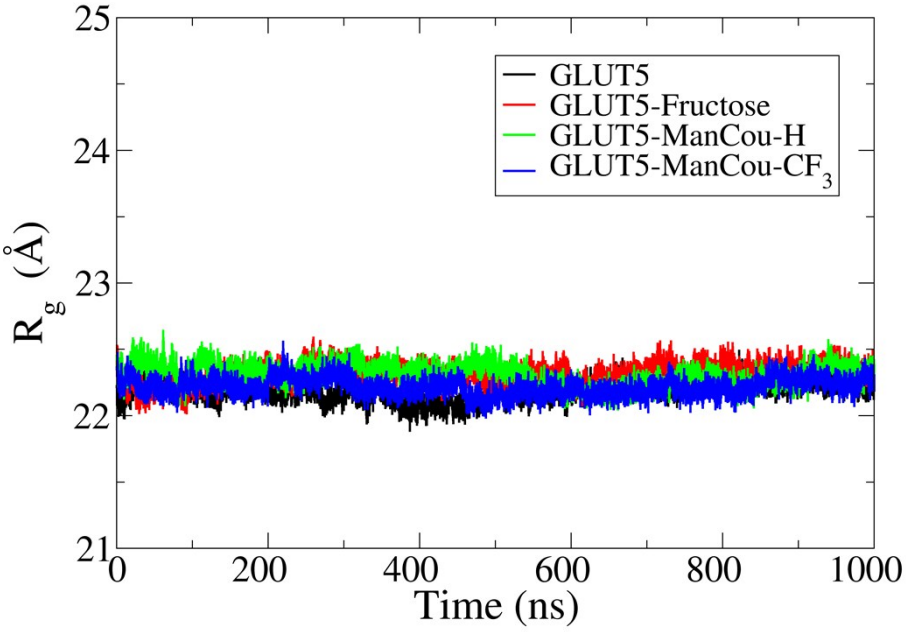

Fig. S6. Radius of Gyration of (i) free Glut-5 (ii) GLUT5-fructofuranose complex (iii) GLUT5-ManCou-H complex (iii) GLUT5-ManCou-CF_3 complex. The profiles suggest about stable simulations.

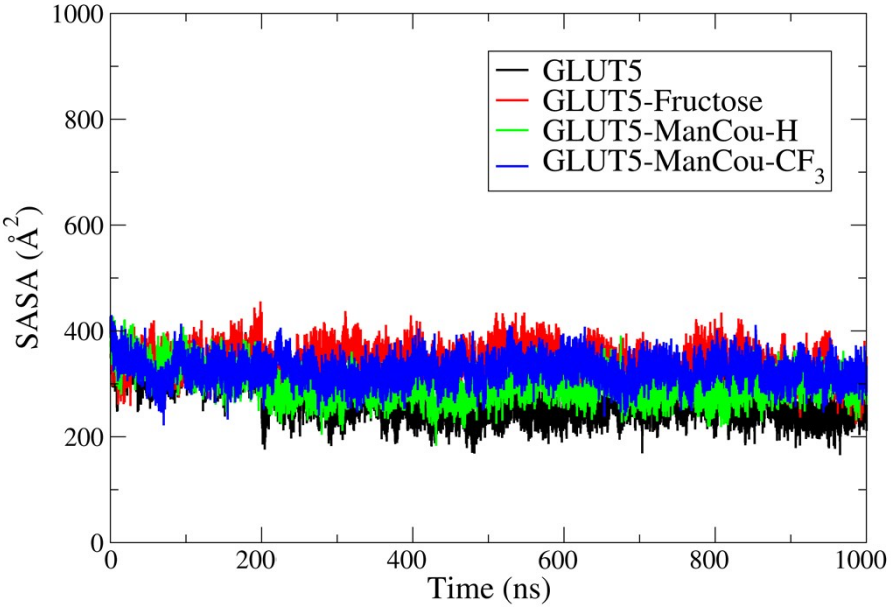

Fig. S7. Solvent Accessible Surface Area (SASA) for (i) free GLUT5, (ii) GLUT5-fructofuranose complex, (iii) GLUT5-ManCou-H complex, (iv) GLUT5-ManCou-CF_3 complex. The time fluctuations of SASA indicate about stable simulations.

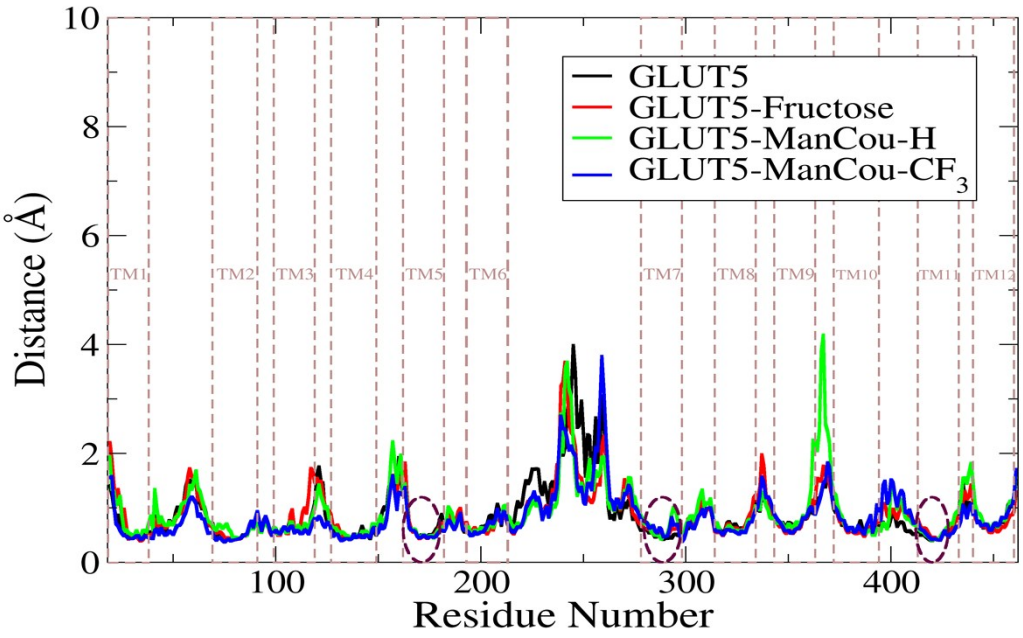

Fig. S8. RMSF of (i) free GLUT5 (ii) GLUT5-fructofuranose complex (iii) GLUT5-ManCou-H complex and (iv) GLUT5-ManCou-CF_3 complex. Dotted circles indicate binding site residues. Loop regions between the TM helices are characterized with higher fluctuations. Lower fluctuations can be found for the GLUT5 complex with ManCou-CF_3 in the loops between TM3 and TM4, extracellular region between TM1 and TM2, the cytoplasmic region betweenTM4 and TM5, and an extracellular region connecting TM11 and TM12.

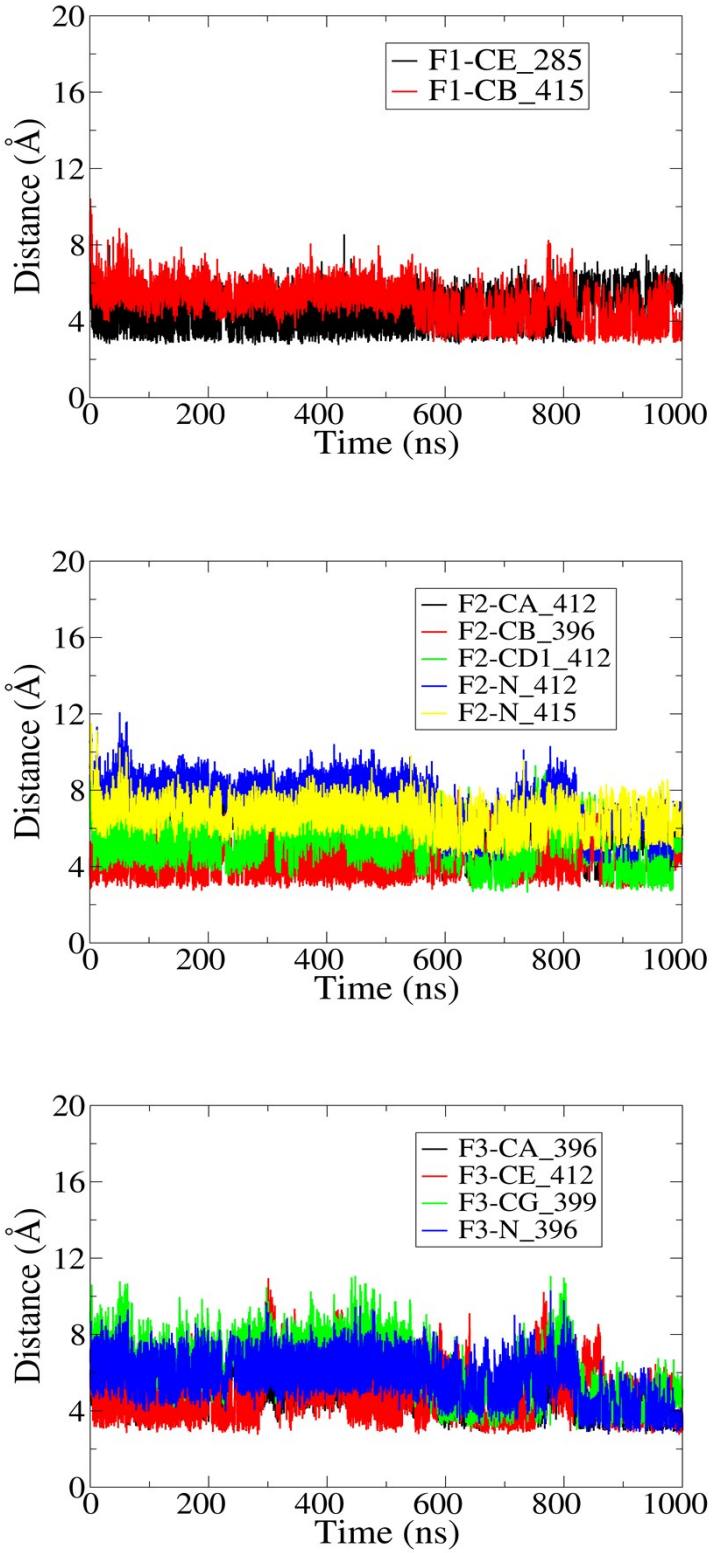

Fig. S9. Interactions of the three fluorine atoms of the CF_3 group of ManCou-CF_3 with GLUT5 residues. The plots suggest about the stability of the binding interactions between the fluorine atoms and the residues in the binding site.

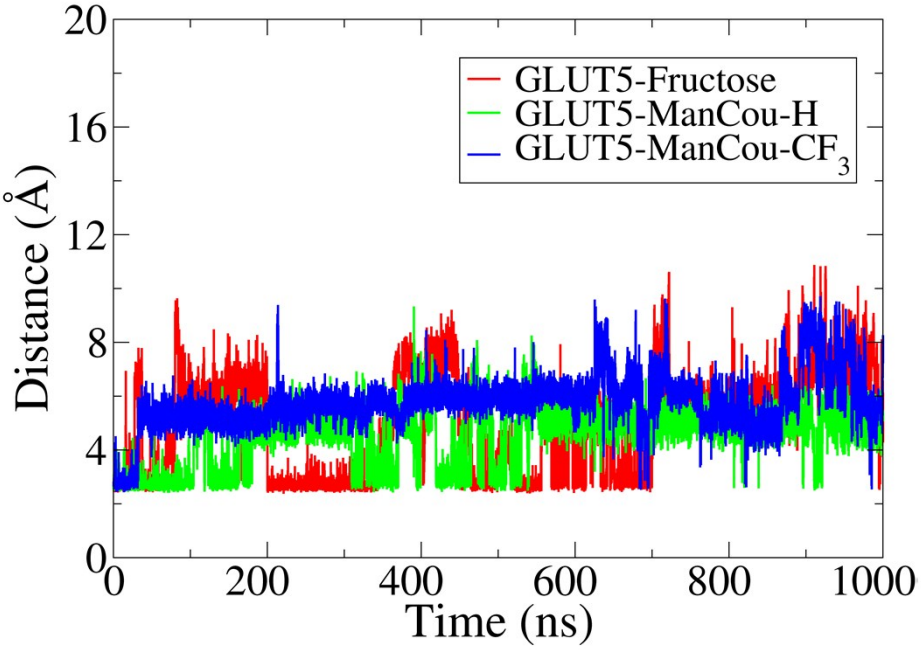

Fig. S10. Interactions between GLUT5 (Q167) and different ligands. The average distance in Mancou CF_3 is stabilized at larger value in respect to the other two ligands.

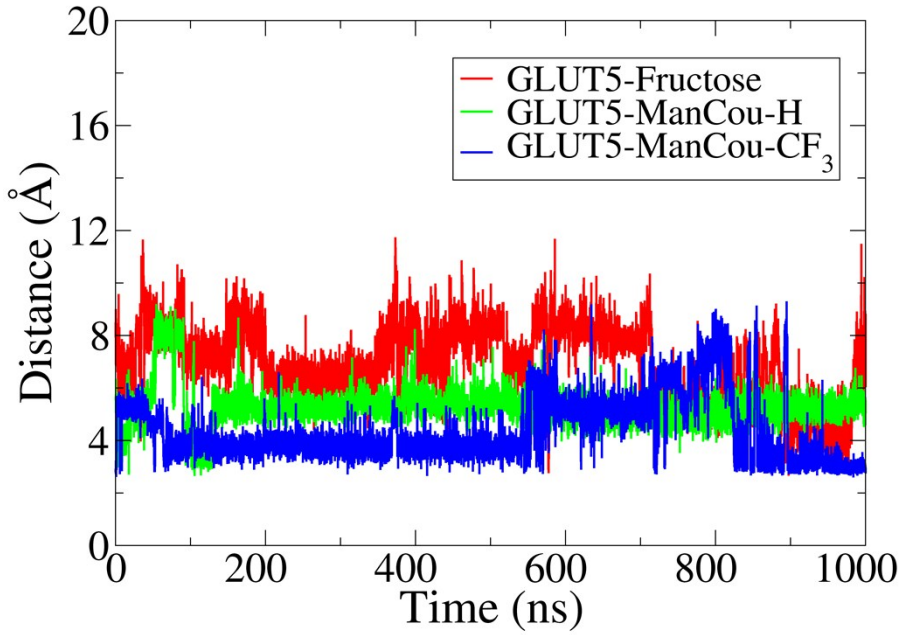

Fig. S11. Interactions between GLUT5 (N294) and different ligands as a function of the simulation time.

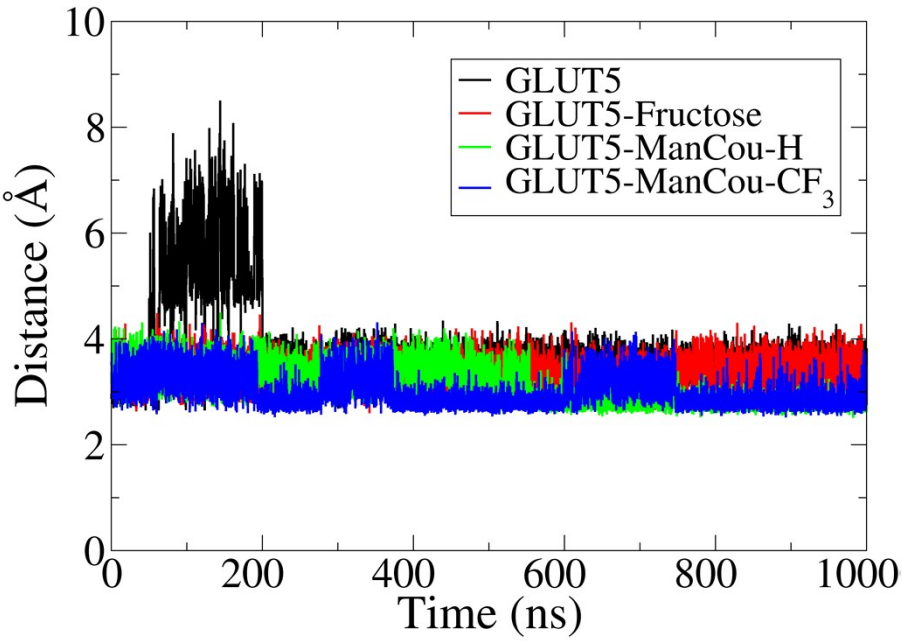

Fig. S12. Salt bridge interactions between E152 and R98 for free GLUT5, GLUT5-fructofuranose, GLUT5-ManCou-H and GLUT5-ManCou-CF_3 complexes. The distance between oxygen atom of E152 and nitrogen atom of R98 is plotted in this graph. This salt bridge is present in the inward-open conformation of GLUT5 (both crystal structure and MD simulations).

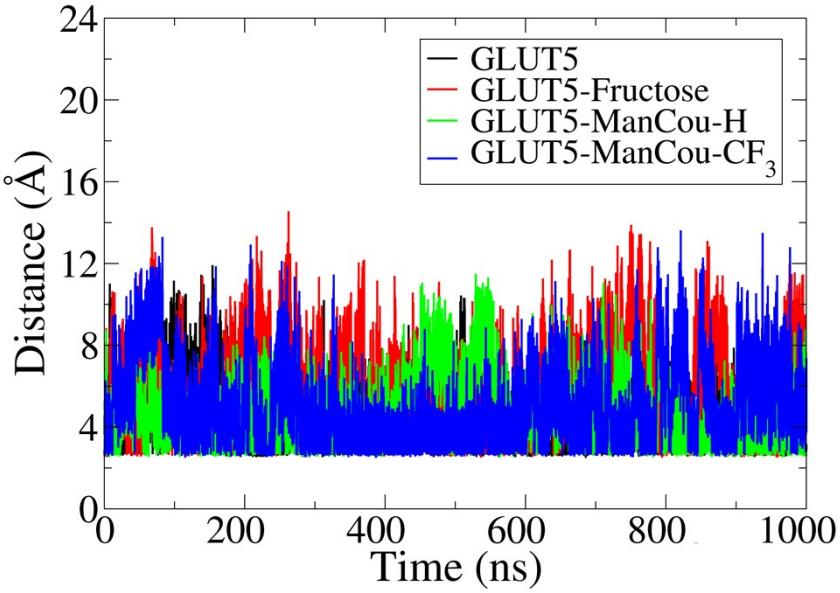

Fig. S13. Salt bridge interactions between residues E401 and R341 for free GLUT5, GLUT5-fructofuranose, GLUT5-ManCou-H and GLUT5-ManCou-CF_3 complexes. The distance between oxygen atom of E401 and nitrogen atom of R341 is plotted in this graph. This salt bridge is presented in the crystal structure of the inward-open conformation of GLUT5 and the MD simulations.

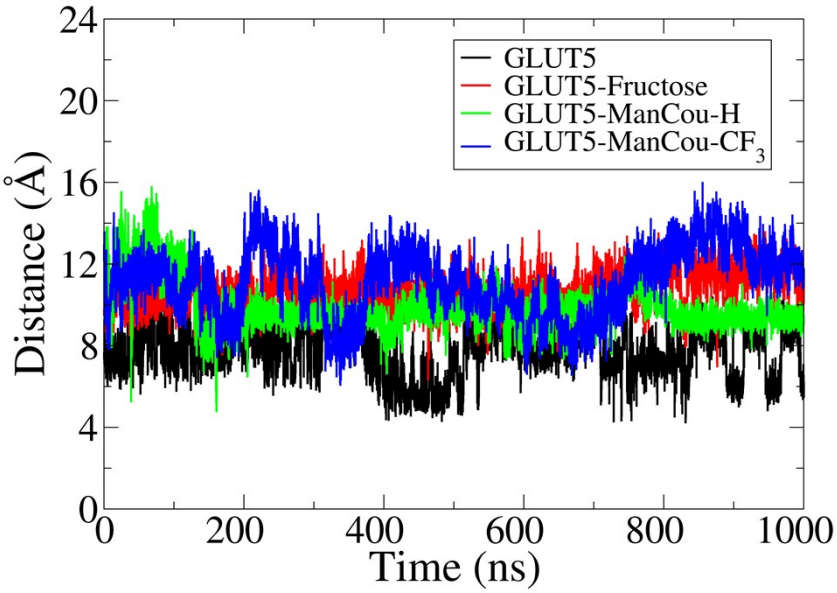

Fig. S14. Salt bridge interactions between residues E152 and R408 for free GLUT5, GLUT5-fructofuranose, GLUT5-ManCou-H and GLUT5-ManCou-CF_3 complexes. The distance between oxygen atom of E152 and nitrogen atom of R408 is plotted in this graph. The higher fluctuations in the complex with Mancou-CF3 might be related with the stabilization of the inward-occluded conformation.

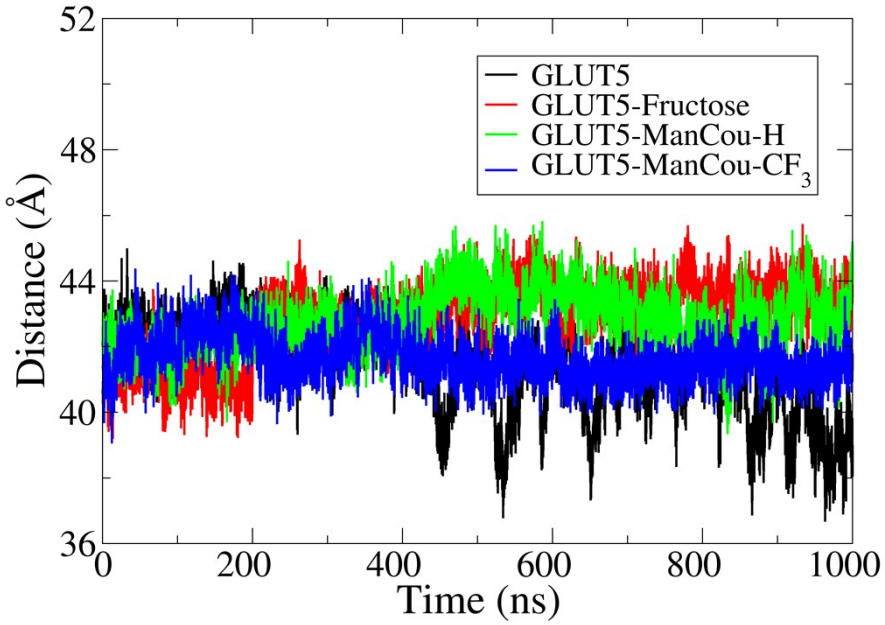

Fig. S15. Distance between the C-alpha atoms of the first residues of TM4 and TM10 (M127 and M372) for (i) Free GLUT5, (ii) GLUT5-Fructose complex (iii) GLUT5-ManCou-H complex (iv) GLUT5-ManCou-CF_3 complex. The reduced average distance in Mancou-CF_3 complex would indicate about stabilization of the inward-occluded conformation.

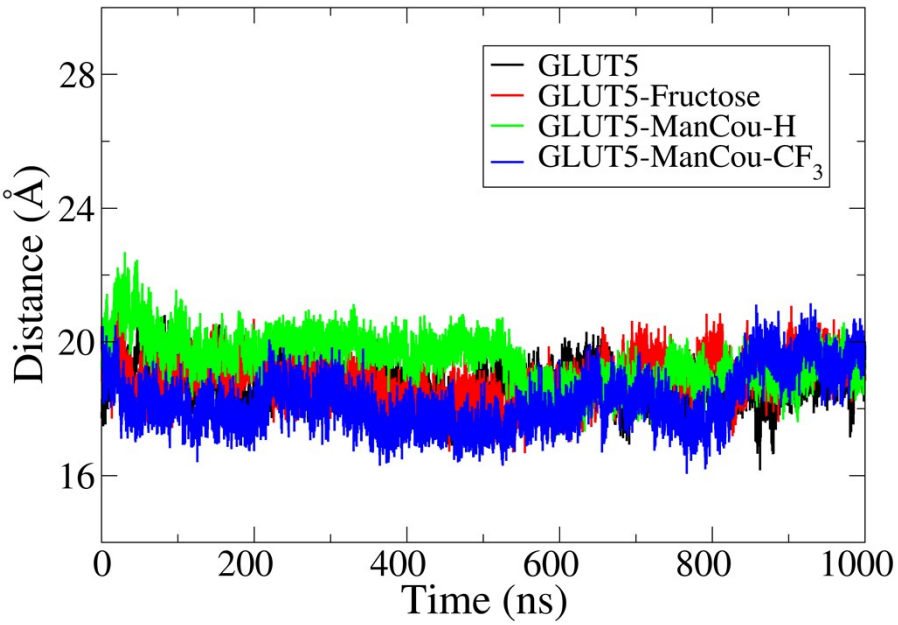

Fig. S16. Distance between the C-alpha atoms of the last residues of TM4 and TM10 (Y149 and I394) for (i) Free GLUT5, (ii) GLUT5-Fructose complex, (iii) GLUT5-ManCou-H complex, (iv) GLUT5-ManCou-CF_3 complex. The reduced average distance in Mancou-CF_3 complex would indicate about stabilization of the inward-occluded conformation.

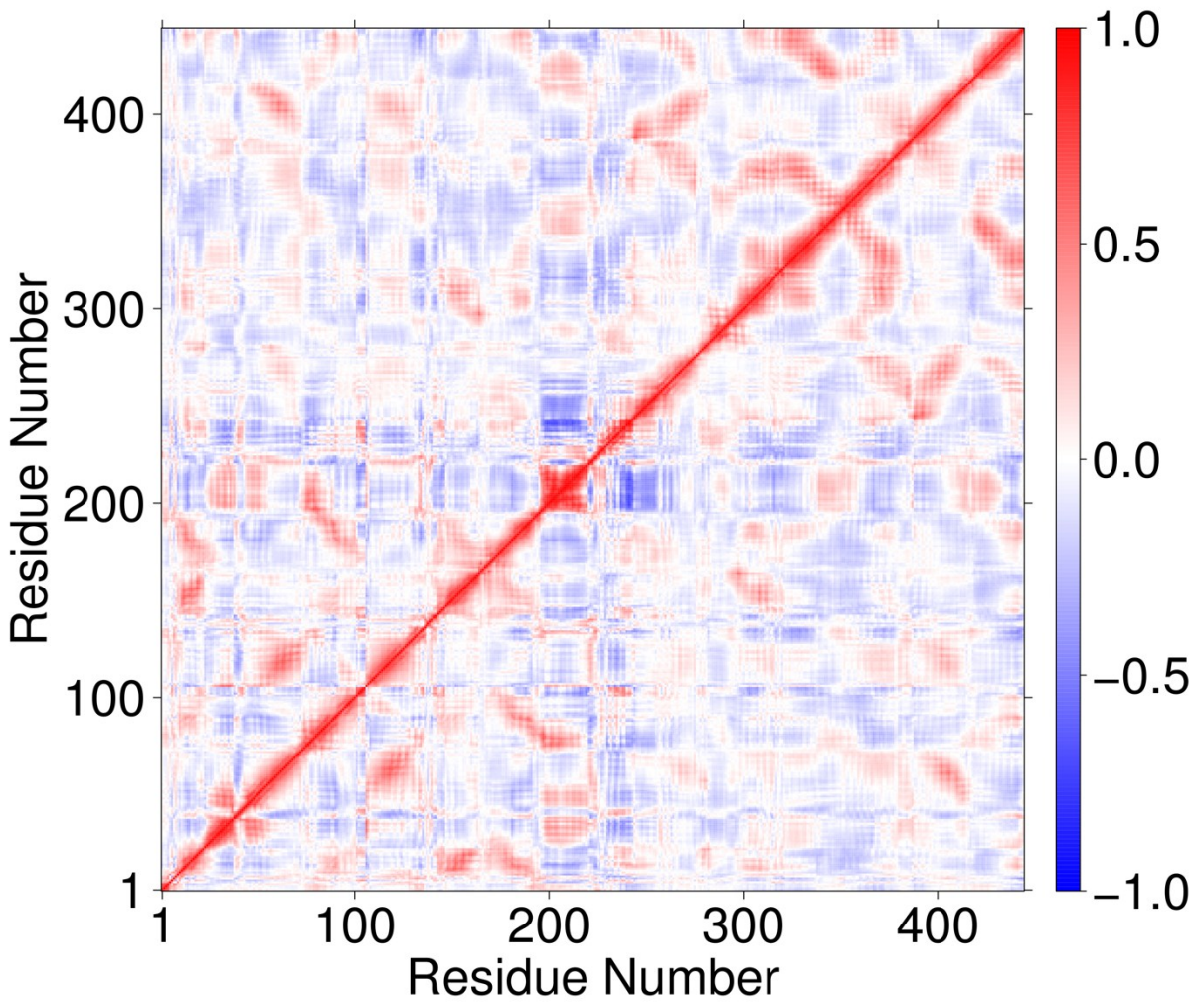

Fig. S17. Dynamic Cross Correlation for free GLUT5 based on C-alpha atoms (Note Residues labels are numbered from 1-444 instead of 18-462 as in crystal structure). Red color indicates positive correlation in the movement of residues and blue color indicates negative correlation in the movement of residues.

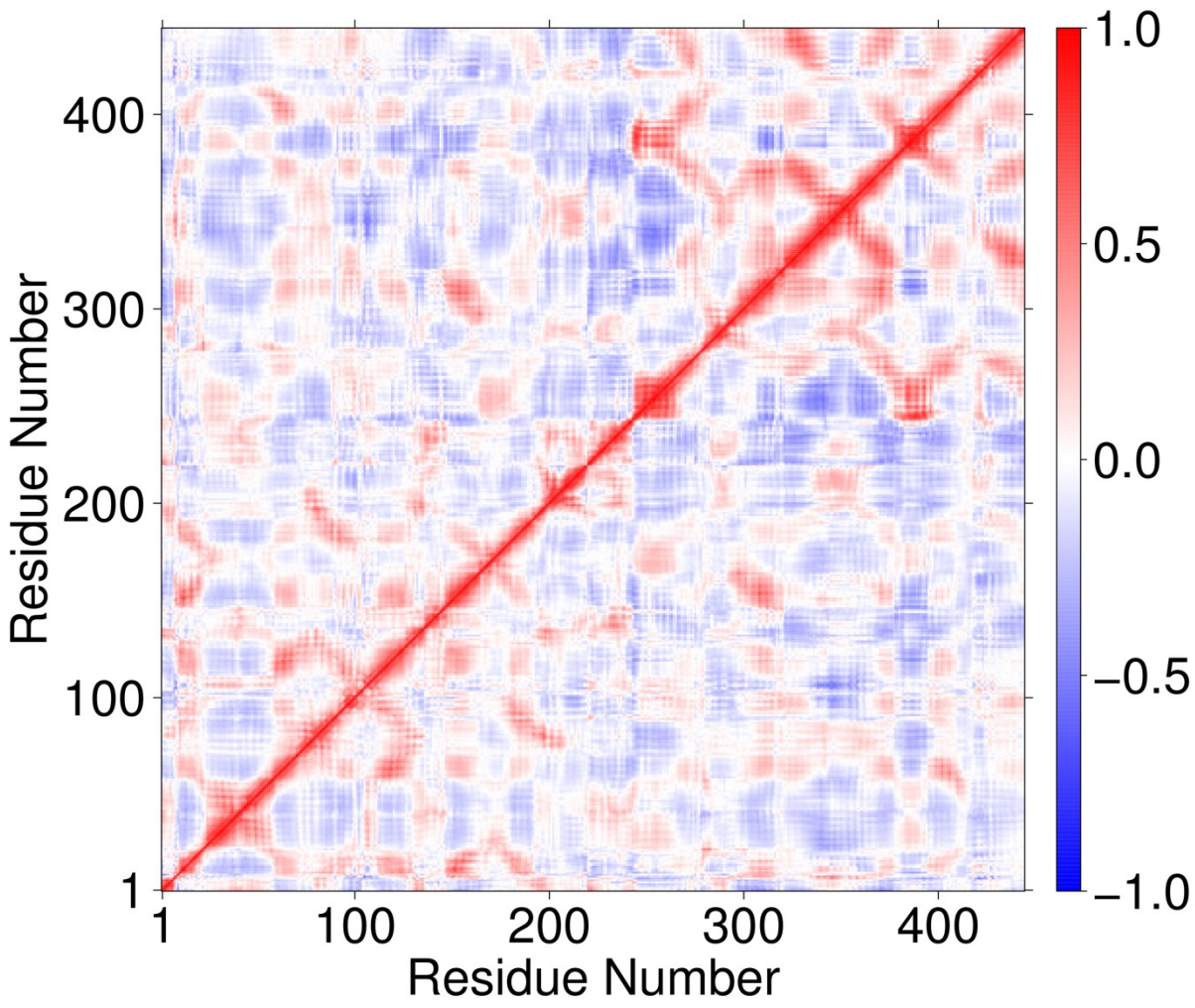

Fig. S18. Dynamic Cross Correlation for GLUT5-fructose complex based on C-alpha atoms (Note Residues labels are numbered from 1-444 instead of 18-462 as in crystal structure). Red color indicates positive correlation in the movement of residues and blue color indicates negative correlation in the movement of residues.

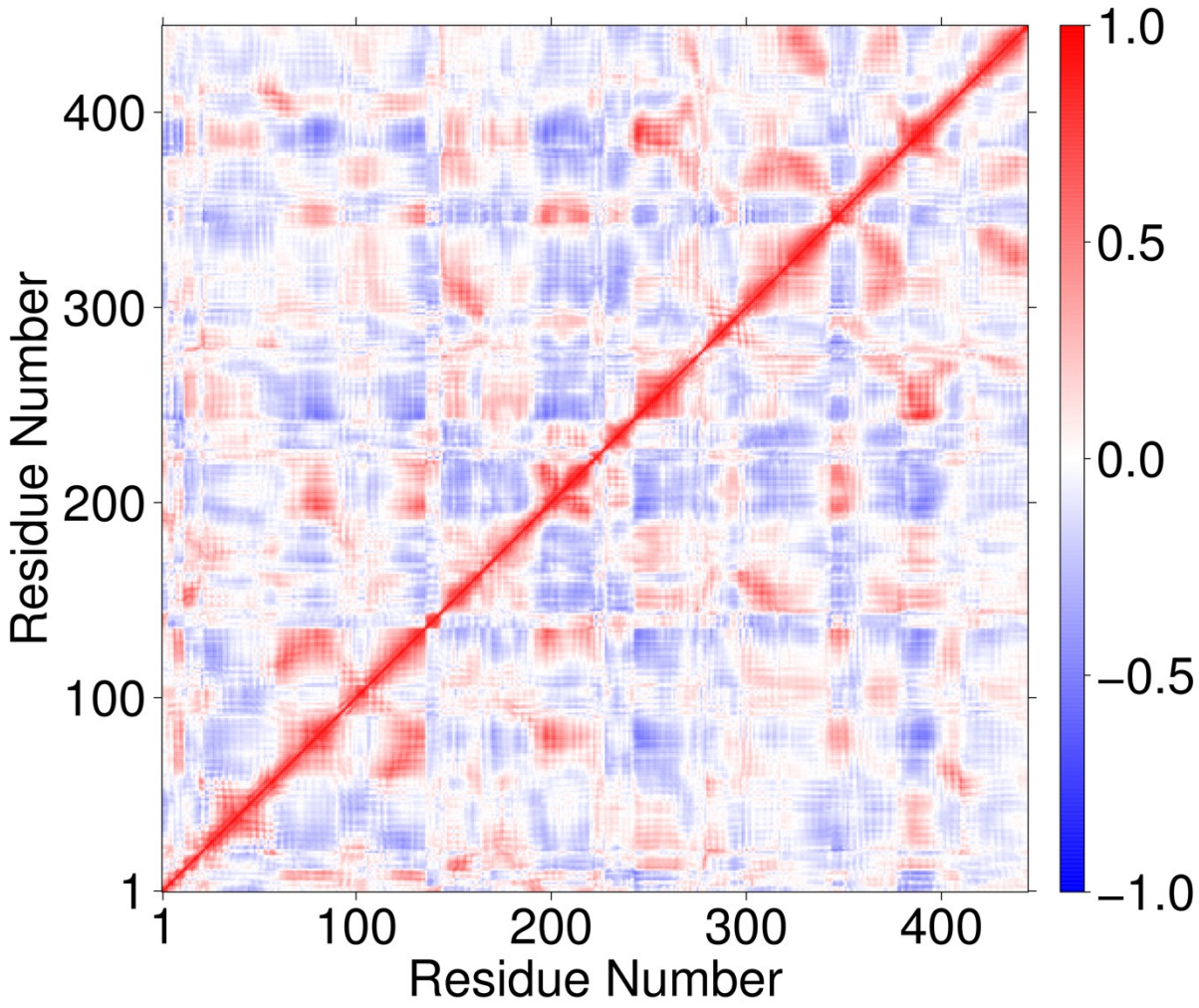

Fig. S19. Dynamic Cross Correlation for GLUT5-ManCou-H complex based on C-alpha atoms (Note Residues labels are numbered from 1-444 instead of 18-462 as in crystal structure). Red color indicates positive correlation in the movement of residues and blue color indicates negative correlation in the movement of residues.

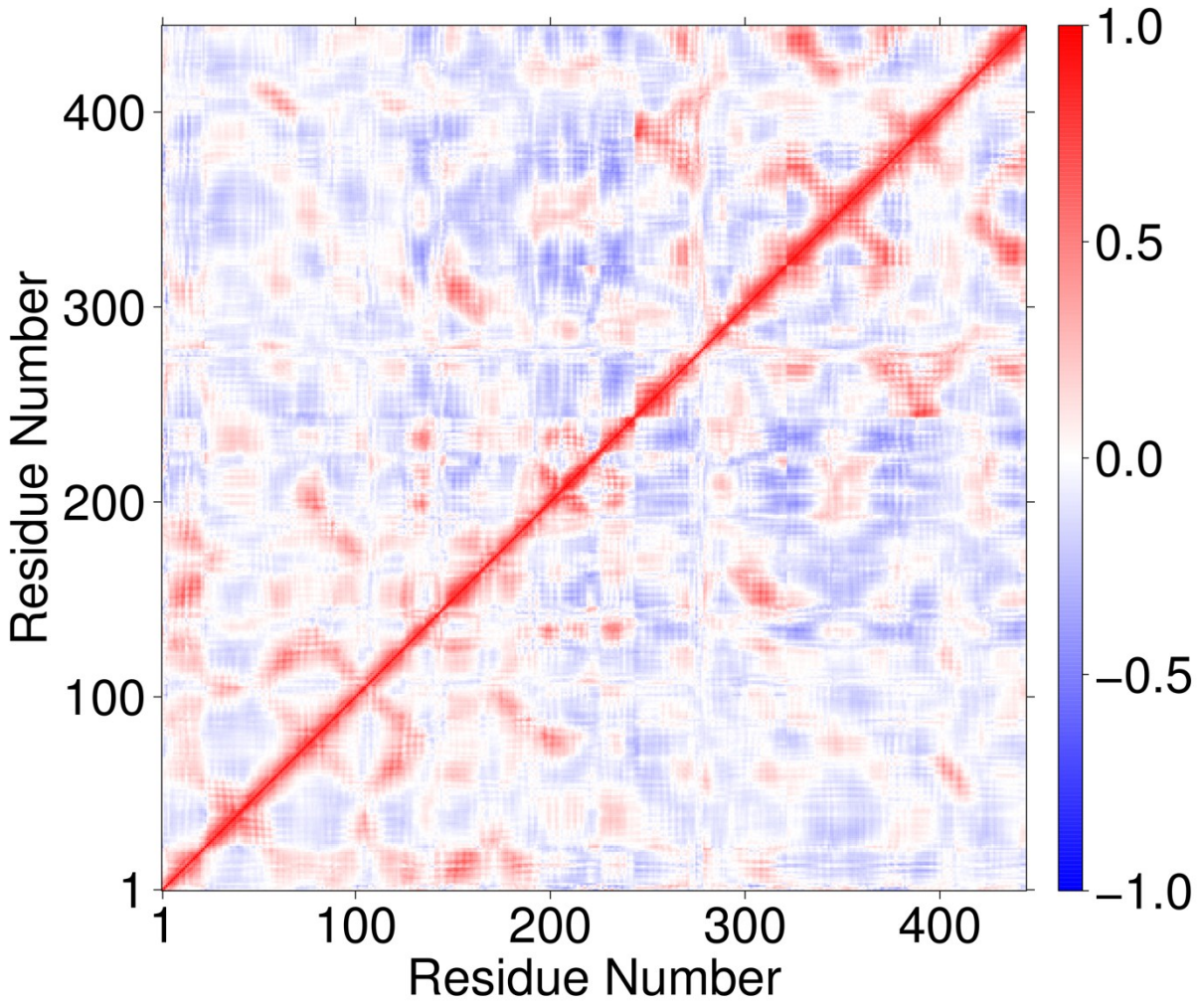

Fig. S20. Dynamic Cross Correlation for GLUT5-ManCou-CF_3 complex based on C-alpha atoms (Note Residues labels are numbered from 1-444 instead of 18-462 as in crystal structure). Red color indicates positive correlation in the movement of residues and blue color indicates negative correlation in the movement of residues.

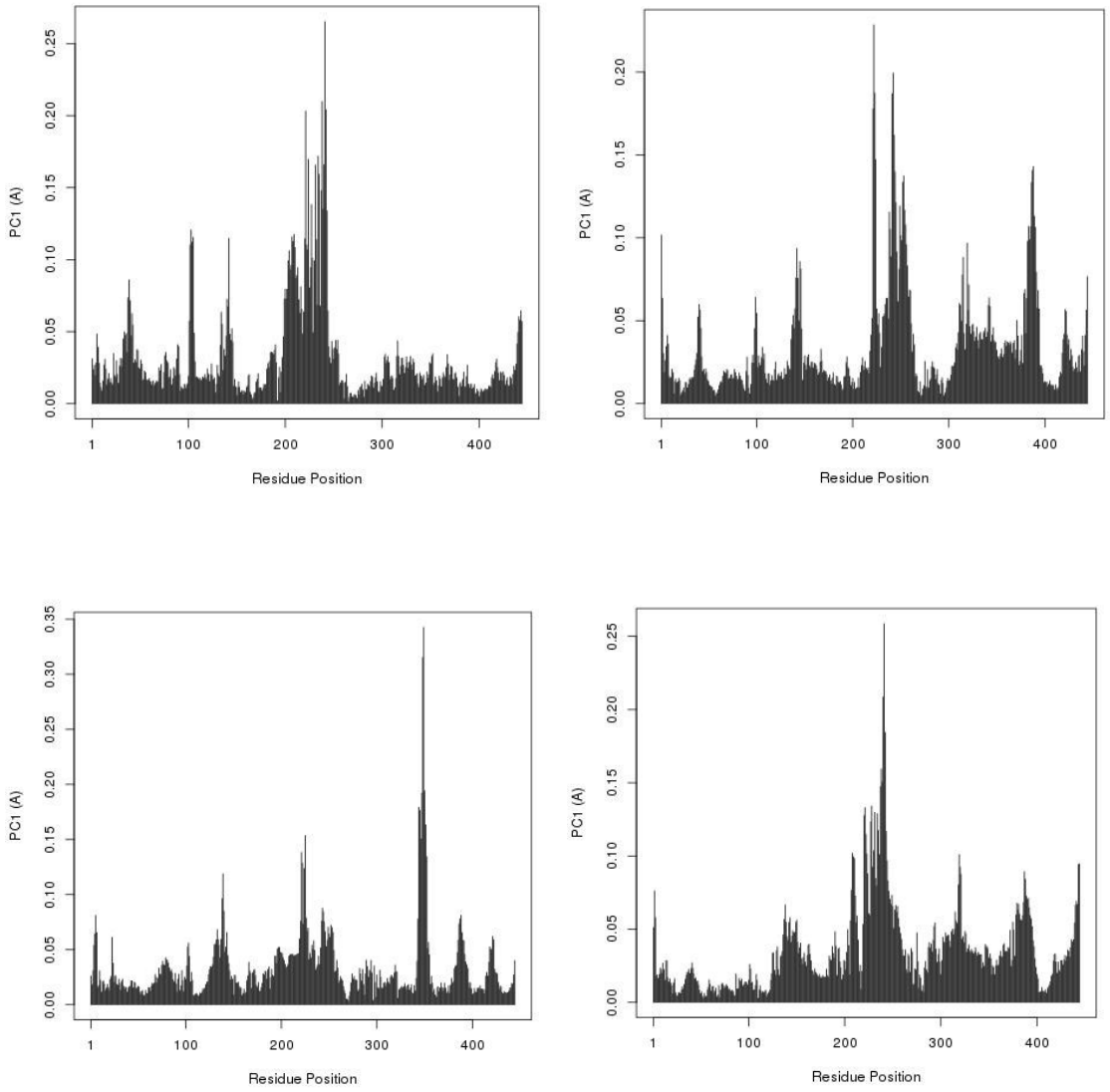

Fig. S21. PC1 variation with GLUT5 residues based on C-alpha atoms for (i) Free GLUT5 (top left), (ii) GLUT5-Fructose complex (top right), (iii) GLUT5-ManCau1 complex (bottom left), (iv) GLUT5-ManCou-CF_3 complex (bottom right). Note: Residues labels are numbered from 1-444 instead of 18-462 as in crystal structure.

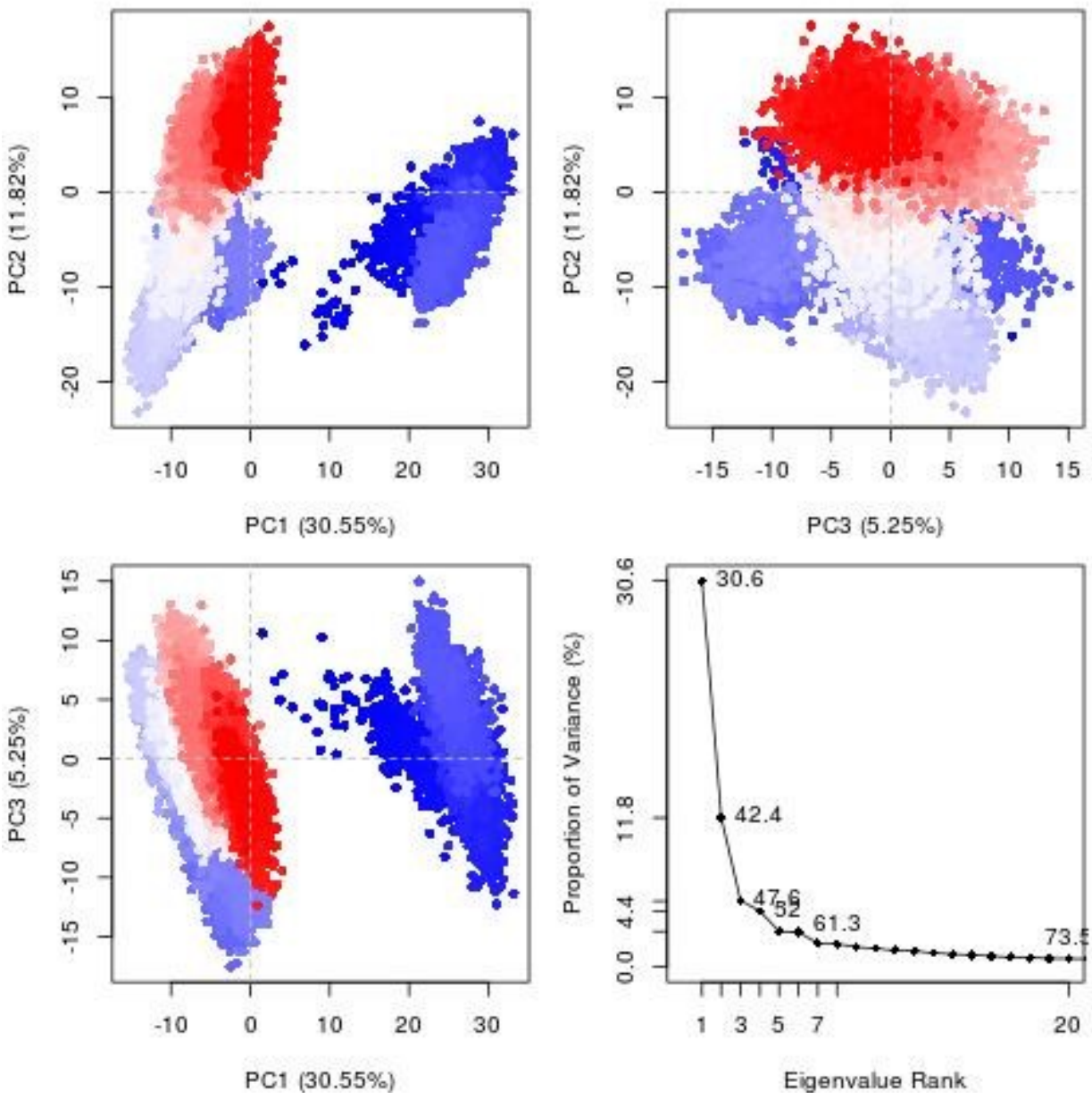

Fig. S22. PCA for the MD simulation of the free Glut5. The percentage of the total mean square displacement of residue positional fluctuations captured in each dimension is characterized by corresponding eigenvalue (PCs). The continuous color scale from blue to white to red indicates that there are periodic jumps between these conformers through the trajectory.

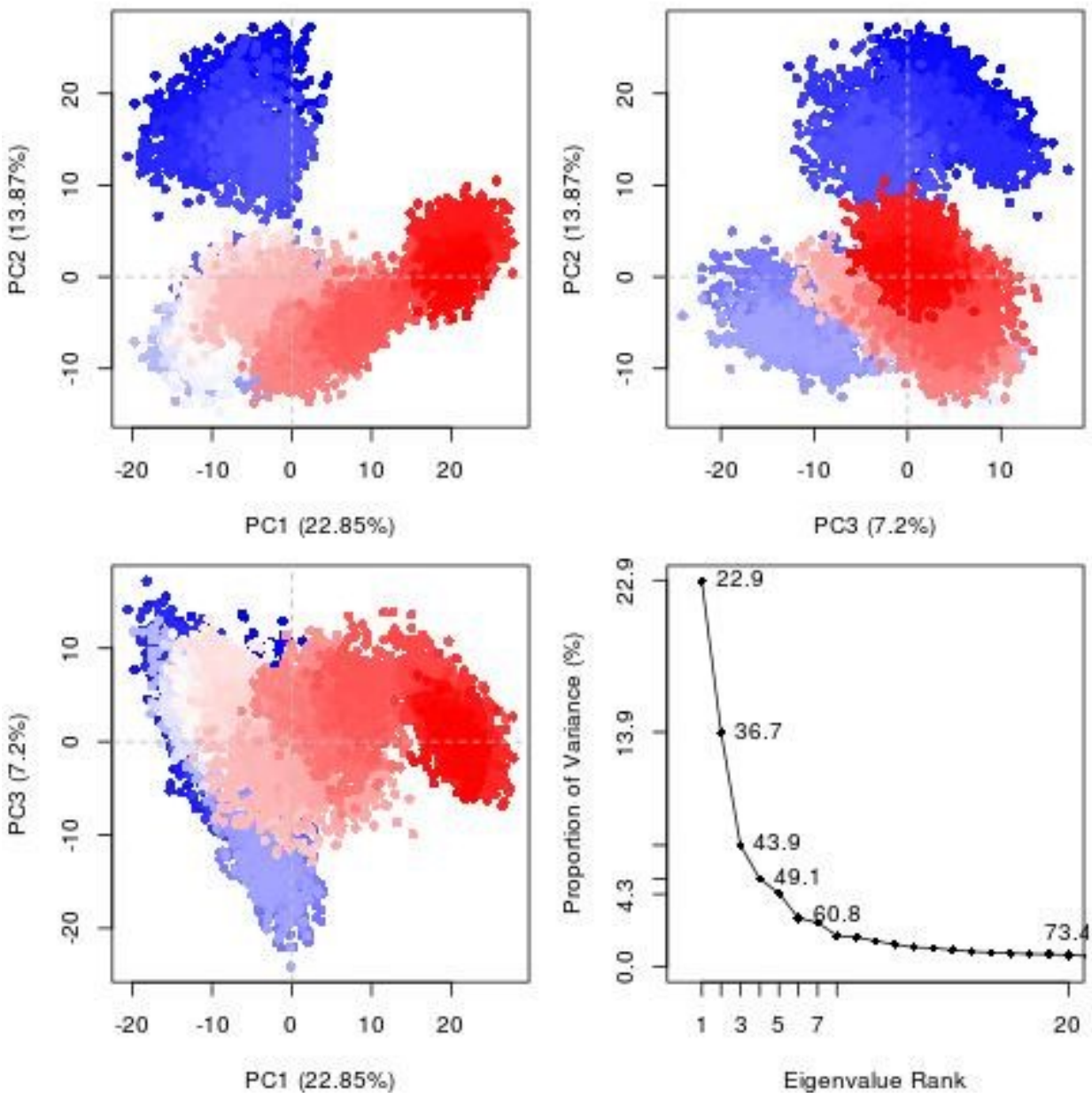

Fig. S23. PCA for the GLUT5-Fructose complex MD simulation. The percentage of the total mean square displacement of residue positional fluctuations captured in each dimension is characterized by corresponding eigenvalue (PCs). The continuous color scale from blue to white to red indicates that there are periodic jumps between these conformers through the trajectory.

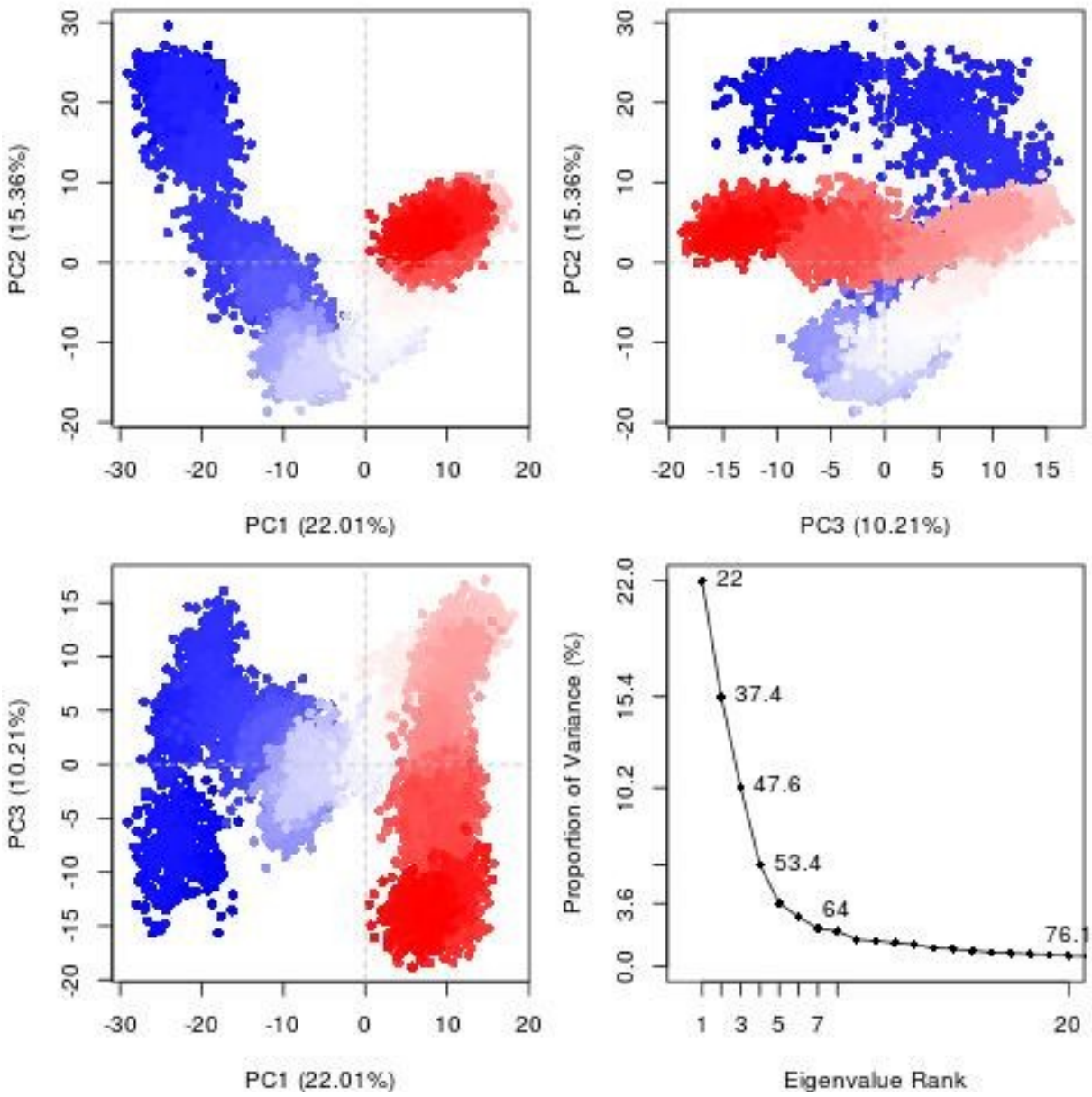

Fig. S24. PCA for the GLUT5-ManCou-H complex MD simulation. The percentage of the total mean square displacement of residue positional fluctuations captured in each dimension is characterized by corresponding eigenvalue (PCs). The continuous color scale from blue to white to red indicates that there are periodic jumps between these conformers through the trajectory.

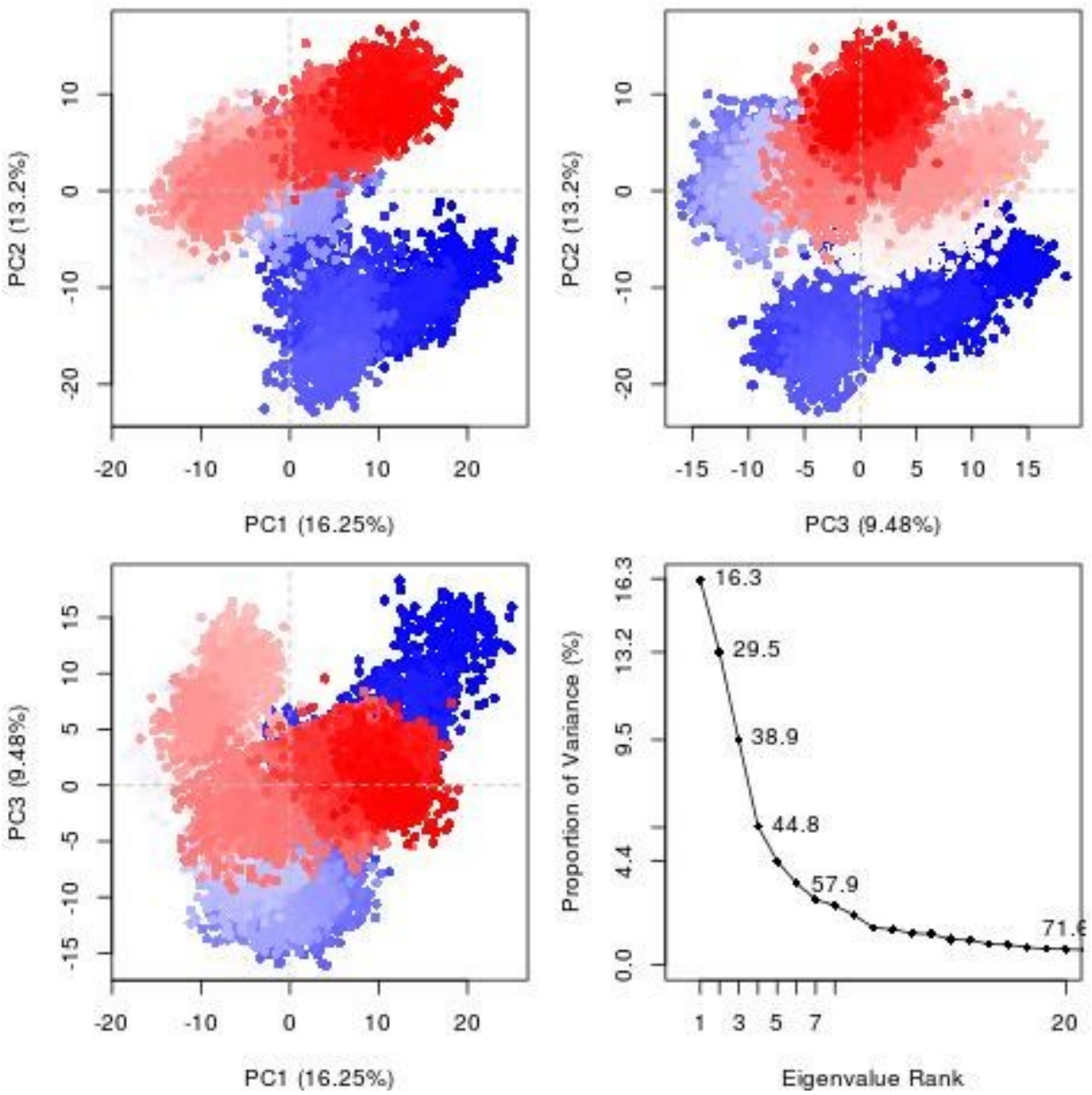

Fig. S25. PCA for the GLUT5-ManCou-CF_3 complex MD simulation. The percentage of the total mean square displacement of residue positional fluctuations captured in each dimension is characterized by corresponding eigenvalue (PCs). The continuous color scale from blue to white to red indicates that there are periodic jumps between these conformers through the trajectory.

2. Methods

2.1. Microplate uptake and inhibition assays

Breast adenocarcinoma MCF7 cells were procured from American Type Cell Culture and were grown in RPMI media supplemented with 5% fetal bovine serum (FBS) and 1% penicillin/streptomycin to lower chances of bacterial contamination. Cells were maintained at 37 °C, at 65% relative humidity, and under 5% CO_2. For microplate assay, at ~80% confluence cells were collected and plated in 96-well flat bottom plates (20,000 cells/well) and allowed to grow for 24 hours. Cells were then washed with warmed (37 ˚C) PBS solution and treated with PBS solutions of ManCou probes (20 μM) supplemented with varying concentrations of fructose. After 10 min incubation at 37 ˚C and 5% CO_2, the ManCou-fructose solutions were removed, and cells were carefully washed with warmed PBS (3 x 100 μl). Fluorescent data were immediately collected using Victor3 plate reader and using WallacTM umbelliferon (excitation 355 nm, emission 460 nm, 1.0 s) protocol. All trials were done in triplicate on each plate.

2.2. Molecular Simulations

The crystal structure of the GLUT5 receptor in the inward-open conformation (PDBID: 4YB9) was used as the starting structure. The missing loop in the C-terminal domain (residues 165-181) was modelled using the MODELLER tool in chimera 1.12.[1-4] Out of the 25 homology modelled structures, the one with the lowest Discrete Optimized Protein Energy (DOPE) score was selected for further use. The protonation states of the titratable amino acid residues were determined using the PROPKA webserver at a neutral pH of 7.0.[5] The structures of Fructose, ManCou-H and ManCou-CF_3 were modelled using Gausview and geometry optimized using the B3LYP/6-31G functional and basis set in Gaussian09.[6-9] The ligands were each docked to the a rigid GLUT5 structure using AutoDock4.[10] The grid for the docking was positioned based on the binding site proposed by Nomura *et al* in the crystal structure paper, with grid dimensions of 50Åx60Åx50Å and a grid spacing of 0.375 Å.[1] The docking used a Lamarckian genetic algorithm with a population size of 150 with only the best structure of each population progressing to the next generation, the maximum number of generations was set at 27,000. ManCou-H and ManCou-CF_3 were docked using the same protocol and the lowest energy conformation of each ligand was used. The lowest energy docked conformation of fructose was found to align well with the proposed orientation of fructose and its interacting residues suggested in the crystal structure paper (Fig. S4).[1]

To embed the ligand bound GLUT5 receptor in a lipid membrane the lipid builder function of the CHARMM-GUI webserver was used.[11] POPC lipids were chosen, owing to their use by other membrane simulation studies where their properties have been shown to correlate well

with experimental results,[12] additionally they have previously been successfully used in MD simulations of the structurally similar GLUT1 and GLUT4 receptors. [13, 14] The GLUT5 structure was inserted into a comparably sized hole in the center of the constructed membrane. The system was solvated with TIP3P water molecules and K^+ and Cl^- ions were added to create a neutral system with an ion concentration of 0.15 M and box dimensions of roughly 100Åx100Åx110Å.[15] The system for MD was then setup using leap for Amber14 with the AmberFF14SB forcefield with the additional lipid14 forcefield for the POPC membrane (Fig. S3).[16-18]

The Fructose, ManCou-H and ManCou-CF_3 ligands were parametrized using the Antechamber package in Amber14 using the AM1-BCC charge model with the GAFF forcefield.[19] The solvated systems were subject to 5000 steps of steepest descent minimization followed by 5000 steps of conjugate gradient minimization using pmemd. Initially the system was heated as an NVT ensemble to 100 K using a Langevin thermostat for 2500 steps whilst the membrane was restrained with a force constant of 10 kcal/mol, the system's pressure was equilibrated with as an NPT ensemble to 1atm with anisotropic pressure coupling whilst a heating to 300 K was performed for 50000 steps, whilst still restraining the lipid membrane. This was followed by a short MD run of 5 ns, without lipid restraints as an NVT ensemble, to equilibrate the PBC box dimensions. The productive 1 microsecond simulations also used a Langevin thermostat and anisotropic pressure coupling, the four different systems were performed on a single GPU of a four GPU workstation using pmemd.cuda for Amber14. For all the MD steps; minimization, heating equilibration and productive dynamics a non-bonded cut off of 10.0 Å was used. For the heating, equilibration and productive MD simulations a step size of 2 fs was used. The analysis of the resultant trajectories was made with a range of programs; CPPTRAJ for Amber 16 and RMSF analysis tools in Visual Molecular Dynamics (VMD), pictures were rendered using Chimera[4, 16, 20]. RMSD, Radius of Gyration, Solvent Accessible Surface Area (SASA) analysis were carried over C-alpha atoms of protein throughout the trajectory using CPPTRAJ for Amber 16. RMSF over C-alpha atoms of proteins was done using VMD.

Dynamic Cross Correlation Analysis was done to understand the extent to which the residual displacements of Glut5 are correlated to each other using Bio3d software.[21] Analysis was carried out on all C-alpha atoms through entire 1 μsec trajectory. To minimize the root mean square differences between the equivalent residues the structures were superimposed using fit.xyz function. Subsequently, dccm function was used to get a matrix of all residue-wise cross-correlations. This matrix was plotted in graphical form using plot.dccm function. Similarly, Principal Component Analysis was also done using Bio3d software.[21] Analysis was carried out on all C-alpha atoms through entire 1 μsec trajectory. To minimize the root mean

square differences between the equivalent residues the structures were superimposed using fit.xyz function. Following superimposition, the pca.xyz function was used for Principal Component Analysis. PCA plots were generated using the function plot(pc) function MGBSA analysis was done on 50 frames taken equidistantly from Molecular Dynamics simulations (200 ns to 1000 ns) using MMPBSA.py in Amber16. The entropic contribution in MMPBSA.py was not included. While entropic contribution is essential for calculating absolute binding free energies, it is considered not important when calculations of relative binding free energies of different ligands to the same protein are performed. In addition, the entropy calculations tend to be very computationally demanding and not accurate [22-25]

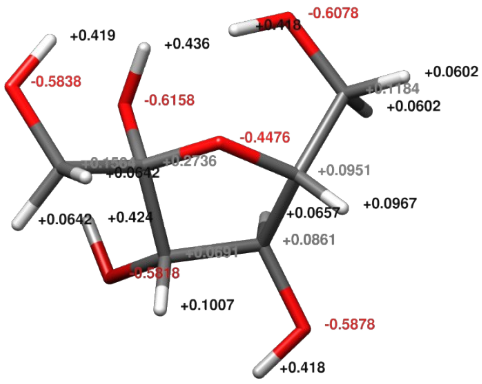

Fig. S26. Charge distribution of the optimized fructose molecule.

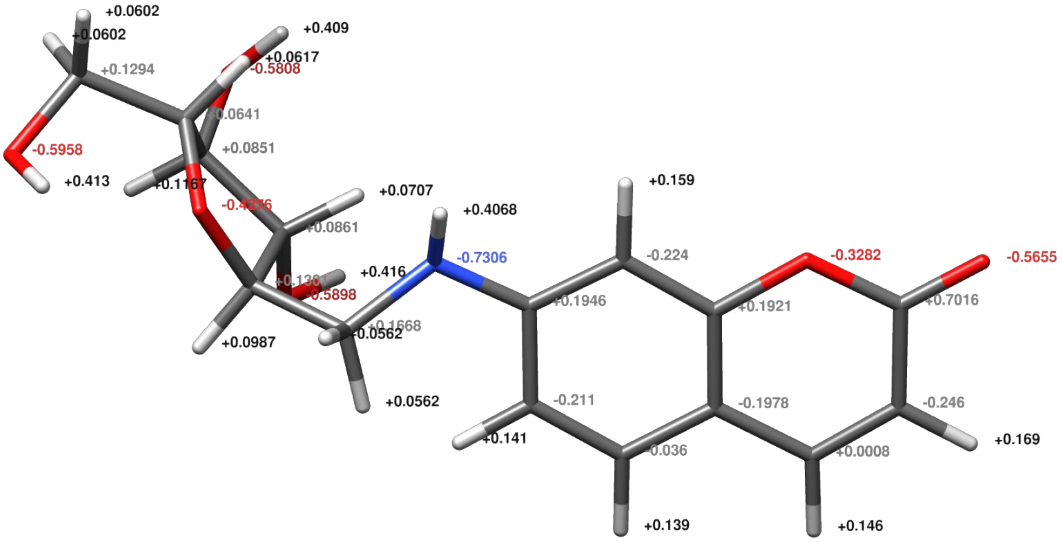

Fig.S27. Charge distribution of the optimized ManCou-H molecule.

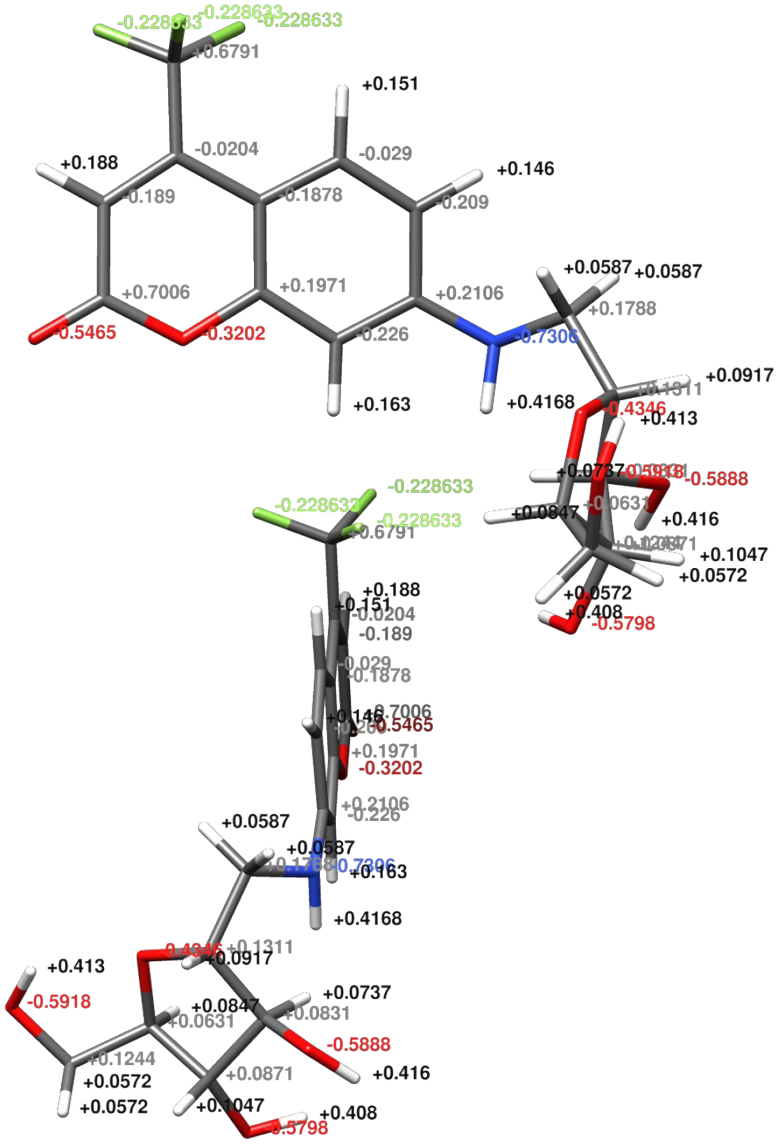

Fig. S28. Charge distribution of the optimized ManCou-CF_3 structure.

3. Cross Correlation Analysis

The end of TM8 and cytoplasmic region connecting Transmembrane Helix (TM) 8 and TM9, spanning from I328 to G348 are positively correlated with the TM10 and extracellular region connecting TM9 and TM10 containing from V368 to L398 in all the cases with some variations. The motion of GLUT5 fragment containing TM2 and cytoplasmic region connecting TM2 and TM3 from residue F78 to residue R98 are positively related with fragment containing extracellular region connecting TM3 and TM4 and TM4 from residue K123 to residue M148, this correlation is less intensive in the free GLUT5.

Similarly, the motion of fragment in TM9 (around residues G348) is positively correlated with TM12 (residues around I448) for all the GLUT5-ligand complexes and free GLUT5.

The cluster of residues in the TM10 and cytoplasmic region connecting TM10 and TM11 (region of residues A388 to V418) are positively correlated in motion with the fragment of cytoplasmic region connecting TM6 to TM7 and upper half of TM7 (residues from L268-Q288). This correlation is extended for GLUT5-ManCou-CF_3 complex till residues N294. The residue N294 of GLUT5 forms a specific and stable hydrogen bond with ManCou-CF_3.

The fragment of GLUT5 in the cytoplasmic region connecting TM10 to TM11 (around L398) are correlated with fragment of TM12 (residues around L458). This correlation is positive for Fructose, ManCou-H and ManCou-CF_3, but also some anti-correlated motions around this region can be seen for GLUT5-ManCou-CF_3 complex. This correlation is very weak for free GLUT5.

Positive correlation in motion of fragments in TM5 (around residues L168) with TM8 (residues around T318) can be seen in the case of Fructose, ManCou-H and ManCou-CF_3 but is very weak in case of free GLUT5.

Correlation in the motion of fragment of protein in TM2 (around the residues from M88 to F118) with the fragment in TM6 and extracellular region connecting TM5 to TM6 (containing the residue from N188 to R218) can be seen with some differences in all the complexes. For ManCou-CF_3 complex the correlation is missing between residues around R98 and L198, for ManCou-H the correlation is strong between regions of residues around R98 and R218. This correlation is variable for all the cases.

Strong anti-correlation in the motion of fragment contain residues in TM6 (from L198 to R218) can be seen in the case of free GLUT5 with residues in cytoplasmic region connection TM6 and TM7 (from K258 to V278). These interactions are partially present in ManCou-H but missing in Fructose and ManCou-CF_3.

Anti-correlated motion can be seen between the fragment of the protein from residue in cytoplasmic region connecting TM6 to TM7 (R218 to K258) with residues in TM8 and cytoplasmic region connecting TM8 to TM9 (from T318 to G348) of GLUT5 when bound to all the ligands, these correlations are weak in free GLUT5.

Anti-correlated motion in the protein fragments in the cytoplasmic region connecting TM6 to TM7 (around residue L268) with the extracellular region connecting TM9 to TM10 (around residue V368) is seen when GLUT5 is in complex with Fructose. This correlation is missing in all other cases.

The residues in cytoplasmic region connecting TM6 to TM7 of protein (from residue R238 to K258) is anti-correlated in motion with cytoplasmic region connecting TM10 to TM11 of protein (from residues L398 to V418) in all the GLUT5-ligand complexes. This interaction of protein is weak for free GLUT5.

There is weak correlated motion between the regions around residues Y298 with residues around V368 for GLUT5-Fructose complex. This is correlation is not present for other ligands and free GLUT5.

4. Principal Component Analysis

For the free GLUT5, the PC1 reflects motions in Intercellular Helix 1-3 (ICH1-3) region (residues R218 to L268) and also in the extracellular turn connecting TM1 and TM2 (residues T53 to V68). The starting residues of TM3 (residues around A122) also shows strong impact on PC1.For the complex with Fructose, the highest impact on PC1 is from the ICH3 and cytoplasmic region connecting ICH3 to TM7 (residues from R238 to V278) and cytoplasmic region connecting TM10 and TM11 (residues from P395 to A411). Protein fragments towards the end of TM8 (residue from I328 to L338) and cytoplasmic region connecting TM4 and TM5 (residue L153 to V165) have some impact on PC1.For GLUT5-ManCou-H complex, PC1 exhibits contributions mainly from the extracellular region connecting TM9 and TM10 (around D367) and protein fragment in ICH3 and cytoplasmic region connecting ICH3 to TM7 (from residue R238 to M271). For ManCou-CF_3 complex, the main contributions to PC1 is from ICH1, ICH2, ICH3 and cytoplasmic region connecting ICH3 to TM7 (residues Q223 to V278 and from the end of TM8 and cytoplasmic region connecting TM8 and TM9 (from residue V330 to R341).

5. References

1. N. Nomura, G. Verdon, H. J. Kang, T. Shimamura, Y. Nomura, Y. Sonoda, S. A. Hussien, A. A. Qureshi, M. Coincon and Y. Sato, *Nature*, 2015, **526**, 397.
2. R. Sánchez and A. Šali, *Prot. Struct. Pred.*, 2000, 97-129.
3. B. Webb and A. Sali, *Protein Struct. Pred.*, 2014, 1-15.
4. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605-1612.
5. T. J. Dolinsky, J. E. Nielsen, J. A. McCammon and N. A. Baker, *Nucl. Acids Res.*, 2004, **32**, W665-W667.
6. K. Kim and K. D. Jordan, *J. Phys. Chem.*, 1994, **98**, 10089-10094.
7. P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623-11627.
8. R. Dennington, T. Keith, J. Millam, K. Eppinnett, W. L. Hovell and R. Gilliland, *GaussView*, Version, 2009.
9. M. Frisch, G. Trucks, H. B. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, B. Mennucci and G. Petersson, *Gaussian 09, revision D. 01*, Gaussian, Inc., Wallingford CT, 2009.
10. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J. Comput. Chem.*, 2009, **30**, 2785-2791.
11. S. Jo, J. B. Lim, J. B. Klauda and W. Im, *Biophys. J.*, 2009, **97**, 50-58.
12. S. W. I. Siu, R. Vácha, P. Jungwirth and R. A. Böckmann, *J. Chem. Phys.*, 2008, **128**, 125103.
13. X. Fu, G. Zhang, R. Liu, J. Wei, D. Zhang-Negrerie, X. Jian and Q. Gao, *J. Chem. Inform. Model.*, 2016, **56**, 517-526.
14. M.-S. Park, *PloS one*, 2015, **10**, e0125361.
15. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.
16. D. A. Case, V. Babin, J. Berryman, R. Betz, Q. Cai, D. Cerutti, T. Cheatham Iii, T. Darden, R. Duke and H. Gohlke, *Amber 14*, University of California; San Francisco, 2014.
17. D. Case, V. Babin, J. Berryman, R. Betz, Q. Cai, D. Cerutti, T. Cheatham III, T. Darden, R. Duke and H. Gohlke, *Amber*, 2014, **14**, 29-31.
18. C. J. Dickson, B. D. Madej, Å. A. Skjevik, R. M. Betz, K. Teigen, I. R. Gould and R. C. Walker, *J. Chem. Theor. Comput.*, 2014, **10**, 865-879.
19. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157-1174.

20. W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33-38.
21. B. J. Grant, A. P. C. Rodrigues, K. M. ElSawy, J. A. McCammon and L. S. D. Caves, *Bioinformatics*, 2006, **22**, 2695–2696.
22. P. Marimuthu and K. Singaravelu, *Biochemistry*, 2018, **57**, 1249–1261.
23. K. Mulholland and C. Wu, *J. Chem. Inf. Model.*, 2016, **56**, 2093–2102.
24. K. Mulholland and C. Wu, *J. Chem. Inf. Model.*, 2016, **56**, 1872–1883.
25. M. Ylilauri and O. T. Pentikäinen, *J. Chem. Inf. Model.*, 2013, **53**, 2626–2633.