

A novel nucleoside-enzyme pair for stringent cell-specific metabolic labeling of RNA

Sarah Nainar¹, Bonnie Cuthbert¹, Nathan M. Lim¹, David L. Mobley¹, Celia Goulding¹, and Robert Spitale^{1,*}

¹Department of Pharmaceutical Sciences, University of California—Irvine, Irvine, California 92697, United States
*rspitale@uci.edu

ABSTRACT

Abstract here.

Contents

1 Computational Methods	1
1.1 Setup	2
MD simulation parameters • Protein preparation • Docking	
1.2 Analysis	2
Defining the metastable binding modes • Distance to key residues • Hydrogen Bond Contacts • Comparisons to X-ray Crystal Structures	
2 Results	3
2.1 Binding Mode Stability	3
Distance to key residues • Hydrogen Bond Contacts • Comparison to X-ray Crystal Structures	
3 Discussion	6
References	7
4 Acknowledgements	8
5 Author contributions statement	8
6 Additional information	8
6.1 MSM and Clustering	8
6.2 Distance to key residues	8

1 Computational Methods

The starting structure used for the molecular dynamics (MD) simulations was taken from the crystal structure of human uridine-cytidine kinase 2 complex with the cytidine substrate (PDBID: 1UEJ)¹. In this study four different MD simulations were carried out, using two different binding modes for each of the respective ligands 2AZU and 2AZC. Here, we differentiate between the two binding modes by the orientation of the ribose moiety on the molecule. The "canonical" binding mode refers to the pose in which the azide (substituted on the 2' position) is pointed inwards into the binding site, see Figure 5a. We refer to this as the canonical binding mode as this pose is similar to the pose found in the crystal structure with the bound cytidine substrate (PDBID: 1UEJ). The "flipped" binding mode refer to the pose in which the azide is pointed outwards from the binding site, see Figure 5b.

A total of 500ns of MD simulation time was conducted for each binding mode by running 5 simulations each, where each copy of the simulation began from the same protein structure. The metastable binding modes sampled during our MD simulations were defined by constructing a Markov State Model (MSM) from our pool of MD simulation data and clustering with perron-cluster cluster analysis (PCCA). We then compared these metastable binding modes with x-ray crystal structures for 2AZU and 2AZC by computing the root-mean-square deviation (RMSD) between the ligand heavy atoms. We would like to note that the x-ray crystal structures of 2AZU and 2AZC were not used for setup or seen prior to conducting our MD simulations. Our simulation results, along with the x-ray crystal structures, support the hypothesis that the ligands must adopt the "flipped" binding mode for catalytic turnover.

1.1 Setup

1.1.1 MD simulation parameters

All simulations in this study were conducted using OpenMM v7.1.1² at T=300K and P=1atm. Here, we use timesteps of 4fs by employing the hydrogen mass repartitioning (HMR) scheme³. This scheme allows us to take larger timesteps by slowing down the fastest motions (i.e. hydrogen bond stretching) in our MD simulations. The HMR scheme constraints the bond length between hydrogens and their connected heavy atoms and reallocates mass from the connected heavy atom to the hydrogens. The protein-ligand systems were placed in a periodic box with explicit TIP3P water molecules using a 10Å solvent padding distance and counter ions (NaCl) were added using a concentration of 150mM. A 10Å cut-off distance was used for the particle-mesh Ewald method for computing long-range (e.g. electrostatic) interactions. Protein atoms were parameterized using the ‘amber99sbildn’ forcefields⁴ and the ligands were parameterized using GAFF2⁵ in which atomic charges were assigned using the AM1-BCC charge model⁶.

1.1.2 Protein preparation

To prepare the protein system (PDBID: 1UEJ) for MD simulations, we used PDBFixer⁷ to model in missing residues, add missing hydrogen atoms, and solvate our system. Sidechains were protonated in accordance with the pH=8.0 environment of the enzyme assay experiments⁸ and as described in other computational studies⁹. With the cytidine molecule bound, we energy minimized the protein-ligand complex for a maximum of 30,000 steps and followed with an equilibration protocol as follows. The equilibration protocol occurs in four 10 ps stages, whereby a progressively declining restraining force was used to help the protein-ligand system gradually relax. First, we apply a restraining force of 2.0 to the heavy atoms of the protein-ligand complex, simulate for 10ps using constant volume (NVT), and follow-up with 10ps at constant pressure (NPT). Next, we decrease the restraining force to 0.5 and then conduct an NPT simulation for 10ps. Last, we use a 0.1 restraining force on the alpha-carbons (protein backbone) and the ligand heavy atoms and then NPT simulate for 10ps.

1.1.3 Docking

After equilibration of the UCK2-cytidine complex, we applied HYBRID docking¹⁰ to dock our nucleoside analogs (2AZU and 2AZC) into the binding site. HYBRID differs from the standard docking approach such that the software will use the co-crystallized ligand as a reference point and attempt to fit the nucleoside analogs within the binding site by overlaying the docked ligands with the crystallographic ligand. From HYBRID docking, we generated up to 50 different conformers for each nucleoside analog and then followed the same equilibration protocol described previously. After equilibration, we found that the conformers tended to converge into two groups: 1 conformer which resembled the “canonical” binding mode and another which had the ribose moiety “flipped”. We then selected 1 conformer from each representative group and carried these forward for our production NPT 100ns MD simulations (no restraints).

1.2 Analysis

For each nucleoside analog (2AZC and 2AZU), we ran five 100ns MD simulations for each of the two binding modes (canonical and flipped). They are denoted as 2AZC_{canc}, 2AZC_{flip}, 2AZU_{canc}, and 2AZU_{flip}. Collectively, over the course of the 100ns of simulation time, the root-mean-square deviation (RMSD) for the ligand atoms began to stabilize after 25ns. Thus, we discard trajectory frames from 0-25ns as additional equilibration time and only perform further analysis from the 25ns-100ns time frame Fig.2a. The RMSD is calculated by:

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^n d_i^2} \quad (1)$$

where d_i represents the distance between the n atom pairs. For construction of our Markov State models (MSM), we use the PyEMMA v2.5.5¹¹ toolkit. For residue contact analyses we use the MDTraj v1.9.1¹² and VMD v1.9.3¹³ toolkits.

1.2.1 Defining the metastable binding modes

In order to define our metastable binding modes and visualize their structures, we construct a Markov State model (MSM)¹⁴ from our five separate MD simulations. Our features for constructing the MSMs consists of the distance between the closest heavy atoms on the ligand and the following 9 residues: ASP62, PHE83, ASP84, TYR112, PHE114, HIS117, ILE137, ARG166, and ARG176. These residues were selected as they have been noted in the literature to play important roles in binding⁹. From this feature space, we apply the time-lagged independent component analysis (TICA) method using a lagtime of 1ns. TICA transforms our 9 dimensional feature space to a new set of reaction coordinates which maximizes the autocorrelation of the transformed coordinates¹⁵. In other words, TICA allows us to extract the slow order parameters and project them into a lower dimensional space; here, we use the first two TICA coordinates (Fig. 1a). Then, we apply k-means clustering to discretize our trajectory frames into discrete microstate and project them into TICA space (denoted by individual Xs). Following, we use perron-cluster cluster analysis (PCCA)¹⁶ to assign each microstate to a metastable macrostate (denoted by color in Fig.1b).

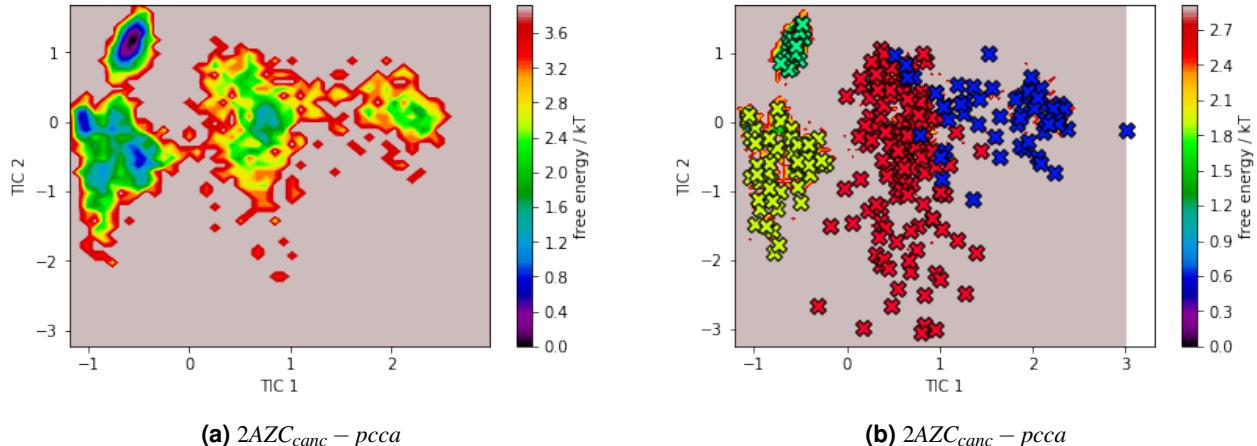


Figure 1. Caption for this figure with two images

From each of our assigned macrostates, we randomly sample 100 frames and then visualize the frame which minimizes the RMSD to the crystallographic ligand.

1.2.2 Distance to key residues

Using the ‘compute_contacts’ tool from MDTraj, we compute the distance between the closest heavy atoms in the ligand and 4 residues: ASP62, TYR112, HIS117 and ARG176 (Fig. 3). These residues were chosen in particular as ASP62 is known to be the catalytic residue, while TYR112, HIS117, and ARG176 are believed to play a key role in substrate specificity between uridine and cytidine as they have been found to bind to the nucleobase moiety⁹. Here, we calculate the frequency in which the distance between the ligand and the residues are less than or equal to 3.0Å, which we define as the minimum distance needed to form an interactive bond.

1.2.3 Hydrogen Bond Contacts

We compute the frequency of hydrogen bond contacts between the ligands and surrounding residues using the HBonds plugin v1.2 in VMD 1.9.3¹³, shown in Figure 4. The criterion used for defining formation of a hydrogen bond is that the cutoff distance between a hydrogen bond donor and acceptor must be less than 3.0Å and the angle is less than 20 degrees. Each colored bar corresponds to the hydrogen bond frequency from an individual MD simulation.

1.2.4 Comparisons to X-ray Crystal Structures

Here, we compared our metastable binding modes from our MD simulations against each subunit found in the experimental x-ray crystal structures and list the values against the subunit which minimizes the computed RMSD. To compare the metastable binding modes from our MD simulations with experimental x-ray crystal structures, we first align the two structures using the protein backbone. Particularly, we align by the protein backbone using residues 19 to 229 but exclude residues 48 to 52 as these were missing in the x-ray crystal structures. Once aligned by the protein backbone, we then compute the RMSD between crystallized ligand and the matching ligand heavy atoms from our defined metastable binding modes.

2 Results

2.1 Binding Mode Stability

Table 1. Calculated RMSDs for ligand heavy atoms from 25-100ns MD simulations

	RMSD	RMSD	RMSD	RMSD	RMSD	AVG	SEM
2AZC_canc	6.66	1.36	1.14	4.12	2.06	3.07	<i>1.04</i>
2AZU_canc	0.94	1.17	2.89	1.31	1.41	1.54	<i>0.35</i>
2AZC_flip	1.83	1.66	4.01	2.61	4.58	2.94	<i>0.58</i>
2AZU_flip	2.43	3.08	2.35	2.23	2.61	2.54	<i>0.15</i>

The averages and the standard errors of the mean (SEM) for the RMSD of each ligand binding mode are shown in Fig.2 and Table1. They are 3.07 ± 1.04 , 1.54 ± 0.35 , 2.94 ± 0.58 , and 2.54 ± 0.15 for $2AZC_{canc}$, $2AZU_{canc}$, $2AZC_{flip}$, and $2AZU_{flip}$,

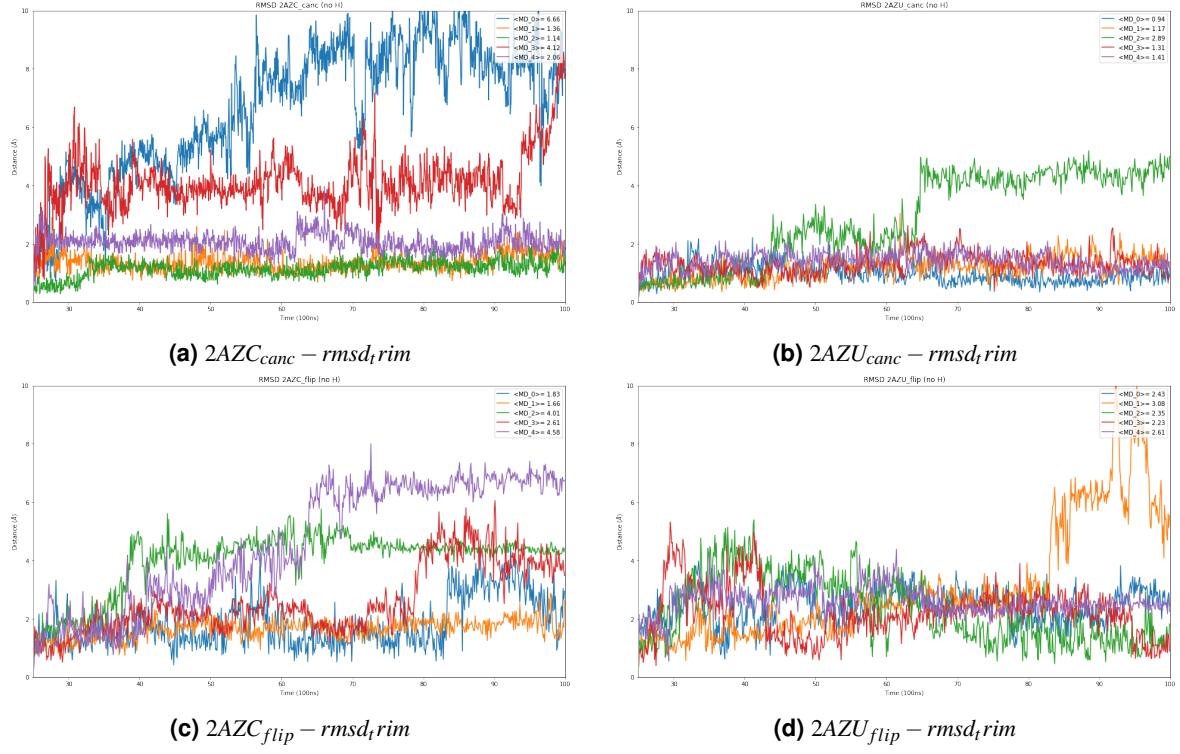


Figure 2. Calculated RMSDs for ligand heavy atoms from 25-100ns MD simulations

respectively.

Comparing the RMSD of the canonical binding modes for 2AZC and 2AZU, we find that the 2AZU ligand to be much more stable than the 2AZC ligand. In $\frac{2}{5}$ of our MD simulations for 2AZC_{canc} (Fig. 2a), the ligand comes nearly unbound after 30ns; while for 2AZU_{canc} (Fig. 2b) the ligand nearly unbinds in only $\frac{1}{5}$ of the simulations. When comparing 2AZC_{flip} and 2AZU_{flip}, the average RMSD suggests both are about equally stable. However, this particular binding mode appears to show slightly more instability than the canonical binding mode with an average RMSD of 2.54 and 2.94 for the 2AZC_{flip} and 2AZU_{flip}, respectively.

2.1.1 Distance to key residues

In Figure 3, we plot the distance between the closest ligand heavy atoms and the 4 residues: ASP62 (blue), TYR112 (orange), HIS117 (green) and ARG176 (red). One plot from each respective binding mode was chosen from the simulation in which contact with the catalytic residue ASP62 was highest. Additional contact distance plots from the other MD simulations can be found in the supporting information. See Figure 10 for 2AZC_{canc}, Figure 11 for 2AZU_{canc}, Figure 12 for 2AZC_{flip}, and Figure 13 for 2AZU_{flip}.

For the canonical binding mode, we see a maximum contact frequency of 3.9% (2AZC) and 2.5% (2AZU) in which the ligand is $d < 3.0\text{\AA}$ away from the catalytic residue ASP62. Figure 3a shows that the 2AZC_{canc} ligand in the canonical binding mode only rarely comes into contact with ASP62 but is in stable contact with HIS117 and ARG176. Interestingly, Figure 3b illustrates that even though the 2AZU_{canc} ligand appears to be stably bound to the key residues (TYR112, HIS117, and ARG176) the ligand is too far away from ASP62.

In stark contrast, we see an enormous increase in contact frequency when simulating the flipped binding mode: 27.9% for 2AZC_{flip} and 80.4% for 2AZU_{flip}. Figure 3c shows the 2AZC_{flip} ligand forms stable contacts with the catalytic residue ASP62 and residues TYR112/HIS117 but does not contact ARG176. Similarly, we see the same stable contacts being formed for 2AZU_{flip} in Figure 3d.

2.1.2 Hydrogen Bond Contacts

In order to gain better insight on the interactions required for catalysis, we profile the hydrogen bond contacts between the ligand and surrounding protein residues. When comparing the canonical binding modes, we see that 2AZC_{canc} forms hydrogen bond contacts mostly with ASP84, PHE114, HIS117, and ARG176; but no virtually no contact with the catalytic residue ASP62 (Fig. 4a). For 2AZU_{canc} (Fig. 4b), we see frequent contacts with TYR112, ARG166, ASP170, and ARG176; also, virtually no

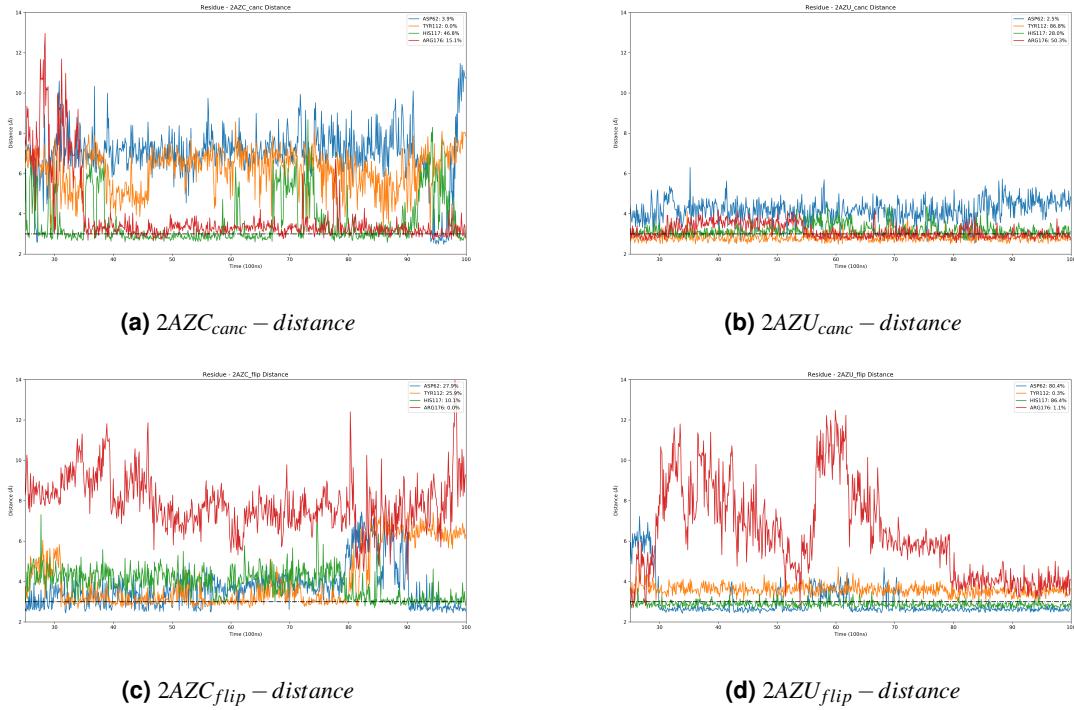


Figure 3. contact-distance

contact with the catalytic residue ASP62. Given that we do not see hydrogen bonding between the ligand and ASP62, this supports the hypothesis that the canonical binding mode may not be the pose required for catalysis.

On the other hand, we see a much higher rate of contact to ASP62 with the flipped binding mode, which supports the hypothesis that this binding mode may be what the ligand adopts for catalytic turnover. In Figure 4c, we see 20% contact with ASP62 and moderate rates of contact with ASP84, PHE114, HIS117, ARG169, and ARG174. For $2AZC_{flip}$, contact with ASP62 (yellow/green/red) appears to be related to hydrogen bonding with ASP84 (yellow/green), TYR112(yellow) and ARG174 (yellow). In Figure 4d, we see 70% contact with ASP62 and moderate contacts with THR29, ASP84, TYR112, and HIS117. For $2AZU_{flip}$, contact with ASP62(green/red) appears to be related to hydrogen bonding with THR29(red), ASP84(green) and HIS117(green/red).

2.1.3 Comparison to X-ray Crystal Structures

	$2AZC_{xtal} - A$	BB_{RMSD}
$2AZC_{canc} - pcca3$	0.72	1.20
$2AZC_{flip} - pcca3$	5.04	1.47
	$2AZU_{xtal} - F$	BB_{RMSD}
$2AZU_{canc} - pcca1$	4.84	2.67
$2AZU_{flip} - pcca0$	2.31	2.81

Here, we compared our defined metastable binding modes from our MD simulations to our x-ray crystal structures by computing the RMSD between matching heavy atoms in the respective ligands. The minimum computed RMSD of $2AZC_{canc}$ was 0.72 Å and was 5.04 for $2AZC_{flip}$. Figure 5a illustrates the crystallized ligand (cyan) directly overlays with the metastable binding mode obtained from our MD simulation (green), except for the additional phosphate group found on the crystallized ligand. The minimum computed RMSD of $2AZU_{flip}$ from our MD simulation against the x-ray crystal structure was 2.31 Å; while for $2AZU_{canc}$ was 4.84 Å. Figure 5b shows that orientation of the nucleobase moiety from our MD simulation (green) differs from the crystallized ligand (cyan). Additionally, the ribose moiety sits nearly perpendicular to the nucleobase moiety in the binding mode from MD; while in the crystal structure, the nucleobase and ribose moieties are nearly in the same plane.

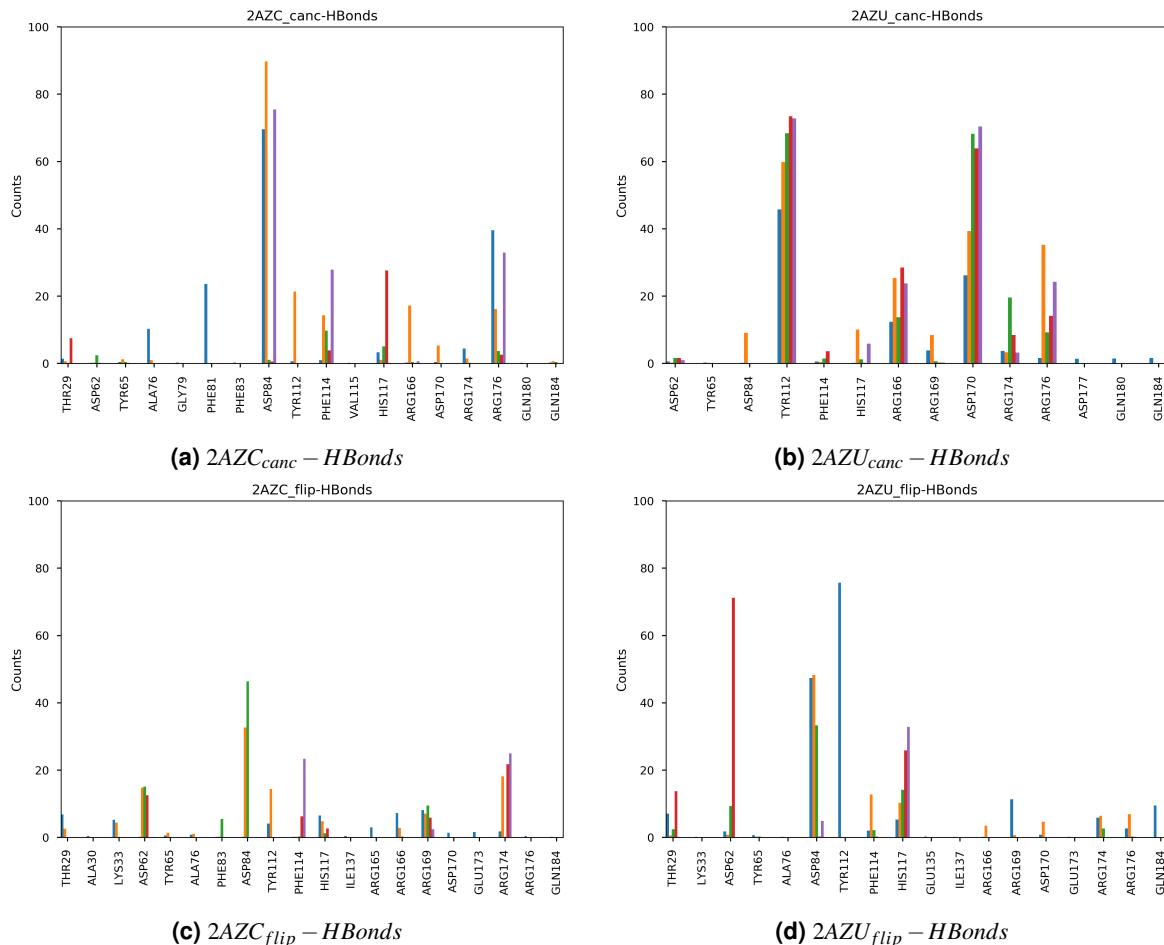


Figure 4. HBonds

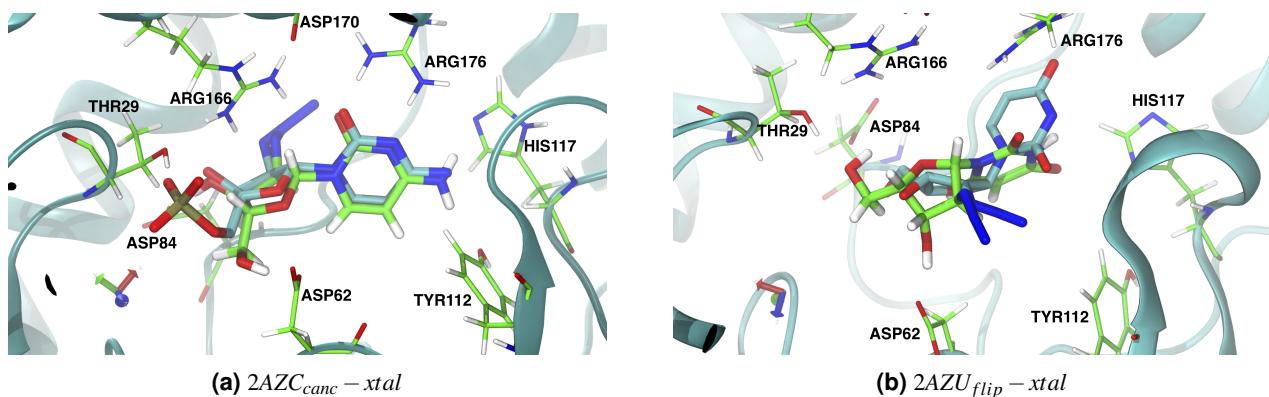


Figure 5. Caption for this figure with two images

3 Discussion

Prior to running our MD simulations, we had no experimental evidence which suggests the existence of an alternative binding mode. We first discovered evidence of a flipped binding mode after our HYBRID docking approach with our short equilibration protocol. After conducting our MD simulations, we later confirmed the existence of the flipped binding mode by comparing our metastable binding modes against the x-ray crystal structures.

From this study, we believe that the canonical binding mode may not be the binding mode the ligand adopts for catalytic turnover. Considering the crystallized 2AZC ligand was found to contain an additional phosphate group—indicative of the

post-catalytic state—and this coincided with our findings from the MD simulations, we have strong evidence that the canonical binding mode represents the pose after catalysis. This is illustrated by the extremely low RMSD value (0.72Å) against the crystal structure for 2AZC_{canc} and the low rate of contact with the catalytic residues ASP62 for both 2AZC_{canc} and 2AZU_{canc} (Fig. 3a and Fig. 3b). Contrary to the literature⁹, our simulations suggest that ARG176 plays no role in substrate specificity, where we see roughly equivalent hydrogen bonding frequency for both 2AZU_{canc} and 2AZC_{canc}. Residue ARG176 seems to only serve a role in stabilizing the ligand in the binding site when in the canonical binding mode. From Figure 12, we observe several moments in which the ligand 2AZC_{flip} comes into closer contact with ARG176 we see the distance with ASP62 increases and vice versa.

Given that we see a much higher contact frequency with ASP62 for the flipped binding mode over the canonical binding mode, we believe that the ligand must adopt the flipped conformation to stably contact ASP62 for catalytic turnover. Based on the hydrogen bond contact profiles (Fig. 4), we believe ASP84 plays a critical role in binding both nucleoside analogs and additional contacts with residues TYR112/HIS117 appear to be necessary for facilitating the interaction with ASP62 (Fig. 3c and Fig. 3d) when in the flipped binding mode. The binding of 2AZC_{flip} appears to specifically require contact with either ARG174/ARG176 residues and potentially forms interactions with TYR112 for additional stability (Fig. 4c). In contrast, binding of 2AZU with TYR112 does not appear to facilitate contact with ASP62 (Fig. 4d) but does appear important for stabilizing the canonical binding mode (Fig. 4b). For specificity of binding 2AZU, it appears that hydrogen bonding with HIS117 is required and that contact with THR29 serves for additional stability (Fig. 4d).

Overall, these results provide strong supporting evidence that the flipped binding mode represents the conformer of the ligand before catalysis; while the canonical binding mode represents the post-catalytic state.

References

- Suzuki, N. N., Koizumi, K., Fukushima, M., Matsuda, A. & Inagaki, F. Structural basis for the specificity, catalysis, and regulation of human uridine-cytidine kinase. *Structure* **12**, 751–764 (2004).
- Eastman, P. *et al.* Openmm 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology* **13**, e1005659 (2017).
- Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-time-step molecular dynamics through hydrogen mass repartitioning. *J. chemical theory computation* **11**, 1864–1874 (2015).
- Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins: Struct. Funct. Bioinforma.* **78**, 1950–1958 (2010).
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. computational chemistry* **25**, 1157–1174 (2004).
- Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. am1-bcc model: II. parameterization and validation. *J. computational chemistry* **23**, 1623–1641 (2002).
- Eastman, P. Pdbfixer version 1.5. <https://github.com/pandegroup/pdbfixer> (2018).
- Tomoike, F., Nakagawa, N., Kuramitsu, S. & Masui, R. A single amino acid limits the substrate specificity of thermus thermophilus uridine-cytidine kinase to cytidine. *Biochemistry* **50**, 4597–4607, DOI: [10.1021/bi102054n](https://doi.org/10.1021/bi102054n) (2011). PMID: 21539325, <https://doi.org/10.1021/bi102054n>.
- Tanaka, W. *et al.* Molecular mechanisms of substrate specificities of uridine-cytidine kinase. *Biophys. physicobiology* **13**, 77–84 (2016).
- McGann, M. Fred and hybrid docking performance on standardized datasets. *J. computer-aided molecular design* **26**, 897–906 (2012).
- Scherer, M. K. *et al.* Pyemma 2: A software package for estimation, validation, and analysis of markov models. *J. chemical theory computation* **11**, 5525–5542 (2015).
- McGibbon, R. T. *et al.* Mdtraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys. journal* **109**, 1528–1532 (2015).
- Humphrey, W., Dalke, A. & Schulten, K. Vmd: visual molecular dynamics. *J. molecular graphics* **14**, 33–38 (1996).
- Prinz, J.-H. *et al.* Markov models of molecular kinetics: Generation and validation. *The J. chemical physics* **134**, 174105 (2011).
- Pérez-Hernández, G., Paul, F., Giorgino, T., De Fabritiis, G. & Noé, F. Identification of slow molecular order parameters for markov model construction. *The J. chemical physics* **139**, 07B604_1 (2013).

16. Röblitz, S. & Weber, M. Fuzzy spectral clustering by pcca+: application to markov state models and data classification. *Adv. Data Analysis Classif.* **7**, 147–179 (2013).

4 Acknowledgements

Christopher Bayly and Gaetano Calabro at OpenEye Scientific Software for development of the MD simulation protocol.

5 Author contributions statement

S.N. conducted the experiment(s), B.C crystallized the system, and N.M.L conducted the computational experiments. All authors reviewed the manuscript.

6 Additional information

6.1 MSM and Clustering

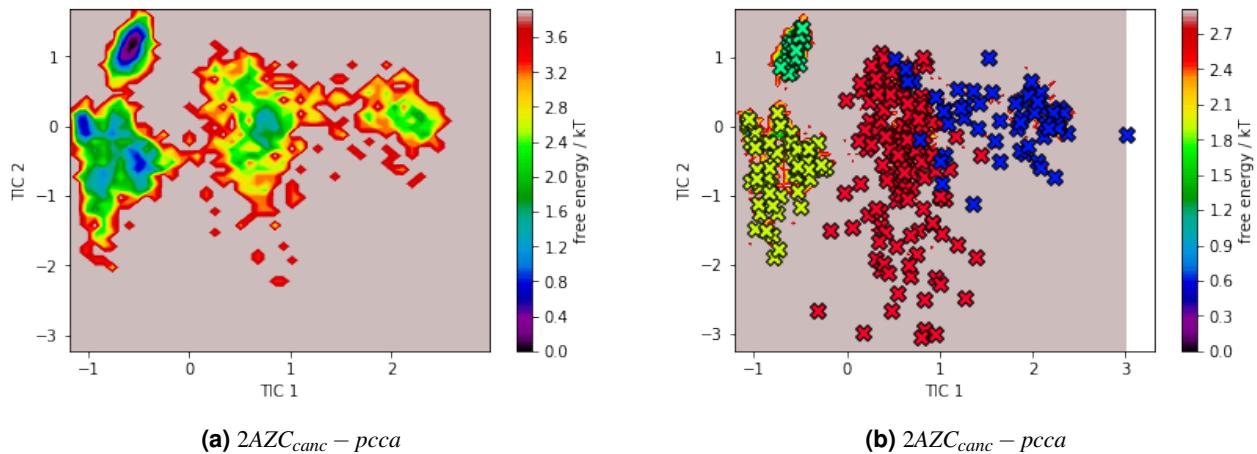


Figure 6. Caption for this figure with two images

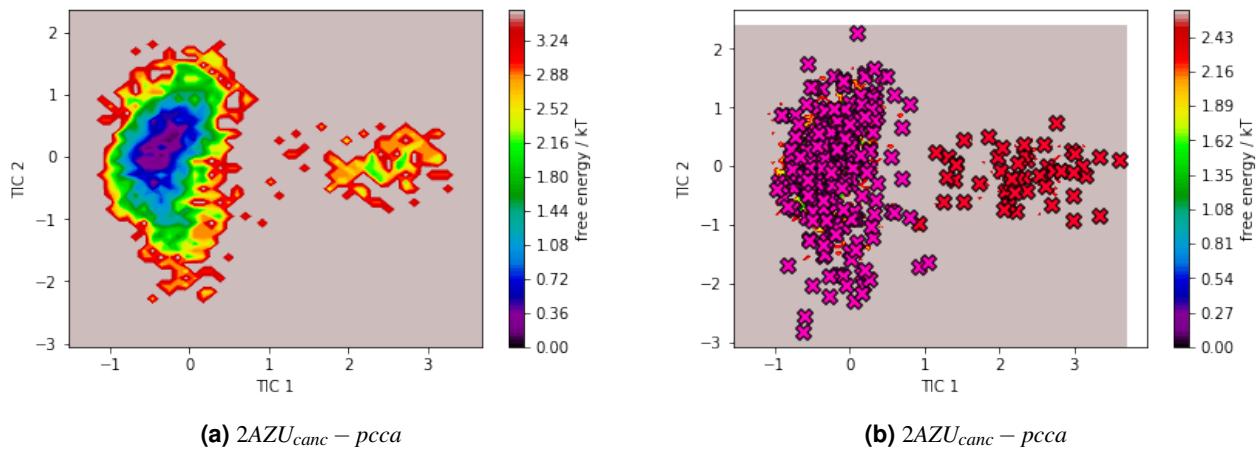


Figure 7. Caption for this figure with two images

6.2 Distance to key residues

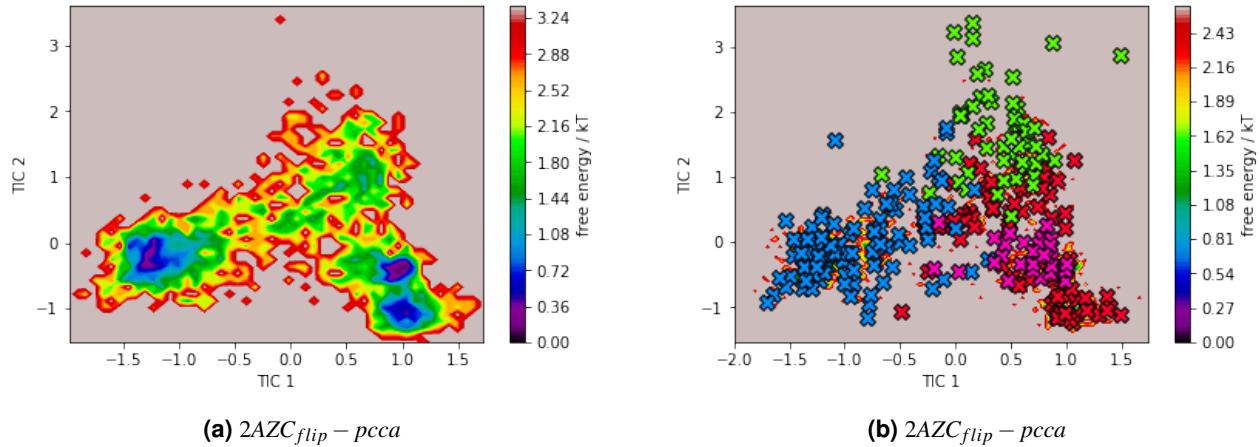


Figure 8. Caption for this figure with two images

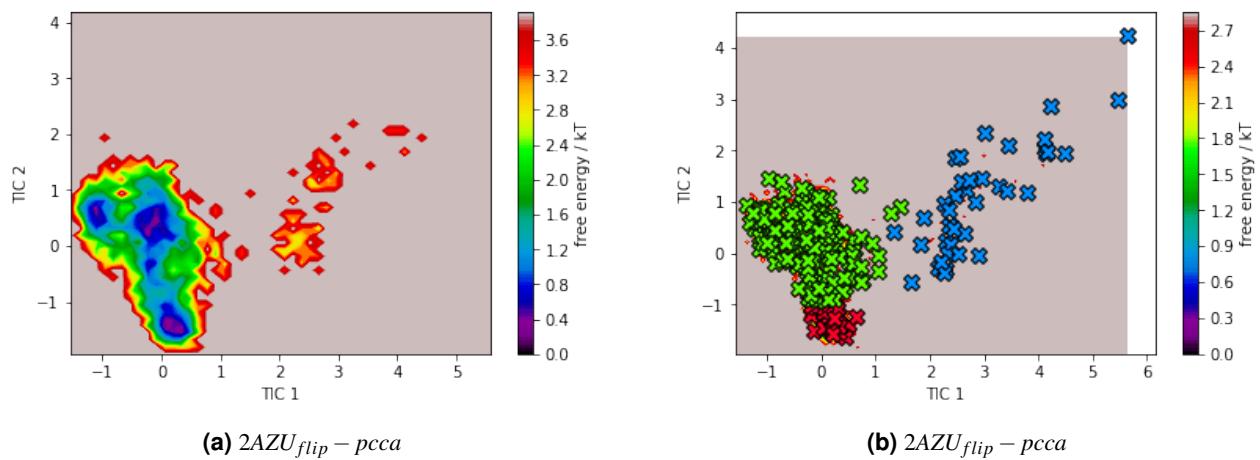


Figure 9. Caption for this figure with two images

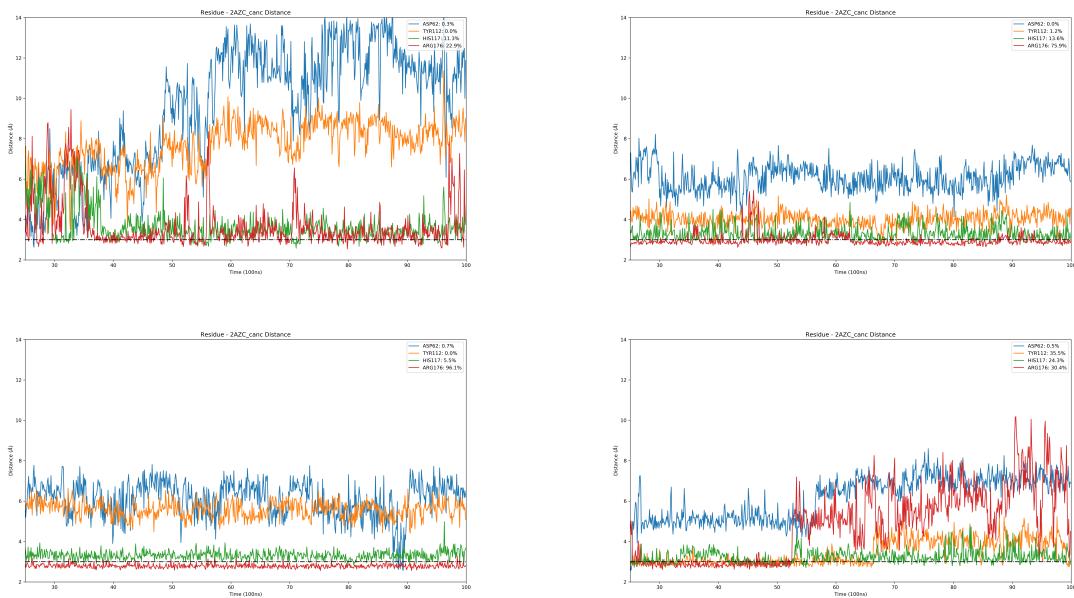


Figure 10. 2AZC_{canc} – distance

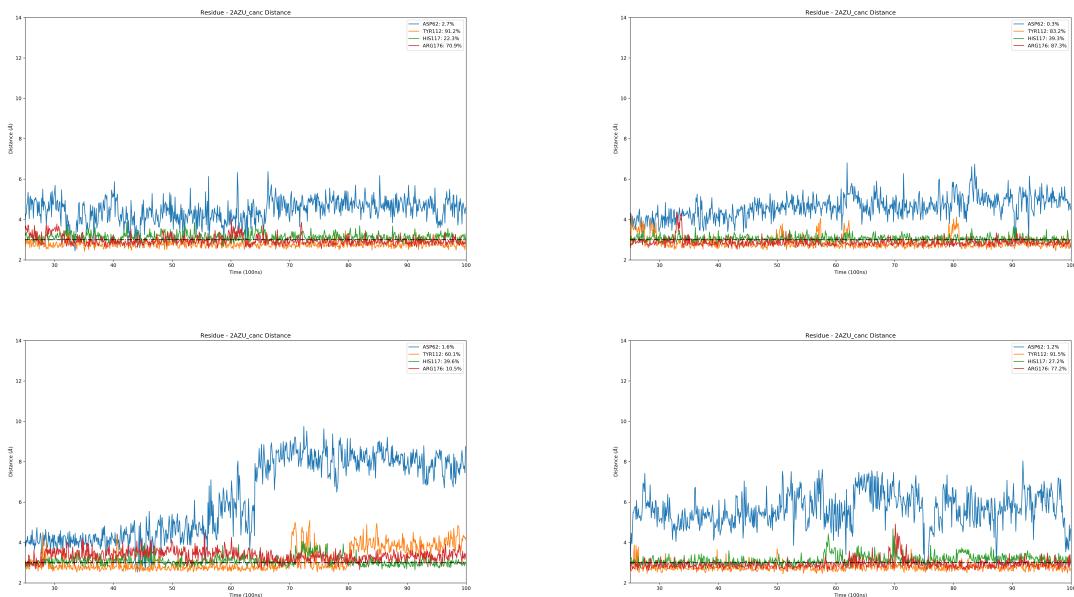


Figure 11. 2AZU_{canc} – distance

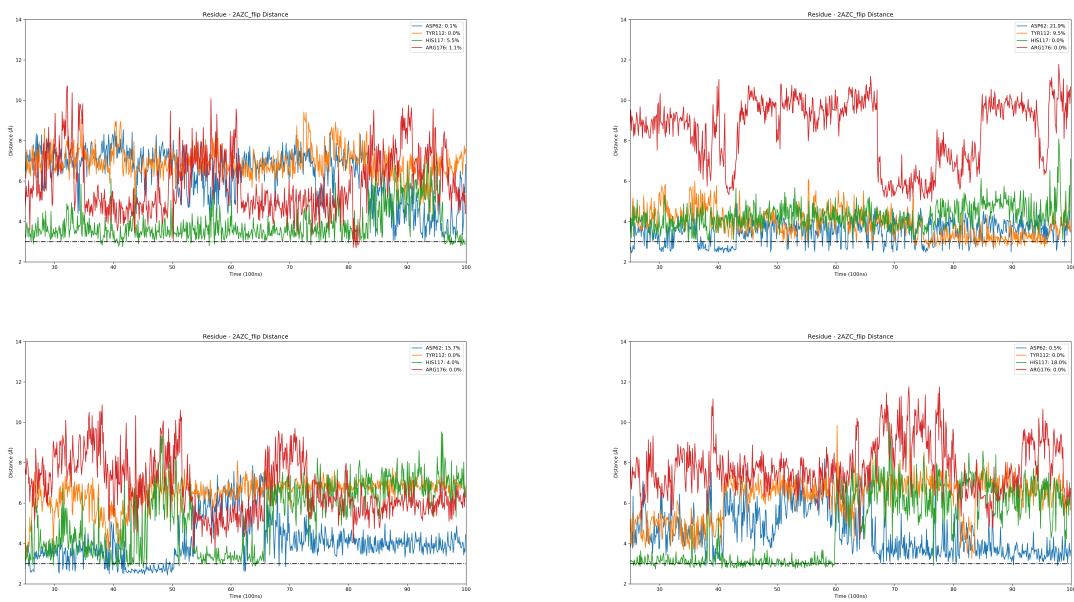


Figure 12. $2AZC_{flip} - distance$

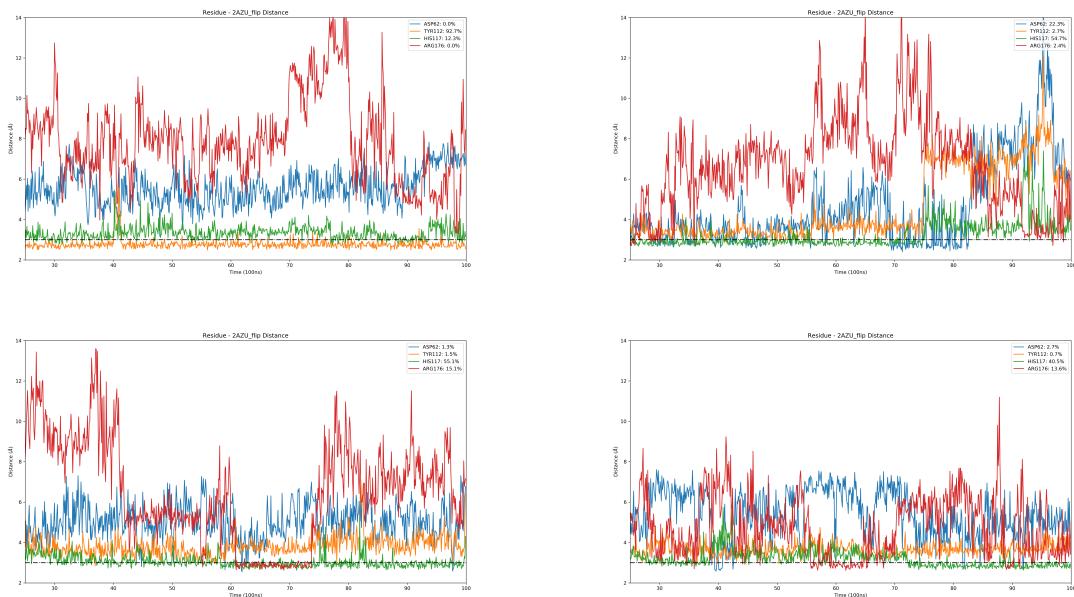


Figure 13. $2AZU_{flip} - distance$