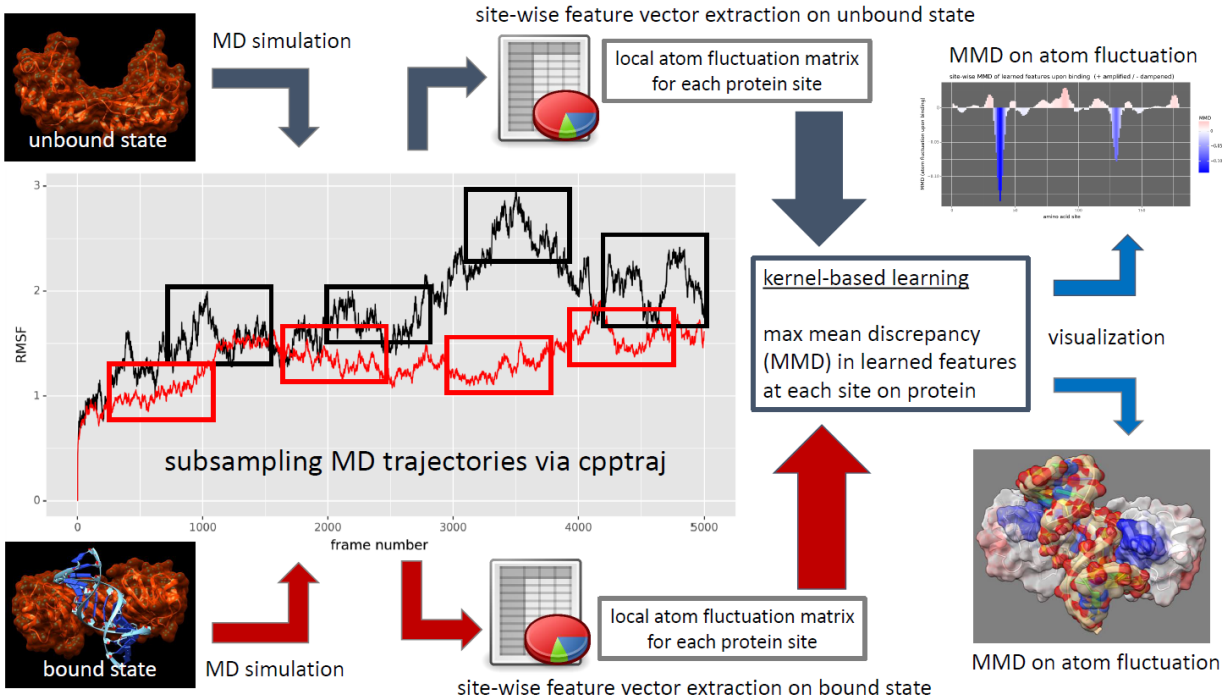


Figure S1 – Overview of the maxDemon 4.0 subroutine in the ATOMDANCE statistical machine learning post-processor for comparative protein dynamics. First two molecular dynamics simulations representing the functional end states are conducted (e.g. drug/DNA/protein bound vs. unbound). The .pdb structure, .prmtop topology, and .nc trajectory files for both states are input to the software and the trajectories are repeated subsampled according to user specification using cpptraj. Site-wise local atom fluctuation matrices are constructed from the subsampling and used to train a Gaussian process kernel (radial basis function). At each site on the protein, the maximum mean discrepancy (MMD) in reproducing kernel Hilbert space is calculated, representing the distance between the learned features in the transformed data space that best captures the functional difference in protein dynamics at the given site. The MMD is signed negative if atom motion is dampened or positive if it is amplified. This MMD is visualized in a variety of plots and can be color-mapped to the .pdb structure file in UCSF ChimeraX (blue indicating regions of dampened motion due to binding interaction).

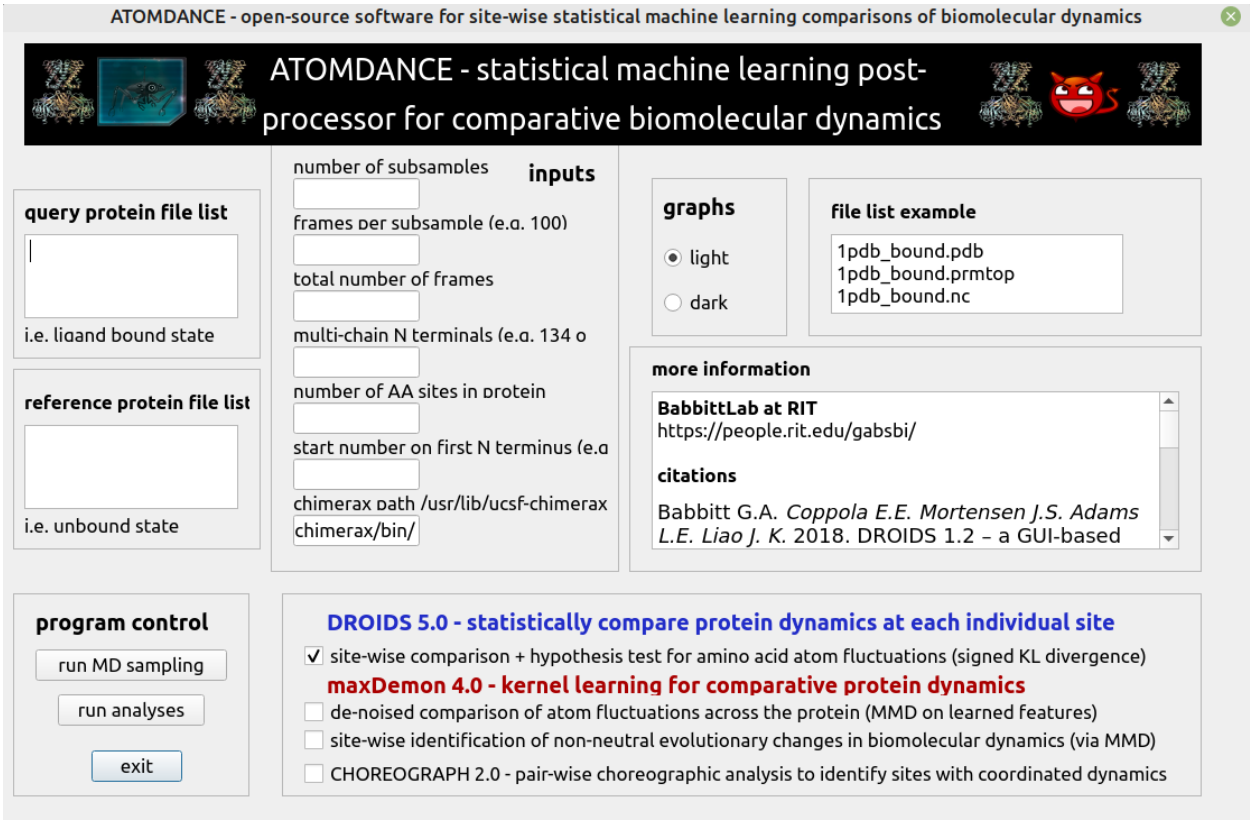


Figure S2 – The Graphical User Interface (GUI) for the ATOMDANCE statistical machine learning post-processor for comparative protein dynamics. Each comparison requires six files; two .pdb structure files, two .prmtop topology files, and two .nc trajectory files…each representing one of the two functional or evolutionary states being compared (i.e. bound vs. unbound or before vs. after mutation). The ATOMDANCE GUI for generating molecular dynamics simulations using open-source software is shown in Supplemental Figure 3.

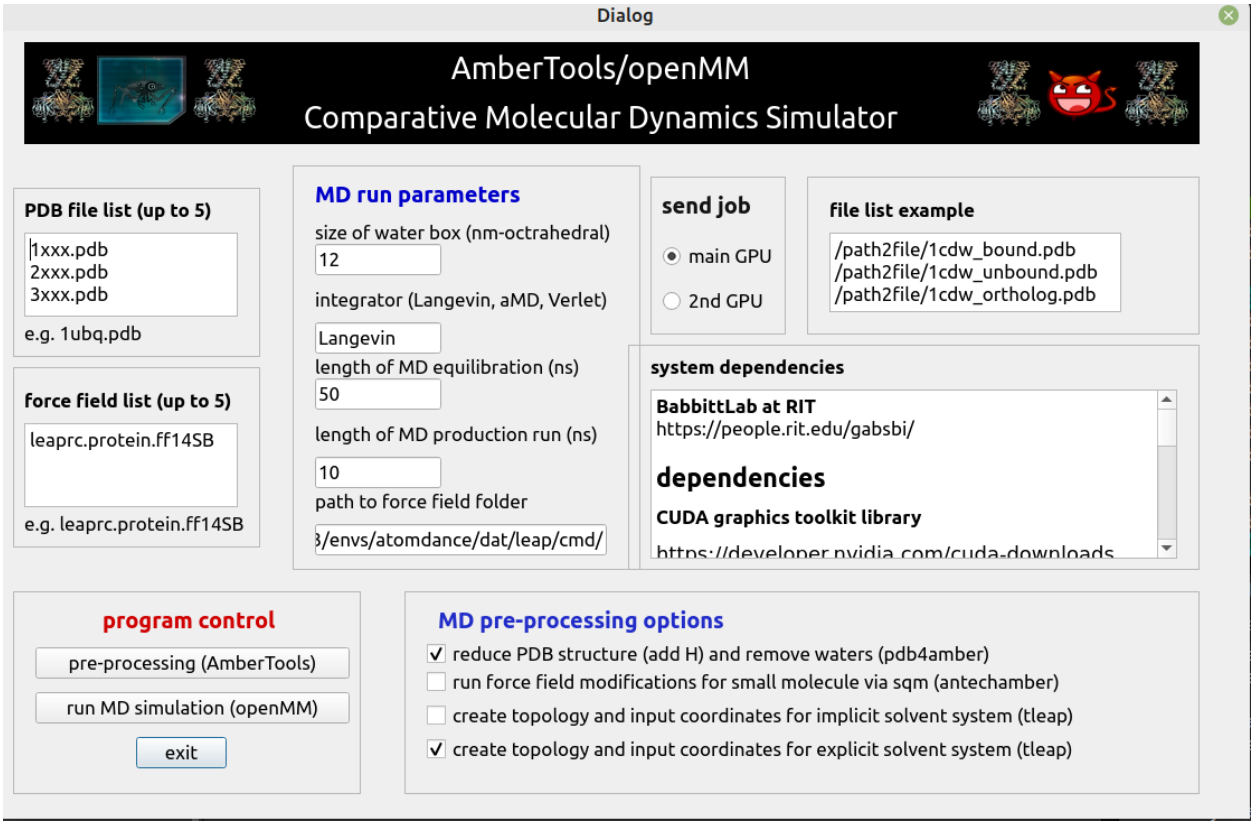


Figure S3 – Overview of the ATOMDANCE supplemental GUI for running MD simulations. This program only requires a conda install for AmberTools and OpenMM. Users can batch run molecular dynamics (MD) simulations for up to 5 PDB structures by listing them in the window in the top left corner. The window below this should contain a list of all the necessary force fields (in AmberTools). Checkboxes for MD simulation pre-processing in AmberTools includes drying and reducing (i.e. adding hydrogens), calculating and defining force field modifications required by small molecule ligands (via antechamber and sqm = scaled quantum mechanical optimization), and preparing a charge neutralized solvated state by adding Na+ and Cl- ions and water molecules via tleap (in AmberTools).

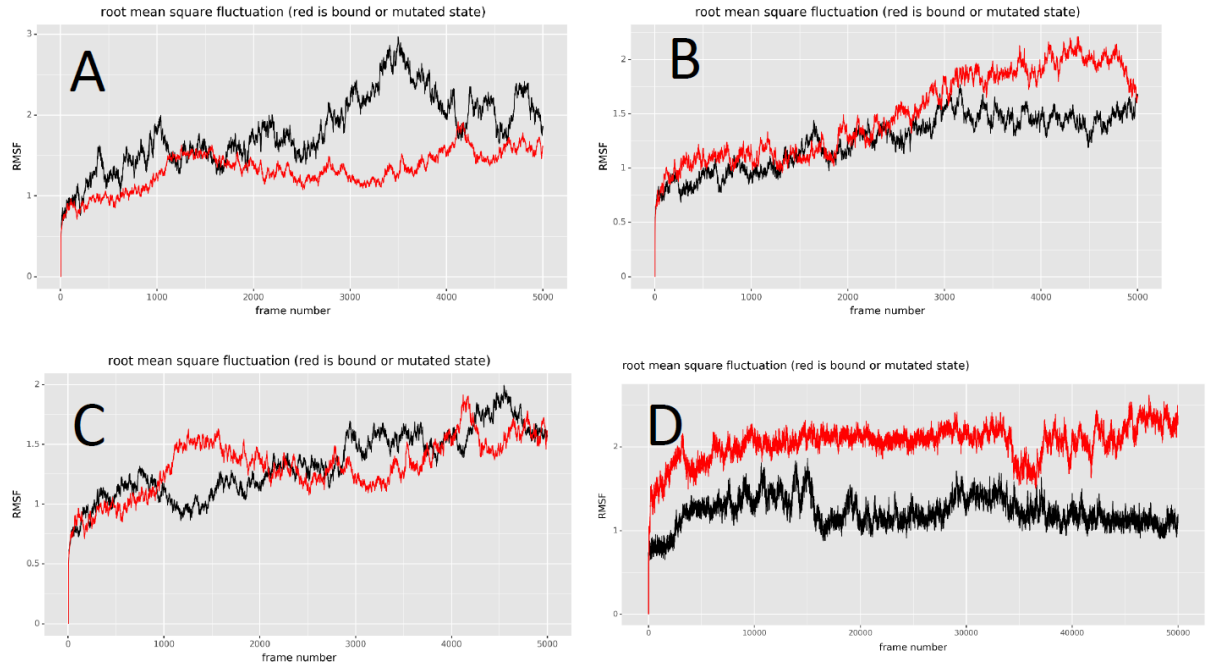


Figure S4 – Root mean square deviation over time (RMSF) plots for Figure 3 showing the 10ns production runs repeatedly sampled for the analyses. The comparisons include (A) DNA-bound vs. unbound TATA binding protein (PDB: 1cdw), (B) sorafenib-bound vs. unbound B-Raf kinase domain (PDB: 1uwh), (C) SARS-CoV-2 viral bound vs. unbound angiotensin-converting enzyme 2 (ACE2) protein (PDB: 6m17), and (D) the allosteric activated (i.e. InsP6 bound) vs inactivated (i.e. unbound) Vibrio cholera toxin RTX cysteine protease domain (PDB: 3eeb). The RMSF for the molecular dynamics trajectory over time for unbound state of the target proteins are shown in black and the bound state is shown in red.

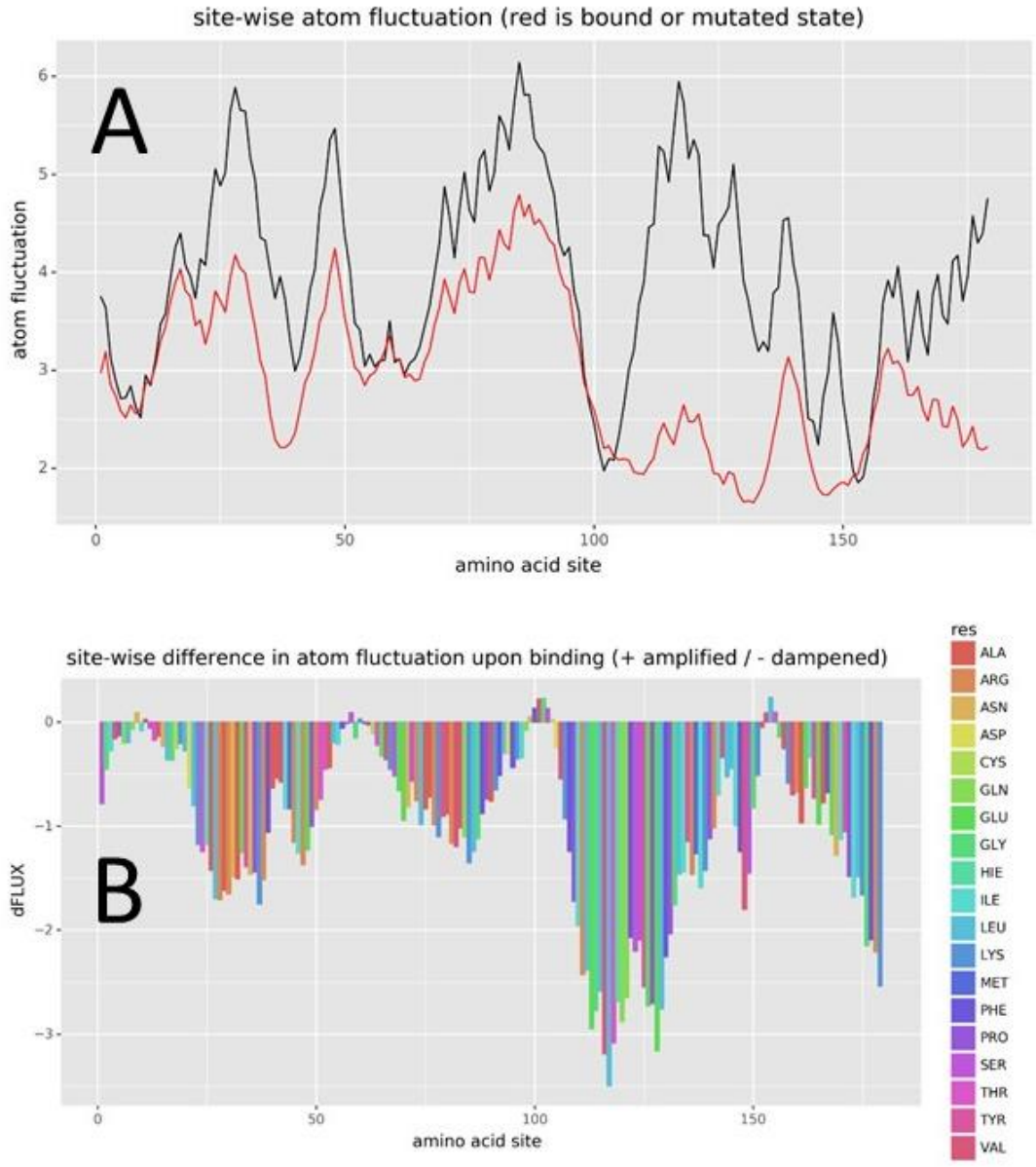


Figure S5 – Alternative plots generated by ATOMDANCE indicating (A) site-wise average atom fluctuation profiles and (B) site-wise average difference in DNA-bound versus unbound TATA binding protein (PDB: 1cdw).

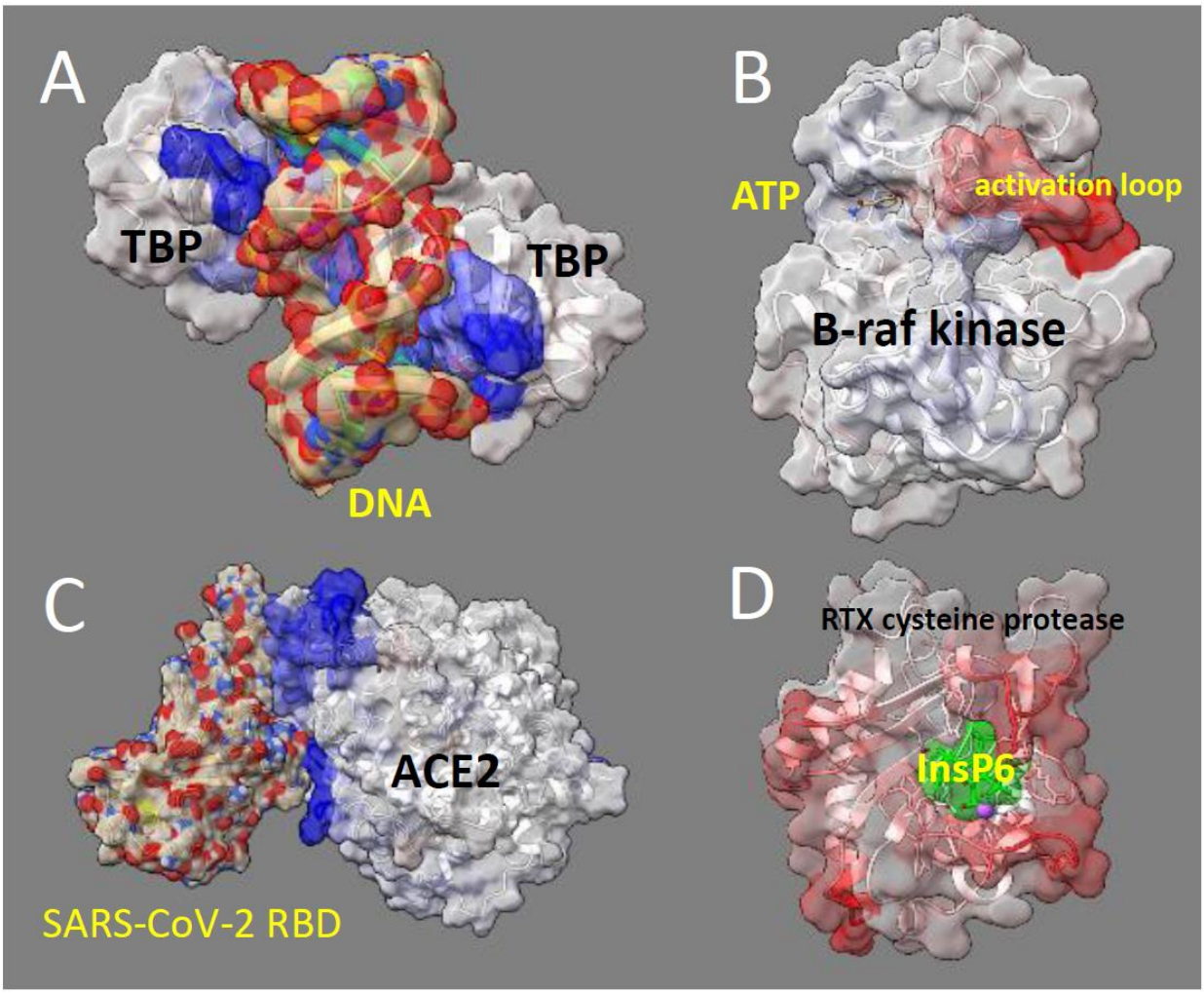


Figure S6 – Close-up views of color-mapped structures from Figure 3. Note red indicates the maximum mean discrepancy (MMD) between learned features where atom motions are amplified in the ligand-bound state and conversely, blue indicates where these atom motions are dampened. The comparisons include (A) DNA-bound vs. unbound TATA binding protein (PDB: 1cdw), (B) sorafenib-bound vs. unbound B-Raf kinase domain (PDB: 1uwh), (C) SARS-CoV-2 viral bound vs. unbound angiotensin-converting enzyme 2 (ACE2) protein (PDB: 6m17), and (D) the allosteric activated (i.e. InsP6 bound) vs inactivated (i.e. unbound) Vibrio cholera toxin RTX cysteine protease domain (PDB: 3eeb).

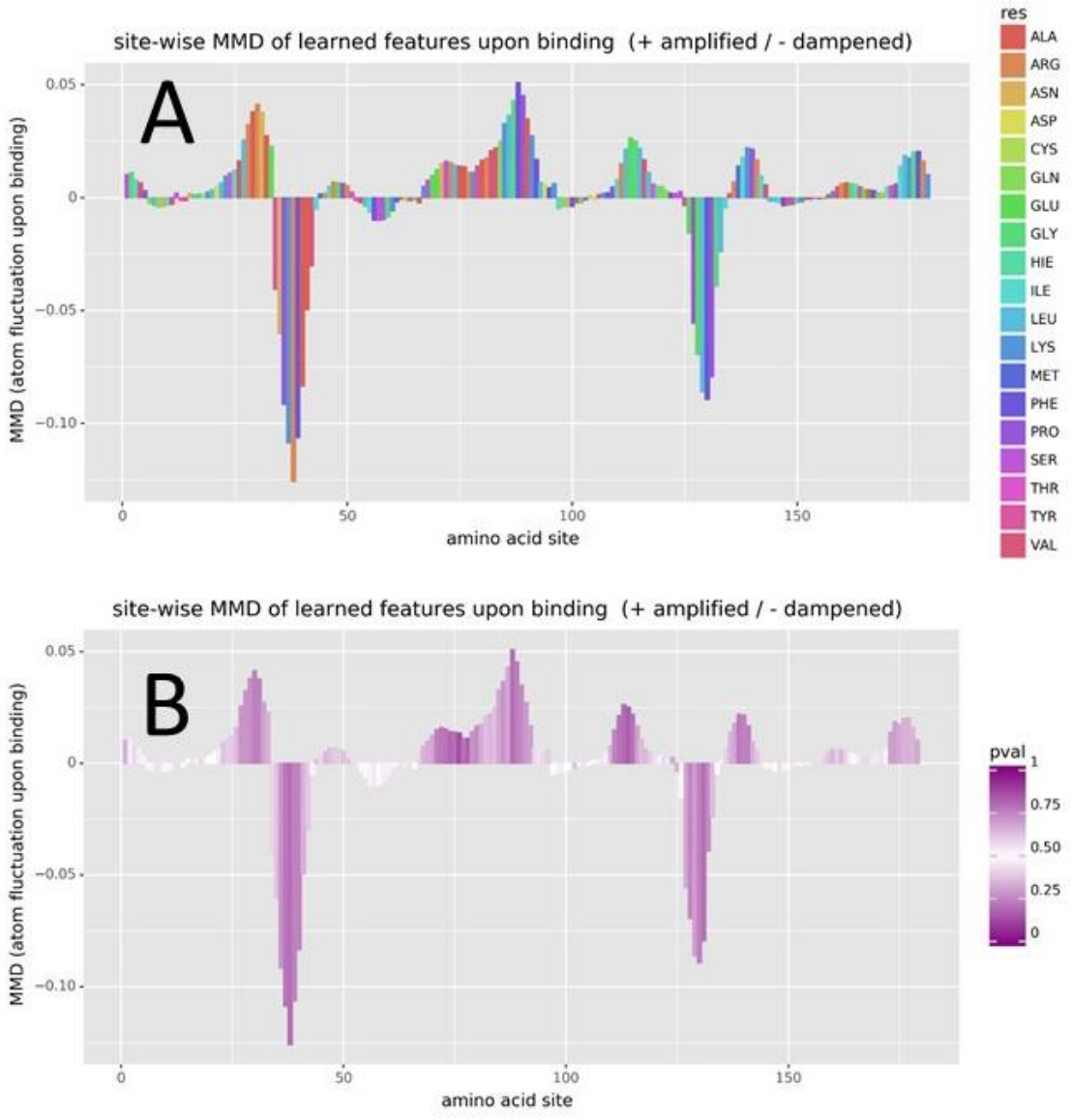
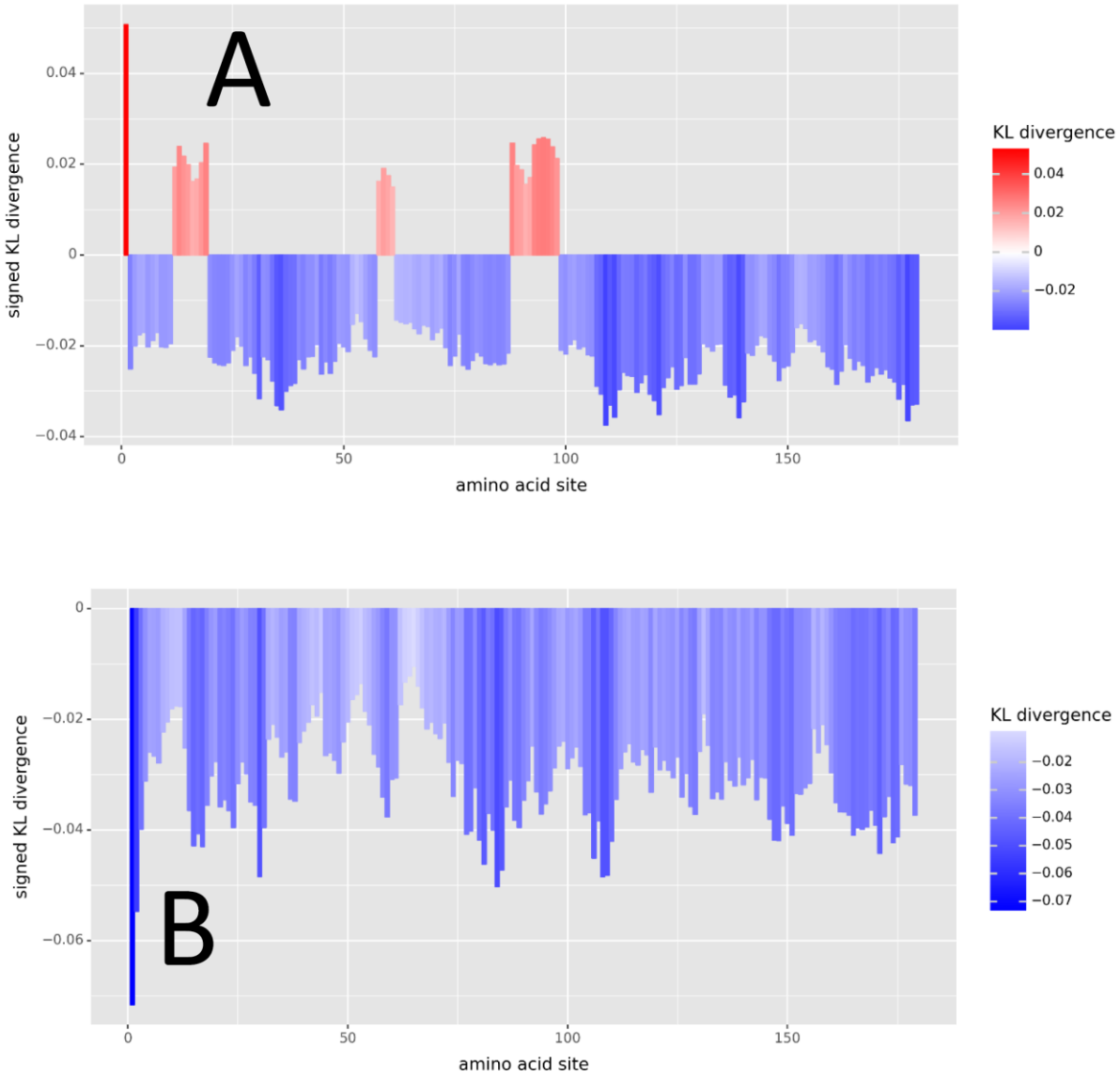


Figure S7 – Alternative plots generated by ATOMDANCE indicating site-wise max mean discrepancy (MMD) in learned features trained on local atom fluctuation showing alternative color plotting for (A) amino acid type and (B) bootstrapped empirical p values.


A
B
signed KL divergence
amino acid site
KL divergence
0.04
0.02
0
−0.02
−0.04
−0.06
0
50
100
150
−0.02
−0.03
−0.04
−0.05
−0.06
−0.07

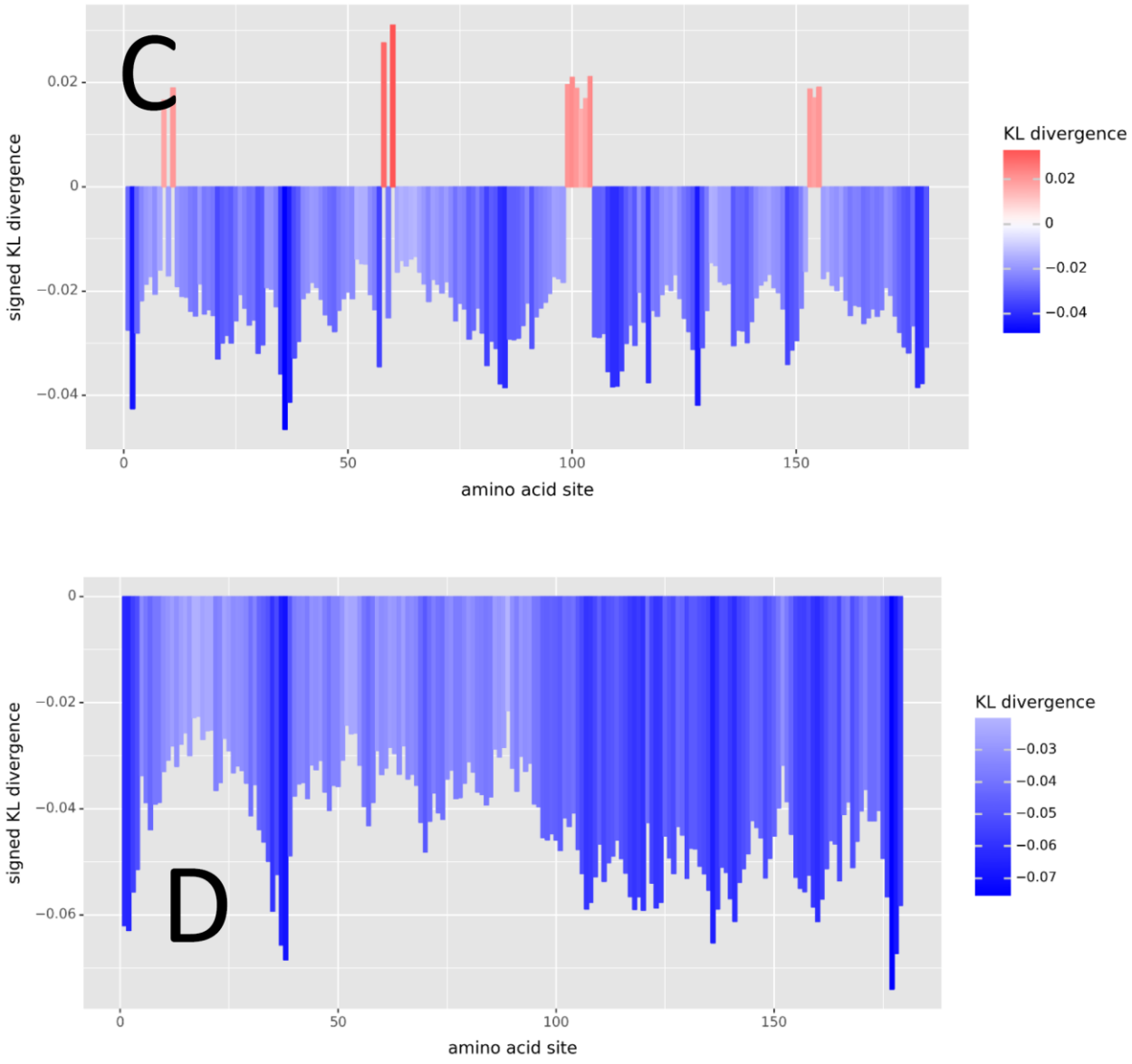
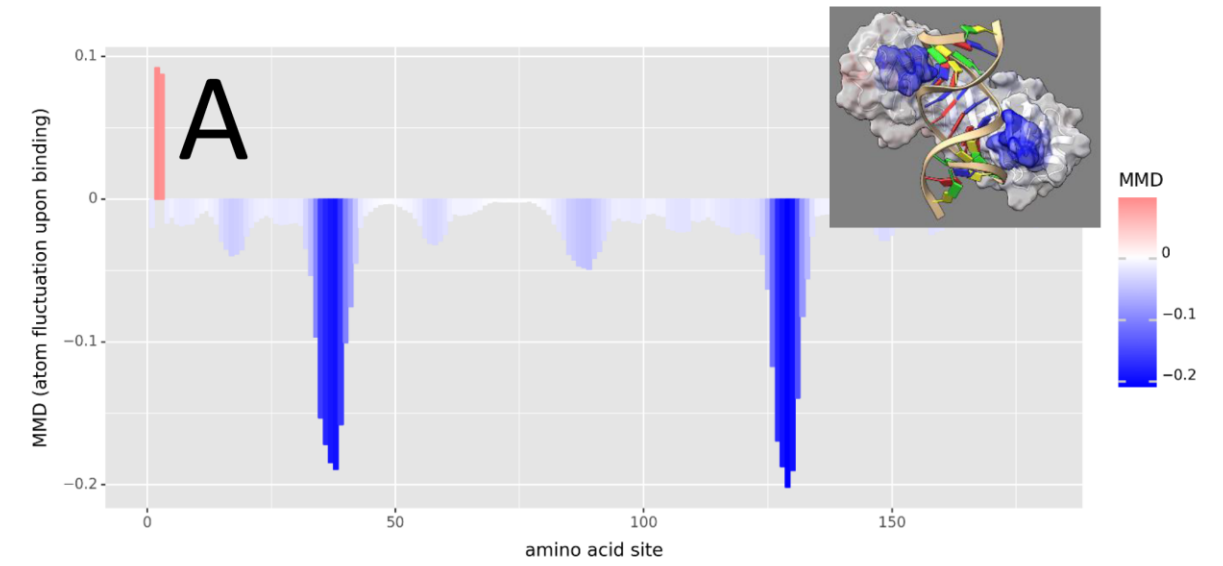
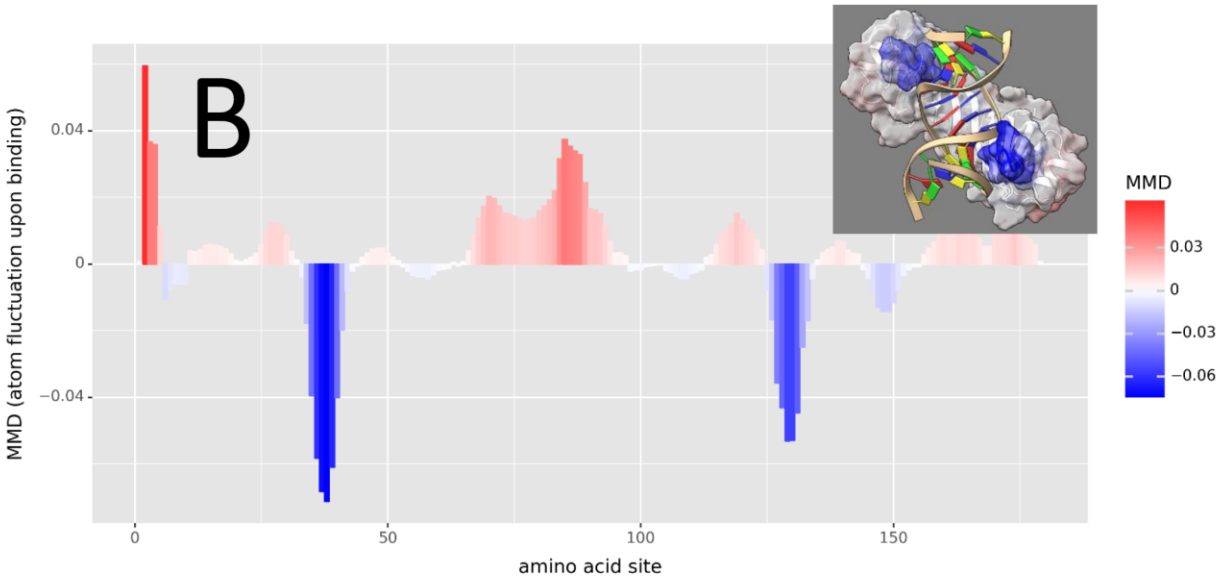


Figure S8 – Comparison of ATOMDANCE: DROIDS 5.0 on various software and methods of molecular dynamics simulations (50ns equilibration and 10ns production). The results of the site-wise signed KL divergence (eqn 3) is shown for (A) GPU-accelerated Verlet integration with an Andersen thermostat in OpenMM, (B) GPU-accelerated Langevin integration in OpenMM, (C) GPU-accelerated particle-mesh Ewald aMD in Amber20 (i.e. pmemd.cuda), and (D) GPU-accelerated aMD integration in OpenMM.


A
MMD (atom fluctuation upon binding)
0.1
0
−0.1
−0.2
0
50
100
150
amino acid site
MMD
0
−0.1
−0.2


B
MMD (atom fluctuation upon binding)
0.04
0
−0.04
0
50
100
150
amino acid site
MMD
0.03
0
−0.03
−0.06

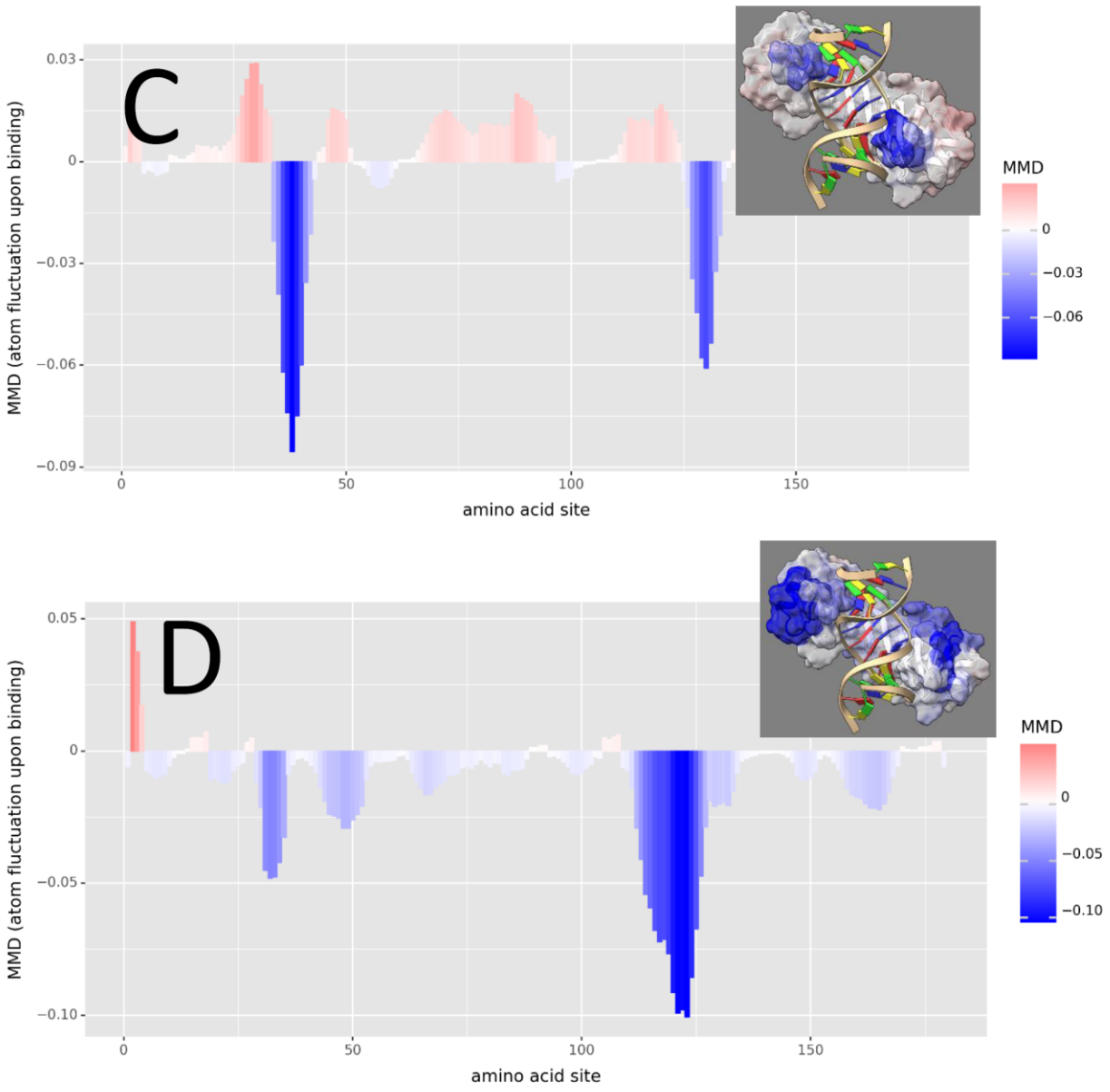


Figure S9 – Comparison of ATOMDANCE: maxDemon 4.0 on various software and methods of molecular dynamics simulations (50ns equilibration and 10ns production)**.** The results of the site-wise signed KL divergence (eqn 3) is shown for (A) GPU-accelerated Verlet integration with an Andersen thermostat in OpenMM, (B) GPU-accelerated Langevin integration in OpenMM, (C) GPU-accelerated particle-mesh Ewald aMD in Amber20 (i.e. pmemd.cuda), and (D) GPU-accelerated aMD integration in OpenMM.

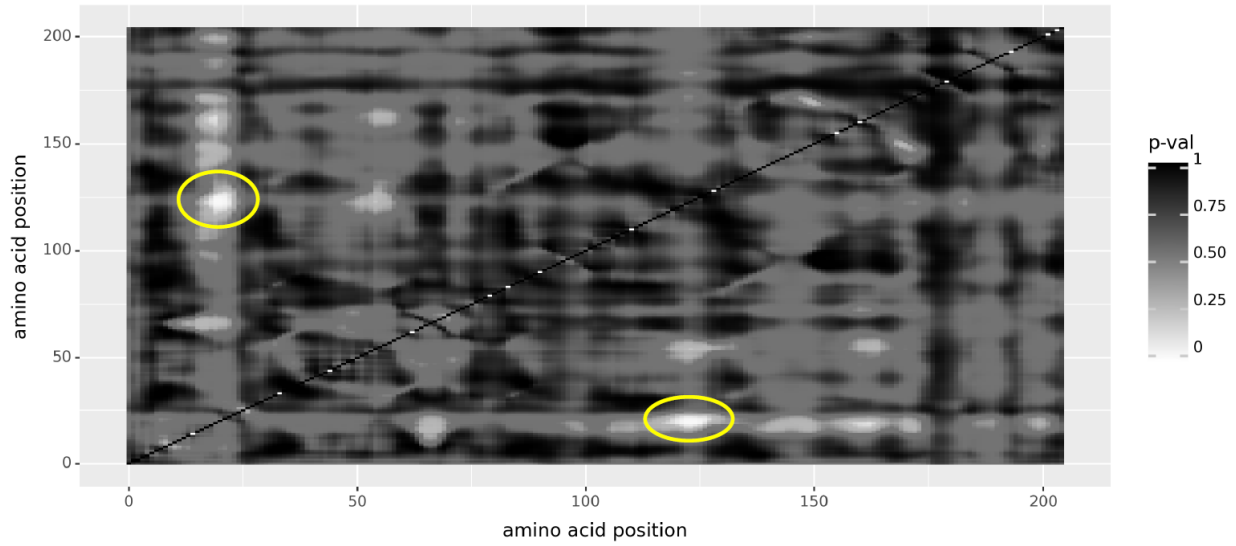


Figure S10 – Close-up of the heatmap of significant site interaction dynamics during allosteric inactivation of RTX cysteine protease in the absence of InsP6 (shown in Figure 5C). Significant coordination between sites is indicated by interaction p-values (corrected for false discovery rate, and highlighted with yellow circles in this symmetric plot) derived from a mixed-effects model ANOVA where atom fluctuations at sites I and j and the fixed effect in the model and the time during the molecular dynamics simulation is the random effect. Note: allosteric activation by InsP6 removes all signatures of significant site interaction.

Supplemental Movie File – video overview with dynamics of DNA-bound TATA binding protein and sorafenib drug-bound B-Raf kinase domain weighted in accordance with maximum mean discrepancy in atom fluctuation. https://people.rit.edu/gabsbi/img/videos/MMDmovie.mp4