

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

Contents

1 Computational Methods	1
1.1 Setup	1
MD simulation parameters • Protein preparation for MD simulations • Docking the nucleoside analogs	
1.2 Analysis	2
Defining the metastable binding modes • Distance to key residues • Hydrogen Bond Contacts • Comparisons to Xtal	
2 Results	3
2.1 Binding Mode Stability	3
Distance to key residues • Hydrogen Bond Contacts • RMSD calculation to Xtal chains	
3 Discussion	6
4 Additional information	6
4.1 MSM and Clustering	6
4.2 Distance to key residues	6

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). 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 sugar group on the molecule; whereby the "canononical" binding mode refers to the same orientation of the sugar as found in the crystal structure and the "flipped" binding mode being the opposite. A total of 500ns of MD simulation time was conducted for each binding mode by running 5 simulations for each binding mode, 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 experimental crystal structures for 2AZU and 2AZC by computing the root-mean-square deviation between the ligand heavy atoms. 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 v7.1.1⁷ at T=300K and P=1atm. Here, we use timesteps of 4fs by employing the hydrogen mass repartitioning scheme⁷. This scheme allows us to take larger timesteps by slowing down the fastest motions (i.e. hydrogen bond stretching) by constraining the bond length between hydrogens and their connected heavy atoms and reallocating 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. Protein atoms were

parameterized using the amber99sbildn forcefields and ligands were parameterized using GAFF2 in which atomic charges were assigned using the AM1-BCC charge model

1.1.2 Protein preparation for MD simulations

To prepare the protein system 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 environment⁷ of the enzyme assay experiments. With the cytidine molecule bound, we energy minimized the protein-ligand complex for a maximum of 30,000 steps and followed with the equilibration protocol as follows. The relaxation 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 the nucleoside analogs

After equilibration of the UCK2-cytidine complex, we applied HYBRID docking to dock our nucleoside analogs (2AZU and 2AZC) into the binding site. HYBRID differs from the standard docking approach such that it 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 ligand and then followed the same equilibration protocol described previously. After visual inspection, we found that the conformers tended to converge into two groups: 1 conformer which resembled the "canonical" binding mode and another which had the sugar group "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

RMSD-Calulation Jupyter Notebook

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 Fig.???. 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 v2.5.5 toolkit. For residue contact analysis we use the MDTraj v1.9.1 toolkit.

1.2.1 Defining the metastable binding modes

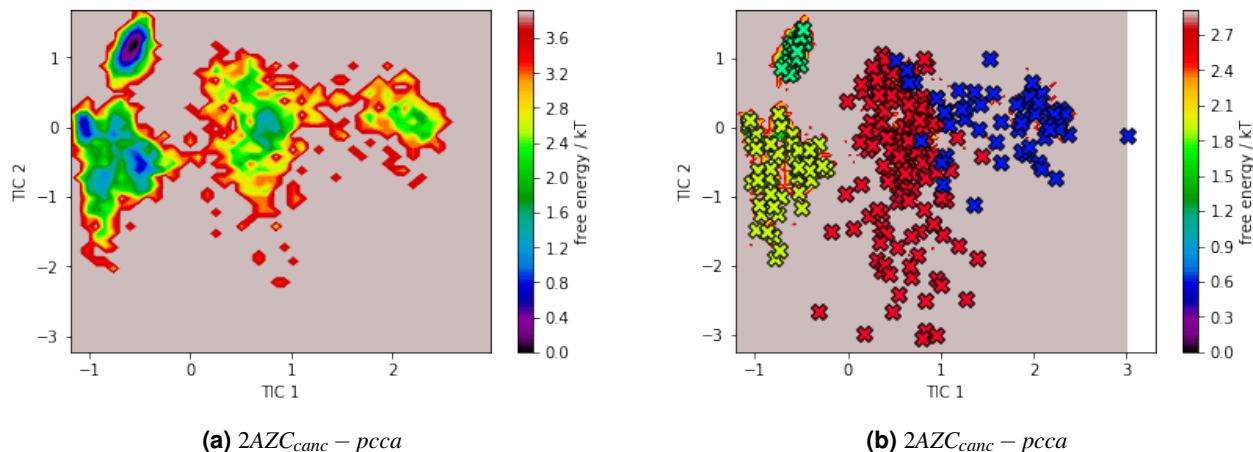


Figure 1. Caption for this figure with two images

In order to define our metastable binding modes and visualize their structures, we construct a Markov State model (MSM) 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 binding⁷. From this feature space, we apply time-lagged independent component analysis (TICA) 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 coordinates. 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. 6a). Then, we apply k-means clustering to discretize our trajectory frames into discrete microstate and project them into TICA space (denoted by the individual X). Following, we use perron-cluster cluster analysis (PCCA) to assign each microstate to a metastable macrostate (denoted by color in Fig.6b). 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

Contact-Distance Jupyter Notebook

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 moiety⁷. Here, we calculate the frequency in which the distance between the ligand and the residues are less than or equal to 3.5Å, 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 surround residues using the HBonds plugin v1.2 in VMD 1.9.3, 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 Xtal

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 structure. Once aligned by the protein backbone, we then compute the RMSD between the ligand heavy atoms using our defined metastable binding modes from our PCCA analysis. The x-ray crystal structure of 2AZC represents the system in it’s post-catalytic state and has a phosphate group attached, thus we exclude the phosphate group from our RMSD calculations as our simulations represent the pre-catalytic state. The x-ray crystal structure of 2AZU represents the system in it’s pre-catalytic state and has no additional phosphate group, so no atoms were excluded from the RMSD calculation.

2 Results

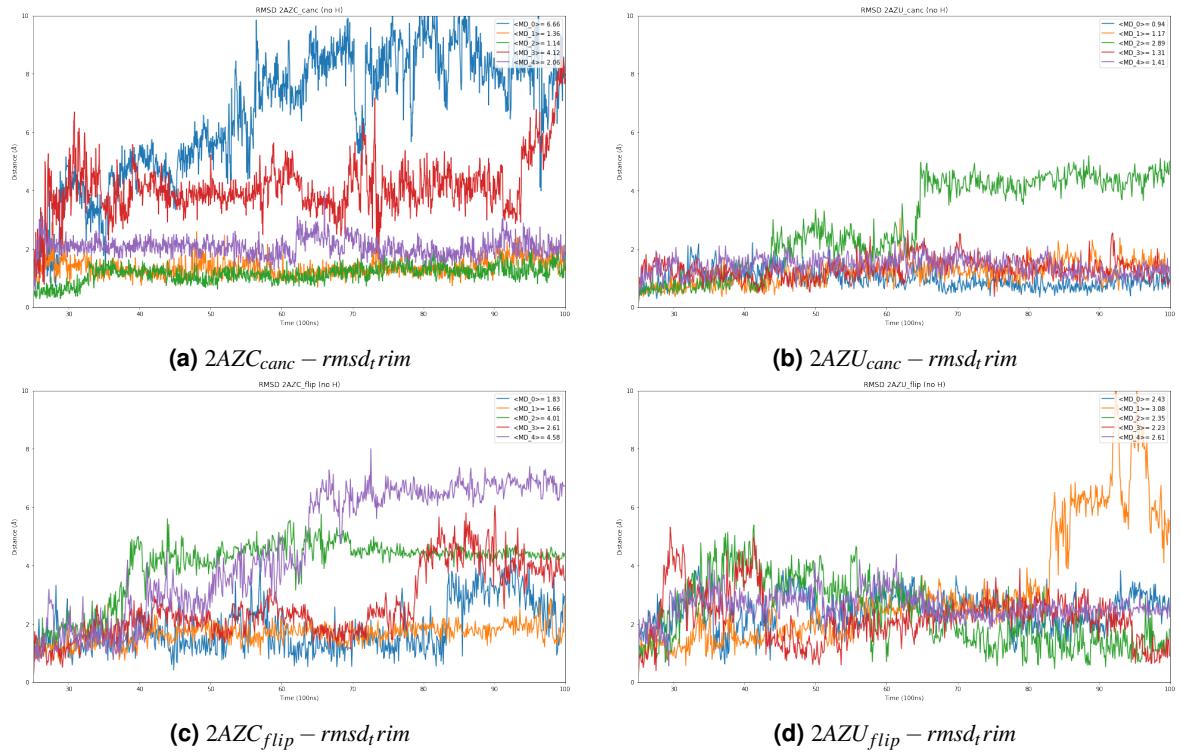
2.1 Binding Mode Stability

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.?? 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 $2AZU_{flip}$, 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.



Prior to running our MD simulations, we had no experimental evidence for the existence of the flipped binding mode. The flipped binding mode was first discovered after our HYBRID docking approach with our short equilibration protocol. 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 plots from the other MD simulations for binding modes 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 $2AZC$ 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 ligand appears to be stably bound to the key residues (TYR112, HIS117, and ARG176) the ligand may be too far away from ASP62 for catalysis.

In stark contrast, we see an enormous increase in contact frequency when simulating the flipped binding mode; 27.9% ($2AZC$) and 80.4% ($2AZU$). Figure 3c shows the $2AZC$ 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 $2AZU$ in Figure 3d.

These results suggest that the canonical binding mode may not be the binding mode the ligand adopts for catalytic turnover. Instead, our results indicate that the ligand must adopt the flipped conformation and that contact with residues TYR112/HIS117 appear to be necessary for catalytic turnover. Here, residue ARG176 seems to only serve a role in stabilizing the ligand in the binding site when in the canonical conformation. We believe the canonical conformation may represent the post-catalytic binding mode and ARG176 plays a role in the post-catalytic process. From Figure 12, there are several moments when the ligand $2AZC_{flip}$ comes into closer contact with ARG176, the distance with ASP62 increases and vice versa.

2.1.2 Hydrogen Bond Contacts

In order to gain some 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

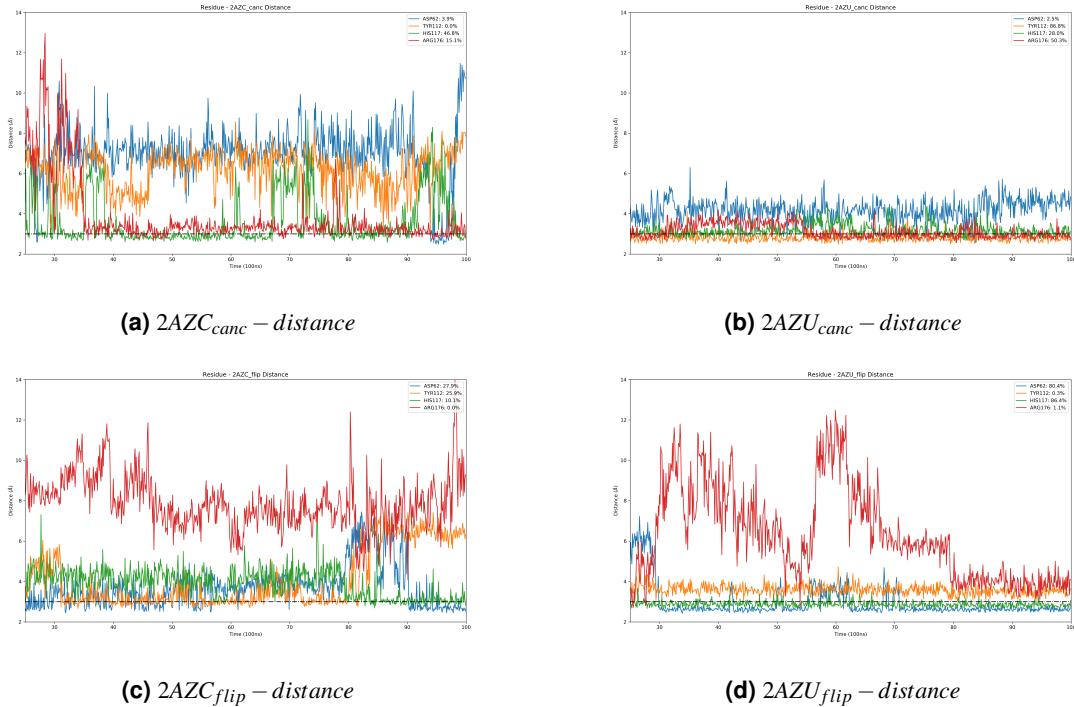


Figure 3. contact-distance

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 frequent contacts with TYR112, ARG166, ASP170, and ARG176; also, virtually no 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.

For the flipped binding mode, we see a much higher rate of contact to ASP62 which supports the hypothesis that this binding mode may be what the ligand adopts for catalytic turnover. In Figure ??, 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).

From our results, we hypothesize that ASP84 plays a critical role in binding the both substrates for catalysis, given that we see high rates of contact for both 2AZU and 2AZC. The binding of 2AZC appears to specifically require contact with either ARG174/ARG176 residues and potentially contact with TYR112 for additional stability. In contrast, binding of 2AZU with TYR112 does not appear to prepare the ligand for catalysis (based on the flipped binding mode) but appears to stabilize the binding mode post-catalysis (based on the canonical binding mode). For specificity of binding 2AZU, it appears that hydrogen bonding with HIS117 is required and potentially contact with THR29 for additional stability.

2.1.3 RMSD calculation to Xtal chains

	2AZC _{xtal} - B	BB _{RMSD}
2AZC _{canc} - pcca2	3.69	1.29
2AZC _{flip} - pcca3	4.97	1.42
	2AZU _{xtal} - F	BB _{RMSD}
2AZU _{canc} - pcca1	4.46	2.73
2AZU _{flip} - pcca0	2.31	2.74

X-ray crystal structures for 2AZU and 2AZC were obtained and visualized after performing our MD simulations. Here, 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 crystal structure containing the 2AZU ligand was found

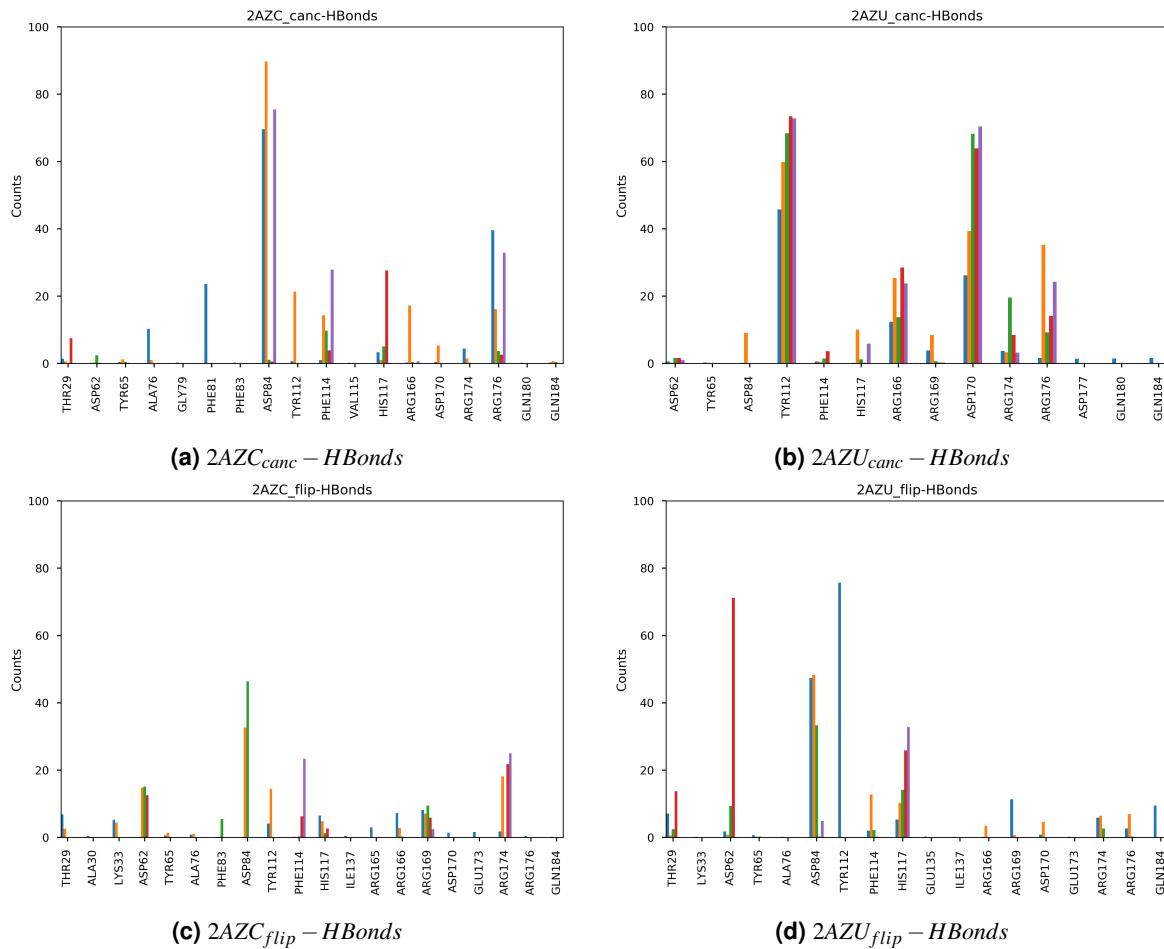


Figure 4. HBonds

to bind similarly in a similar conformation as the flipped binding mode from our MD simulation. The minimum computed RMSD of $2AZU_{flip}$ from our MD simulation against the x-ray crystal structure was 2.31\AA ; while for $2AZU_{canc}$ was 4.46\AA . On the other hand, the crystal structure obtained for $2AZC$ was found to resemble the post-catalytic state, as the bound ligand contained an attached phosphate group. The minimum computed RMSD of $2AZC_{canc}$ was 3.69\AA and was 4.97 for $2AZC_{flip}$. 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.

3 Discussion

Support hypotheses:

- Evidence of flipped binding mode
- $2AZC > 2AZU$
- Number of HBond contacts?
- RMSD for 20 - 100 ns?

Highlight:

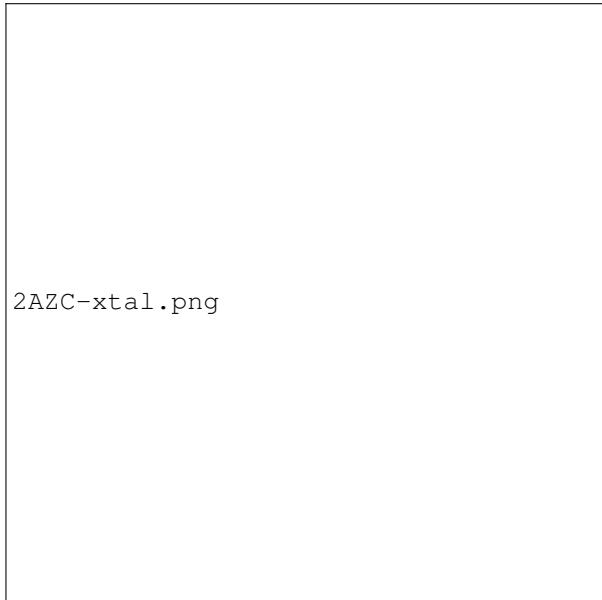
- ASP62 - TYR112 coupling importance for binding

The Discussion should be succinct and must not contain subheadings.

4 Additional information

4.1 MSM and Clustering

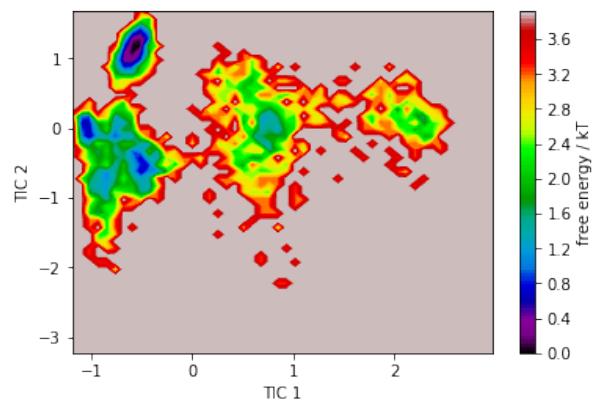
4.2 Distance to key residues



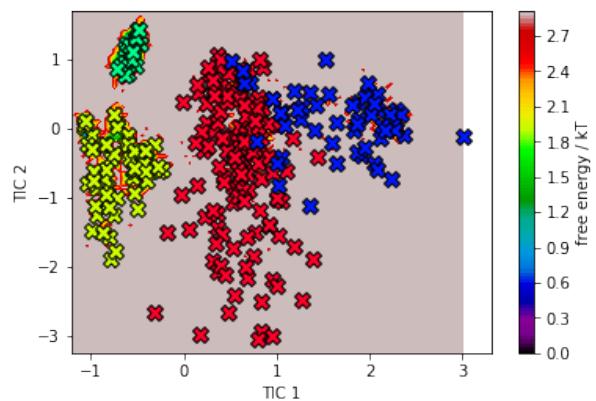
(a) $2AZC - xtal$

(b) $2AZU - xtal$

Figure 5. Caption for this figure with two images



(a) $2AZC_{canc} - pcca$



(b) $2AZC_{canc} - pcca$

Figure 6. Caption for this figure with two images

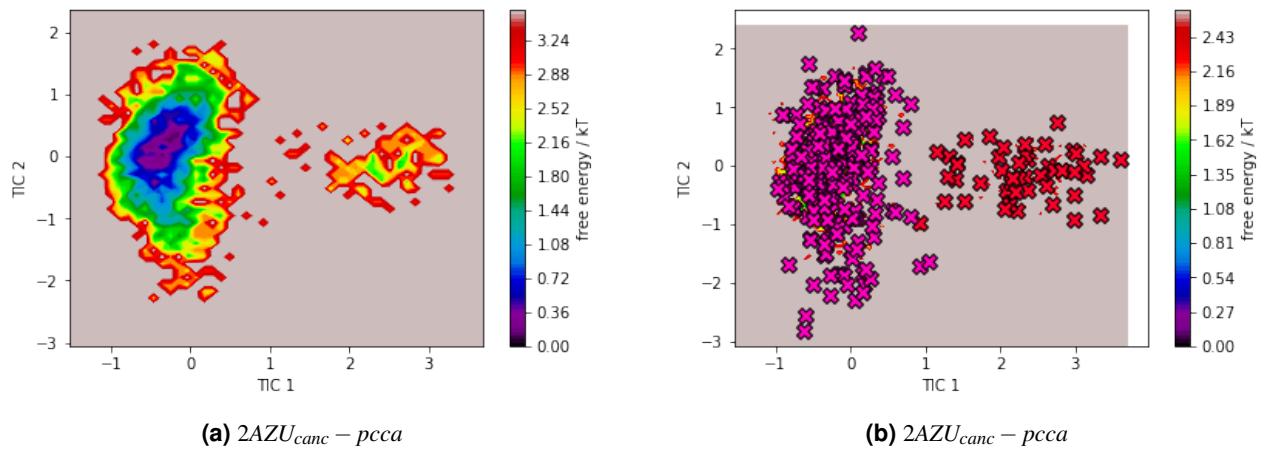


Figure 7. Caption for this figure with two images

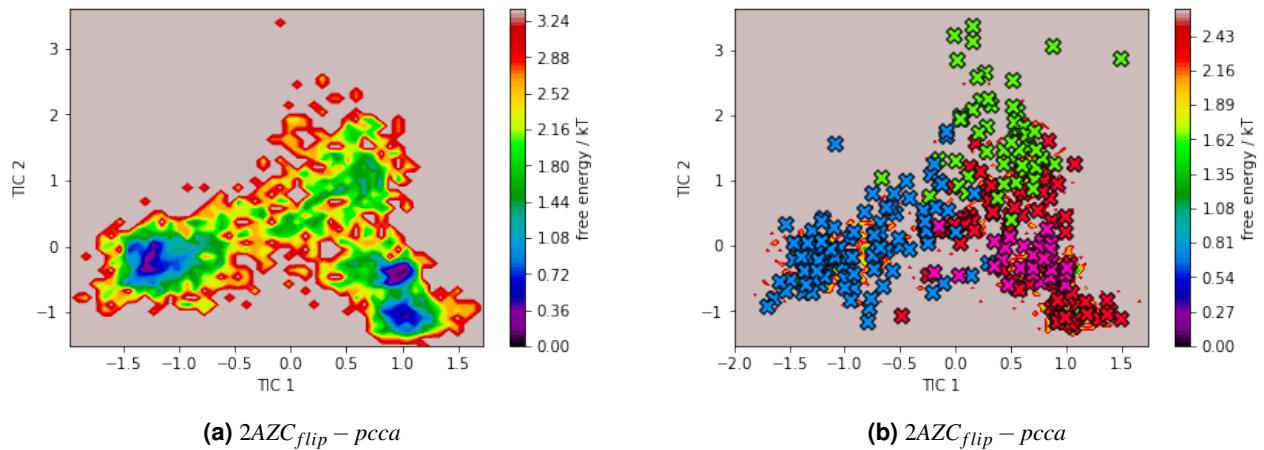


Figure 8. Caption for this figure with two images

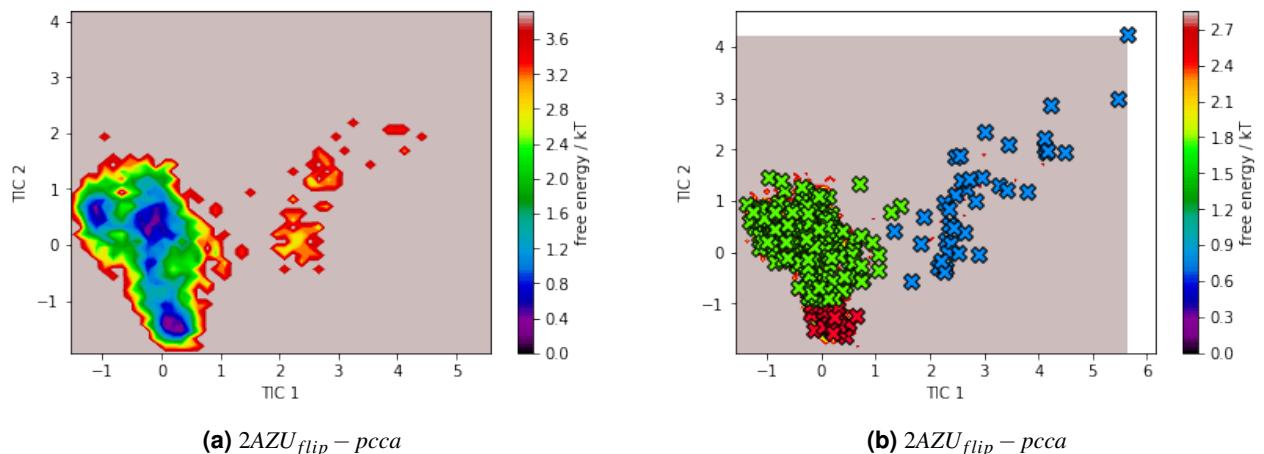


Figure 9. Caption for this figure with two images

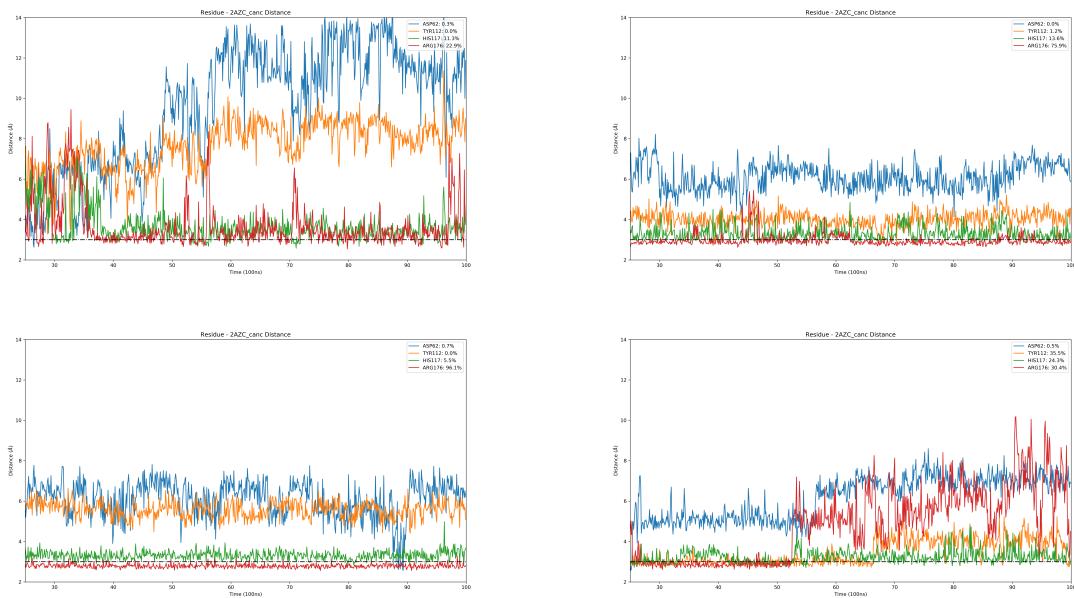


Figure 10. $2AZC_{canc} - distance$

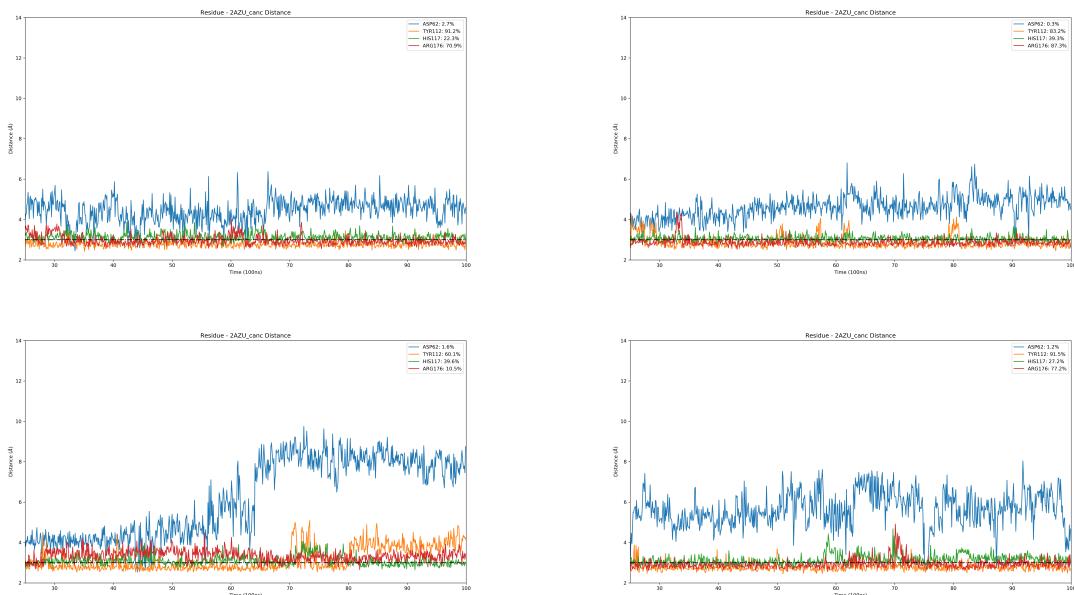


Figure 11. $2AZU_{canc} - distance$

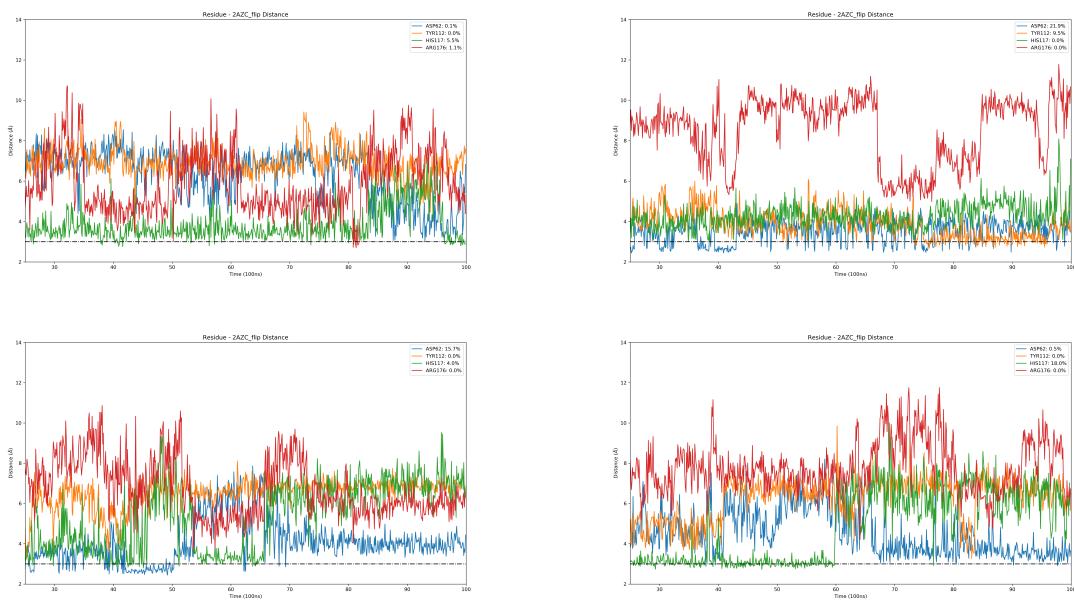


Figure 12. 2AZC_{flip} – distance

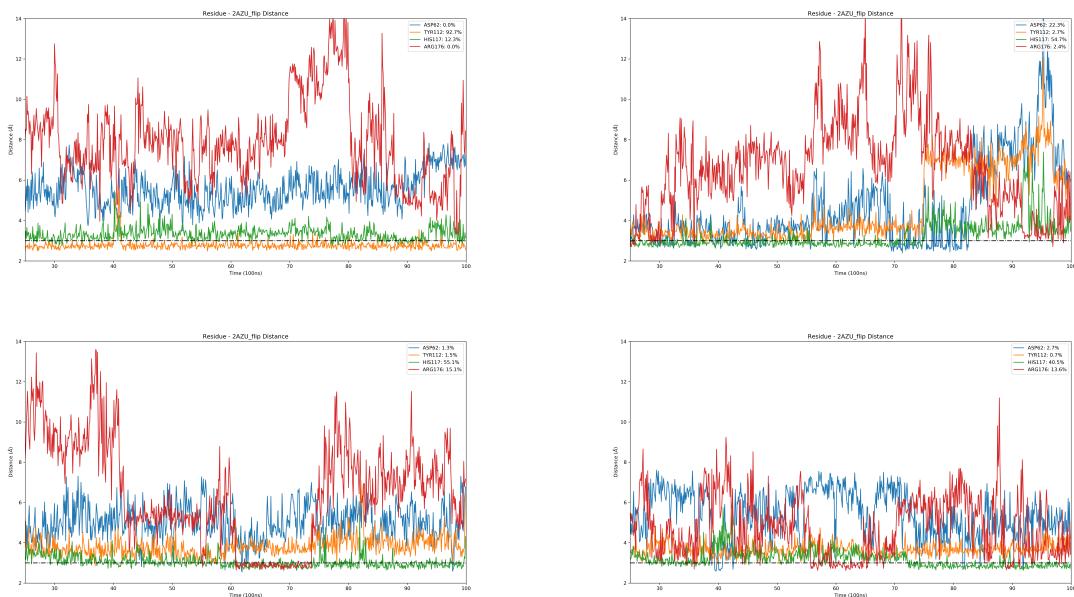


Figure 13. 2AZU_{flip} – distance