Electronic Supplementary Material (ESI) for Soft Matter.

Development of the Computational Antibiotic Screening Platform (CLASP) to Aid in the Discovery of New Antibiotics

Yinghui Dai†, Huilin Ma†, Meishan Wu, Tory Alane Welsch, Soor Rajiv Vora, Dacheng Ren, and Shikha Nangia*

Department of Biomedical and Chemical Engineering, Syracuse University, Syracuse, NY 13244, USA

†contributed equally

*Address for correspondence:
Dr. Shikha Nangia
343 Link Hall
Department of Biomedical and Chemical Engineering
Syracuse University, Syracuse, NY 13244, USA
Phone (315) 443 0571
Email: snangia@syr.edu

Contents

1. Description of the CLASP python scripts

The CLASP program is designed to facilitate high-throughput screening of small molecule solutes that have the potential of being antibiotic drugs. The solute molecules can include compounds available in natural compound libraries and other synthetic compounds that can be transported into the bacterial periplasm through the porin channels. The salient features of CLASP program are:

- **automates** the tedious umbrella sampling simulation set up. The user provides the desired number of umbrella sampling windows (N) and length (L_s) of the translocation coordinate, which roughly corresponds to the length of the porin channel.
- **minimizes** the amount of effort to prepare the simulation set up. Only structure of the solute molecule and the porin need to be provided.
- **streamlines** the input and output workflow. All programs are executed by simple command line. No additional programming is required. CLASP offers user-friendly organization of output as separate labeled folders of the simulation runs will be generated and well labeled. All necessary topologies and parameter files needed for the simulation are added automatically to the simulation folders.
- **accelerates** the simulations of transport mechanisms. CLASP reaches 3 μs/day based on the current resources.
- **analyses** simulation results as CLASP provides a variety of outputs, including the potential of mean force, transport barriers, molecular orientations during translocation, and contact map of molecule-protein residues, etc.

The description of the input and output scripts are as follows:

C_setup.py *generates umbrella sampling configurations.*

The C_setup.py script takes the membrane-embedded, solvated, and equilibrated porin structure as input along with two user defined parameters:

(a) length, L_s, of the translocation coordinate (s), which roughly corresponds to one and a half times of the length of the porin channel; default value = 11 nm
(b) N, the number of umbrella sampling (US) windows, default value is 100. The user can change N to check for convergence.

The C_setup.py script then prepares N independent US windows with spacing of L_s/N nm. In each window, W_i, the script inserts the probe substrate molecule using gmx insert-molecules utility available in GROMACS. The coordinates for the insertion of substrate molecule in each window is computed relative to the center-of-mass (COM) of the porin. Given the hourglass shape of the porin, the COM lies in the middle of the porin channel. The location of the COM is considered $W_{N/2}$, with equal number of windows on either side.

Next, the C_setup.py script generates N separate folders, each with unique location of the substrate molecule defined within W_i, totaling the length L_s along the translocation coordinate. Each folder is set up to execute energy minimization and umbrella sampling simulations independently. The N jobs run concurrently on separate nodes for maximum efficiency and shortest run completion times.

C_PMF.py *generates potential of mean force profile along the transport coordinate.*

The C_PMF.py script combines the N umbrella sampling simulation run-input files (*.tpr* files) into one file (*tpr.dat*) and combines the N umbrella sampling simulation output files (**f.xvg* files) into one file (*fxvg.dat*). The two combined files are used as the two input files with gmx wham, a GROMACS

built-in utility, along with the error analysis (bootstrap) to generate the potential of mean force plot. Bootstrap analysis is the error analysis available from the gmx wham program to estimate statistical uncertainty. In this study, 200 bootstraps were used as recommended from the GROMACS manual to calculate the standard deviations.

C_Dz.py *generates the orientational analysis of the molecule along the transport coordinate.*

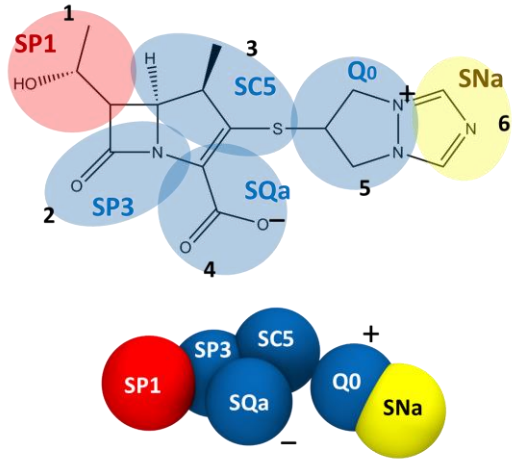
The C_Dz.py script reads the x-, y- and z-axis coordinates of the solute molecule during the transport process using gmx distance, a built-in Gromacs utility with one of its options, -oxyz. Then the script extracts the z-component data and calculates the average and the standard deviation. Finally, the z component is converted to the molecular orientation, which is explained in the main paper.

C_contact.py *generates the contact map of the molecule with the protein residues along the transport coordinate.*

The C_contact.py script calculates the cumulative number of contacts made by the solute molecule with each protein residue throughout the whole trajectory as described in the main paper. Then the script plots the contact map based on the numbers.

In summary, the CLASP is designed to work with GROMACS 5.0 package and later. The CLASP is developed to support outer membrane porins of all bacterial species with porin, provided the porin structure is available. The porins can be embedded in lipopolysaccharide-rich, asymmetric lipid membrane using the *BOB.py* script, which currently has MARTINI coarse-grained parameters of ten Gram-negative bacterial lipids. The CLASP can be easily extended to include new bacterial species once their lipid parameters become available.

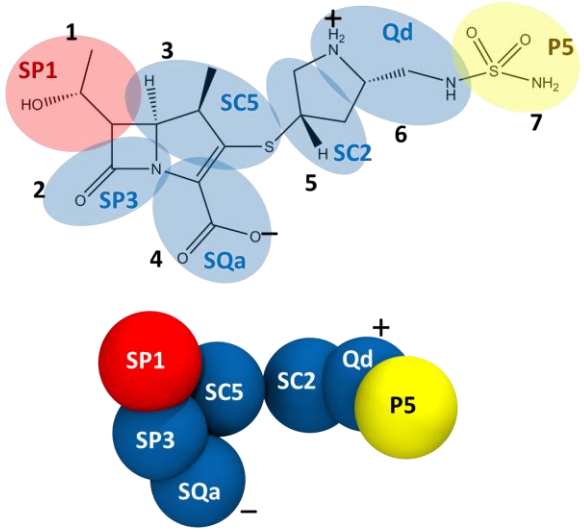
Table S1. CG mapping and parameters for Biapenem



Bond	R(nm)	K_{bond} (kJ mol^{-1} nm^{-2})
1-2	0.297	5000
1-3	0.420	5000
2-3	0.332	5000
2-4	0.295	5000
3-4	0.334	5000
3-5	0.464	5000
5-6	0.252	5000

Angle	θ (deg)	K_{angle} (kJ mol^{-1})
1-2-4	145	50
1-3-4	96	50
1-3-5	148	50
2-3-5	112	50
4-3-5	67	50
3-5-6	152	50

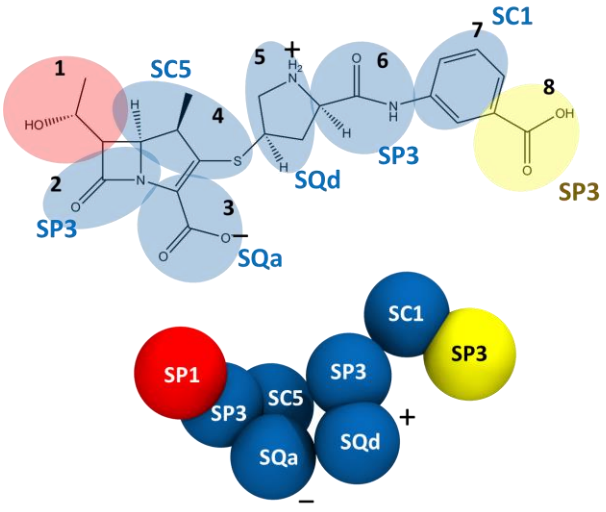
Table S2. CG mapping and parameters for Doripenem



Bond	R(nm)	K_{bond} (kJ mol^{-1} nm^{-2})
1-2	0.292	5000
1-3	0.422	5000
2-3	0.329	5000
2-4	0.300	5000
3-4	0.334	5000
3-5	0.371	5000
5-6	0.247	5000
6-7	0.324	5000

Angle	θ (deg)	K_{angle} (kJ mol^{-1})
1-2-4	146	50
1-3-4	96	50
1-3-5	150	50
2-3-5	129	50
4-3-5	84	50
3-5-6	137	50
5-6-7	134	50

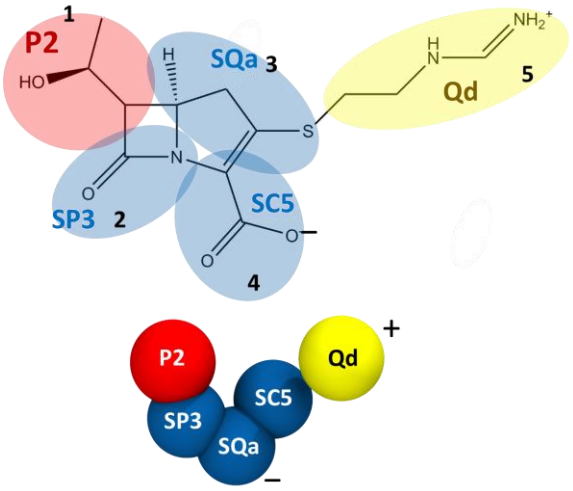
Table S3. CG mapping and parameters for Ertapenem



Bond	R(nm)	K_{bond} (kJ mol^{-1} nm^{-2})
1-2	0.268	5000
1-3	0.566	5000
2-3	0.304	5000
2-4	0.325	5000
3-4	0.339	5000
4-5	0.362	5000
5-6	0.279	5000
6-7	0.385	5000
7-8	0.312	Constrained

Angle	θ (deg)	K_{angle} (kJ mol^{-1})
1-2-3	104	50
1-3-4	56	50
2-4-5	128	50
3-4-5	87	50
4-5-6	119	50
5-6-7	134	50
6-7-8	103	50

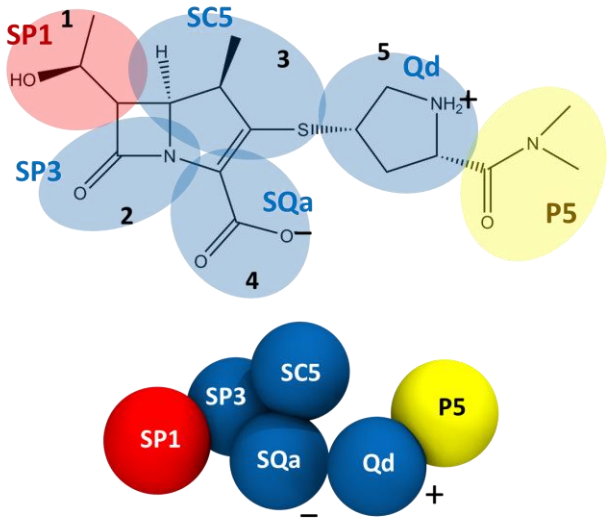
Table S4. CG mapping and parameters for Imipenem



Bond	R(nm)	K_{bond} (kJ mol^{-1} nm^{-2})
1-2	0.293	5000
1-3	0.434	5000
2-3	0.316	5000
2-4	0.301	5000
3-4	0.305	5000
3-5	0.494	5000

Angle	θ (deg)	K_{angle} (kJ mol^{-1})
1-2-4	149	50
1-3-4	99	50
1-3-5	134	50
2-3-5	122	50
4-3-5	85	50

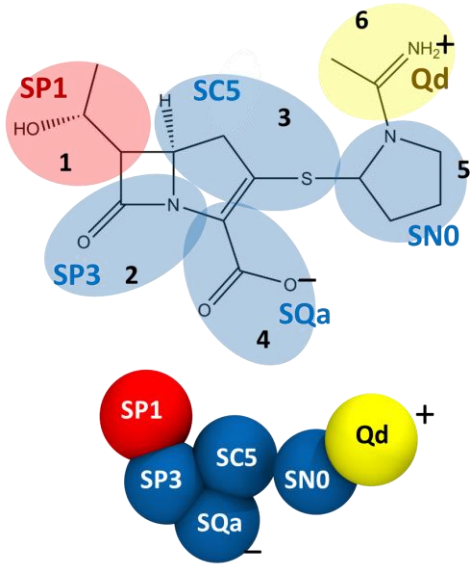
Table S5. CG mapping and parameters for Meropenem



Bond	R(nm)	K_{bond} (kJ mol^{-1} nm^{-2})
1-2	0.297	5000
1-3	0.410	5000
2-3	0.329	5000
2-4	0.303	5000
3-4	0.334	5000
3-5	0.494	5000
5-6	0.343	5000

Angle	θ (deg)	K_{angle} (kJ mol^{-1})
1-2-4	143	50
1-3-4	99	50
2-4-5	114	50
3-4-5	55	50
4-5-6	106	50

Table S6. CG mapping and parameters for Panipenem



Bond	R(nm)	K_{bond} (kJ mol^{-1} nm^{-2})
1-2	0.293	5000
1-3	0.434	5000
2-3	0.316	5000
2-4	0.301	5000
3-4	0.334	5000
3-5	0.403	5000
5-6	0.311	5000

Angle	θ (deg)	K_{angle} (kJ mol^{-1})
1-2-4	146	50
1-3-4	99	50
1-3-5	138	50
2-3-5	143	50
4-3-5	103	50
3-5-6	142	50

Table S7. Composition of the system with embedded OccD3 protein

Antibiotic	outer		inner	W		Ions	
	LPS	DPPE	DPPE	W	WF	Ca^{2+}	Na^{+}
Biapenem	54	6	141	5625	625	58	1
Doripenem	54	6	141	5625	625	58	1
Ertapenem	54	6	141	5625	625	58	1
Imipenem	54	6	141	5625	625	58	1
Meropenem	54	6	141	5625	625	58	1
Panipenem	54	6	141	5625	625	58	1

Table S8. Values of Umbrella sampling windows (*N*) used for the six carbapenem simulations

Antibiotic	*N* value
Biapenem	400
Doripenem	400
Ertapenem	400
Imipenem	400
Meropenem	400
Panipenem	400

An optimal number for N depends on both the size of the drug molecule, and the length of the reaction coordinate. A poor sampling will appear as a gap area in the histogram. When N is inadequate, there will be gaps in the histograms. The user needs to adjust the value of N to achieve better sampling along the entire reaction coordinate. In this study, 100, 200, and 400 for N were tested and the data was converged at 200. In the interest of maximum statistical data, the results for N = 400 are provided.

Table S9. Pairwise sequence alignment of OccD3 and OccD1.

The black dots denote identity. The key pore-lining residues identified in OccD3 are marked in blue outlines.

Figure S1. Histogram of Biapenem with OccD3

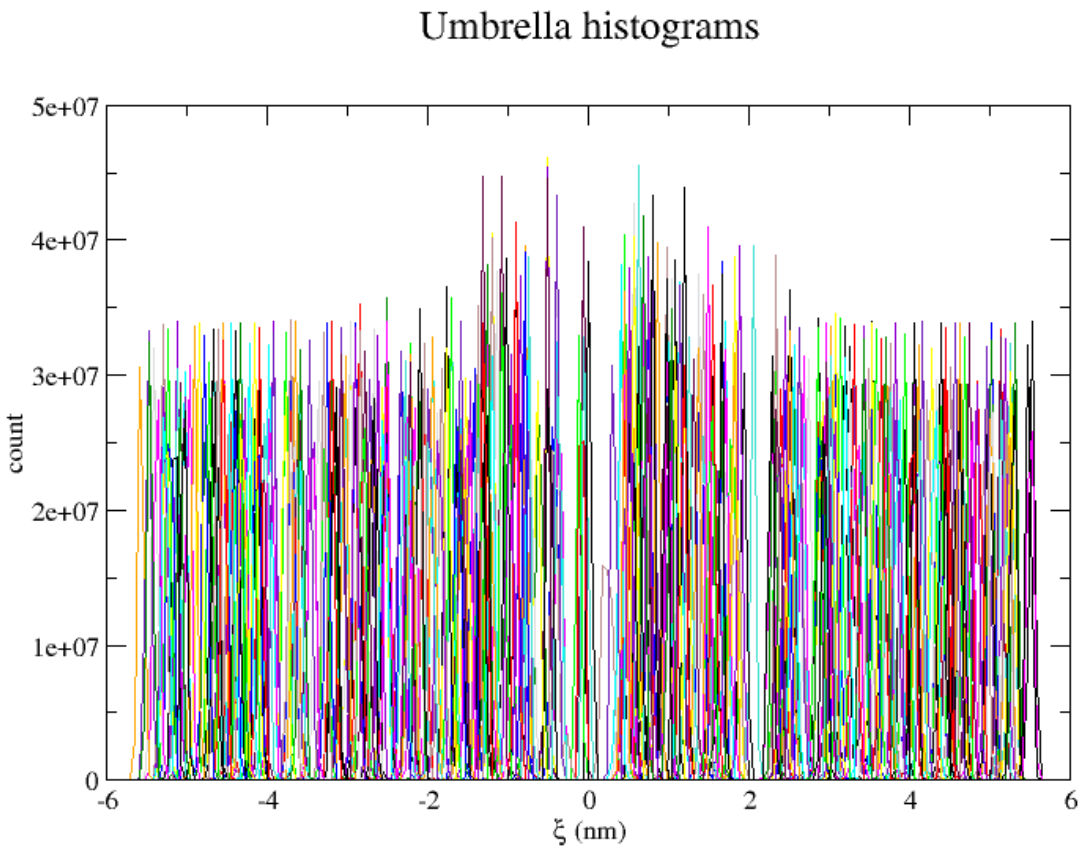


Figure S2. Histogram of Doripenem with OccD3

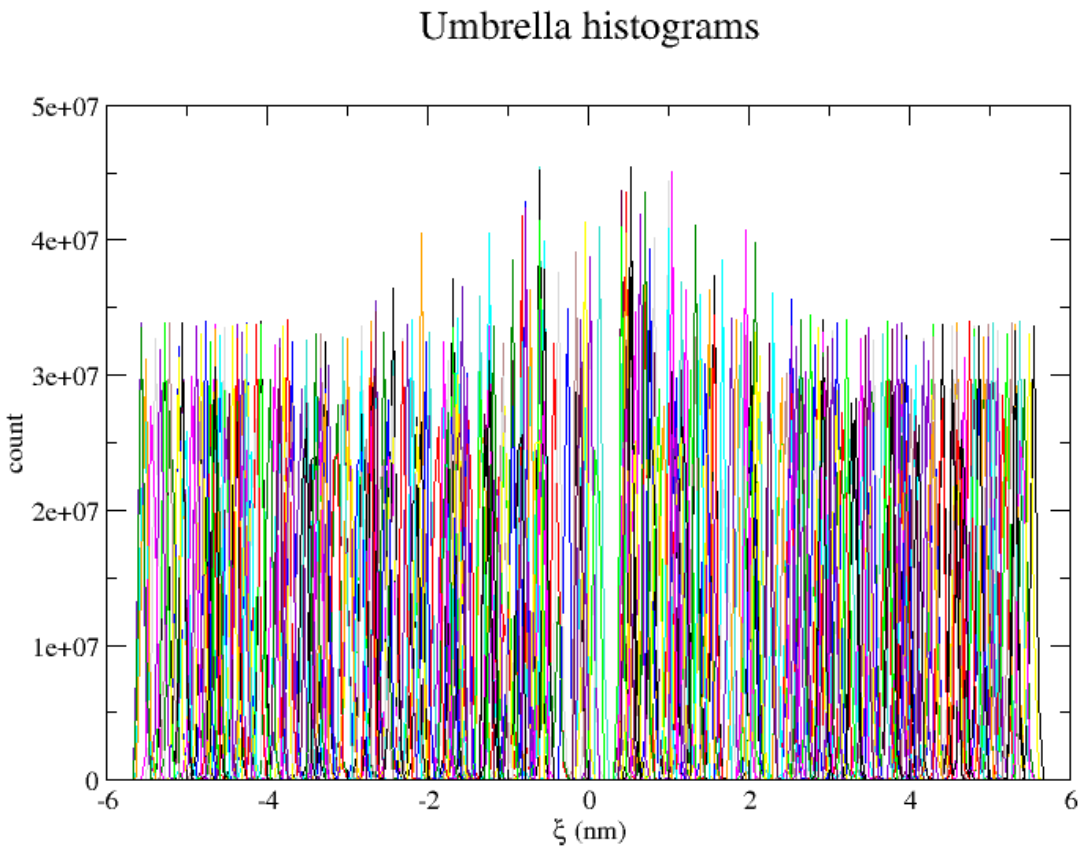


Figure S3. Histogram of Ertapenem with OccD3

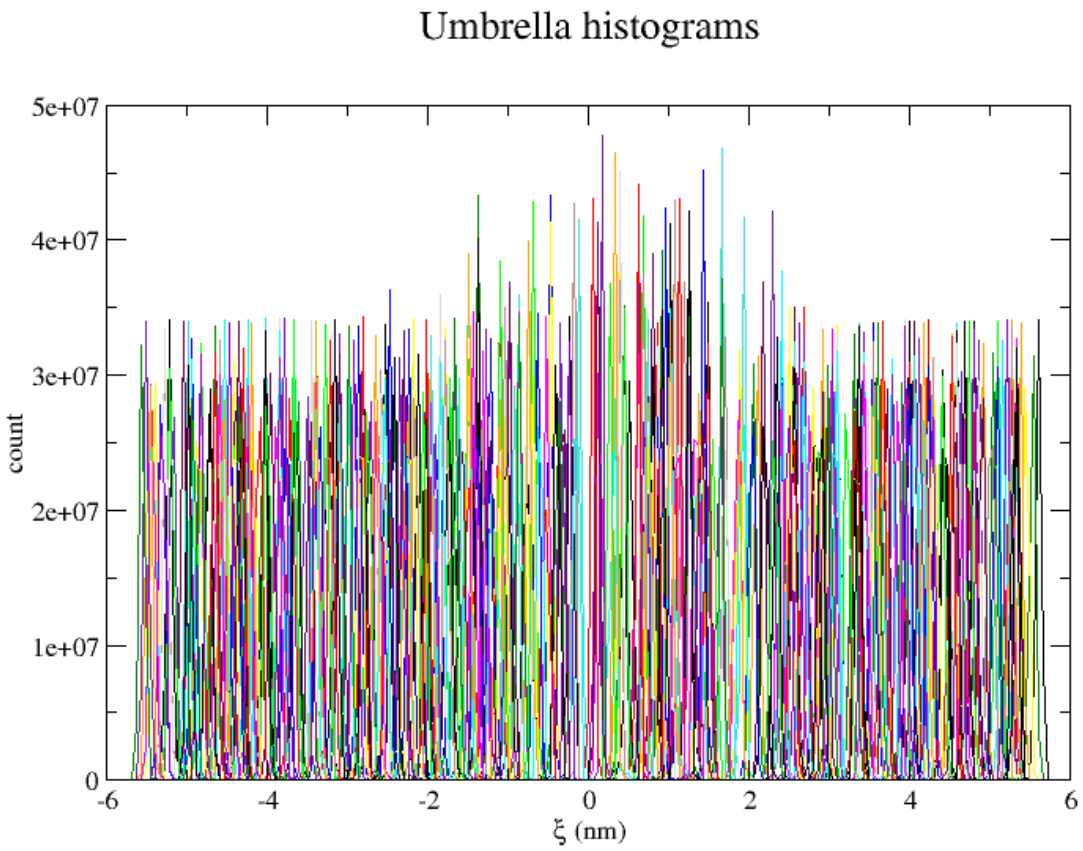


Figure S4. Histogram of Imipenem with OccD3

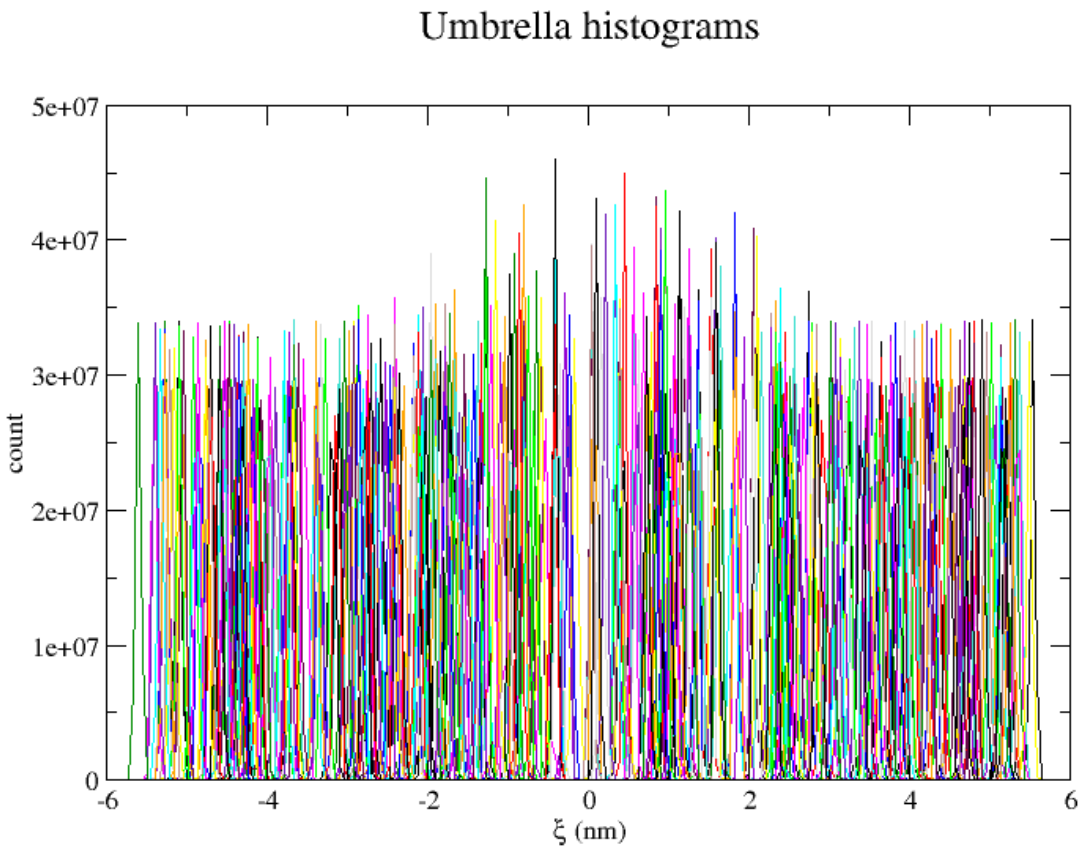


Figure S5. Histogram of Meropenem with OccD3

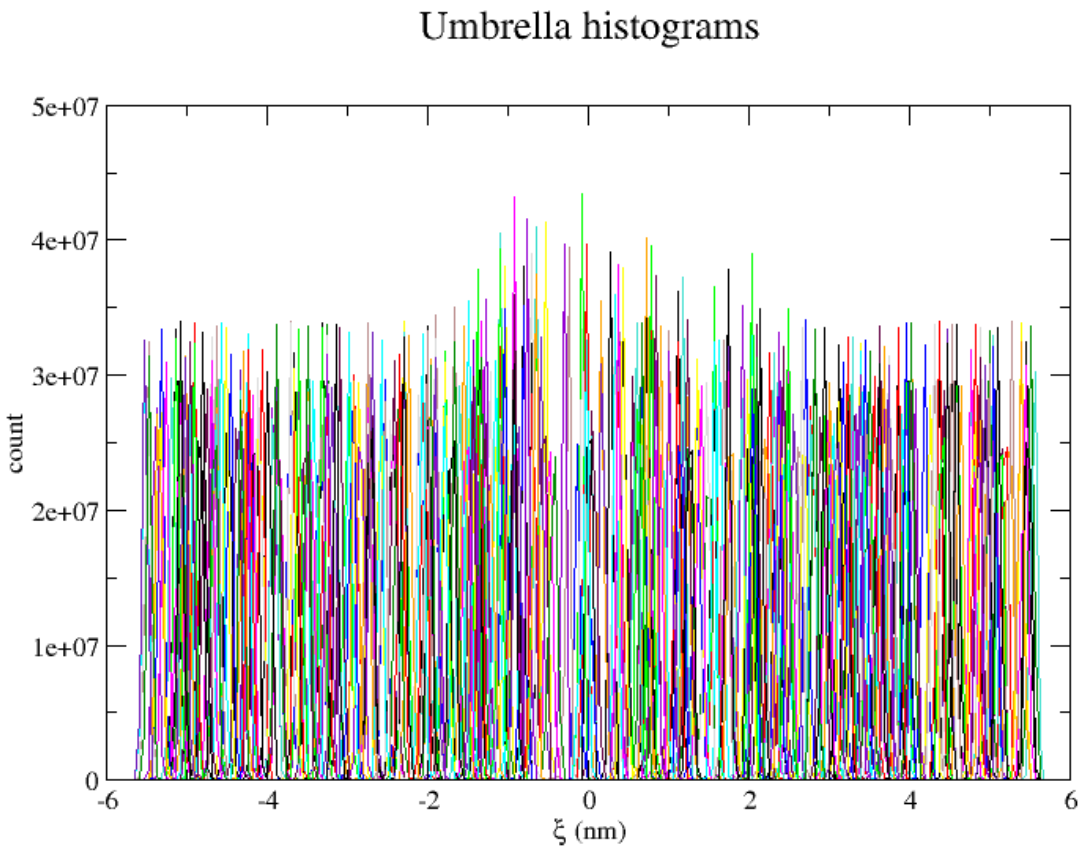


Figure S6. Histogram of Panipenem with OccD3

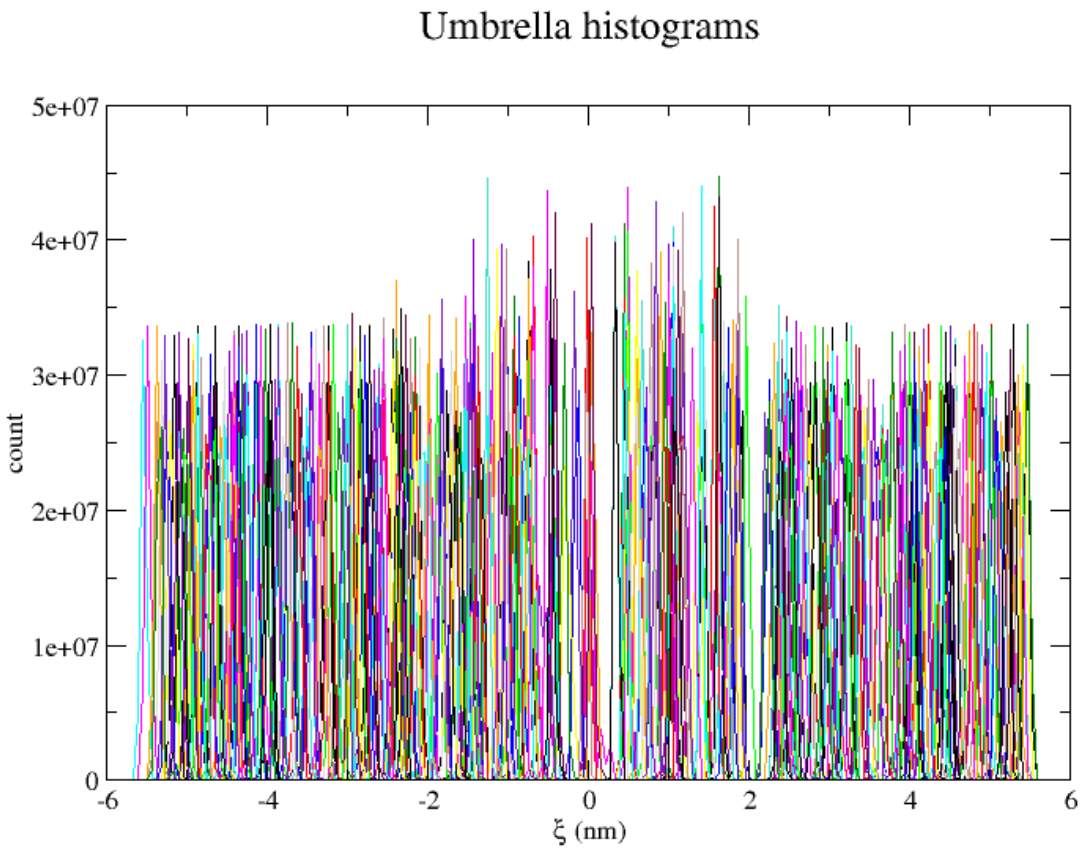


Figure S7. Contact map of Biapenem with (a) OccD3 and (b) OccD3m

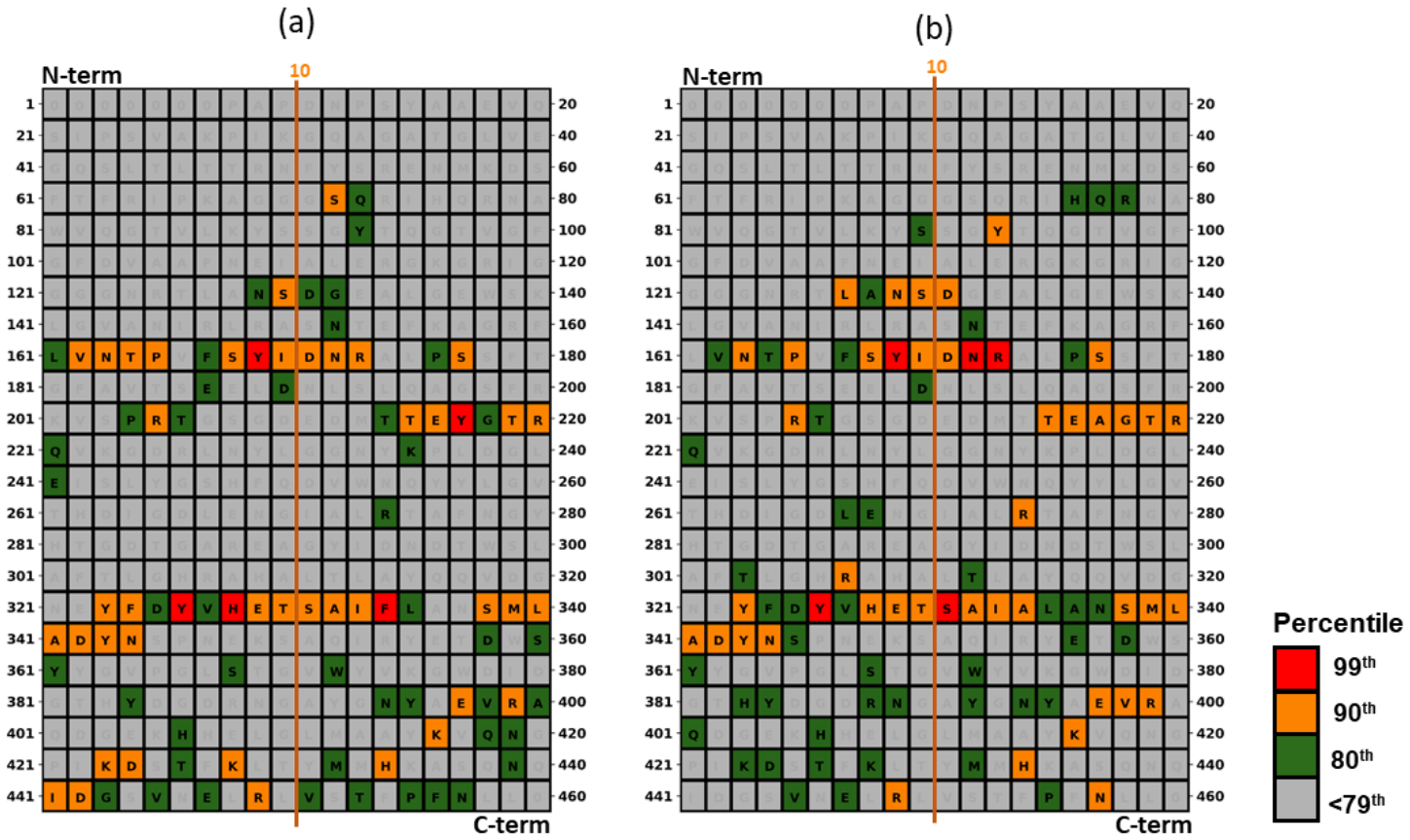


Figure S8. Contact map of Doripenem with (a) OccD3 and (b) OccD3m

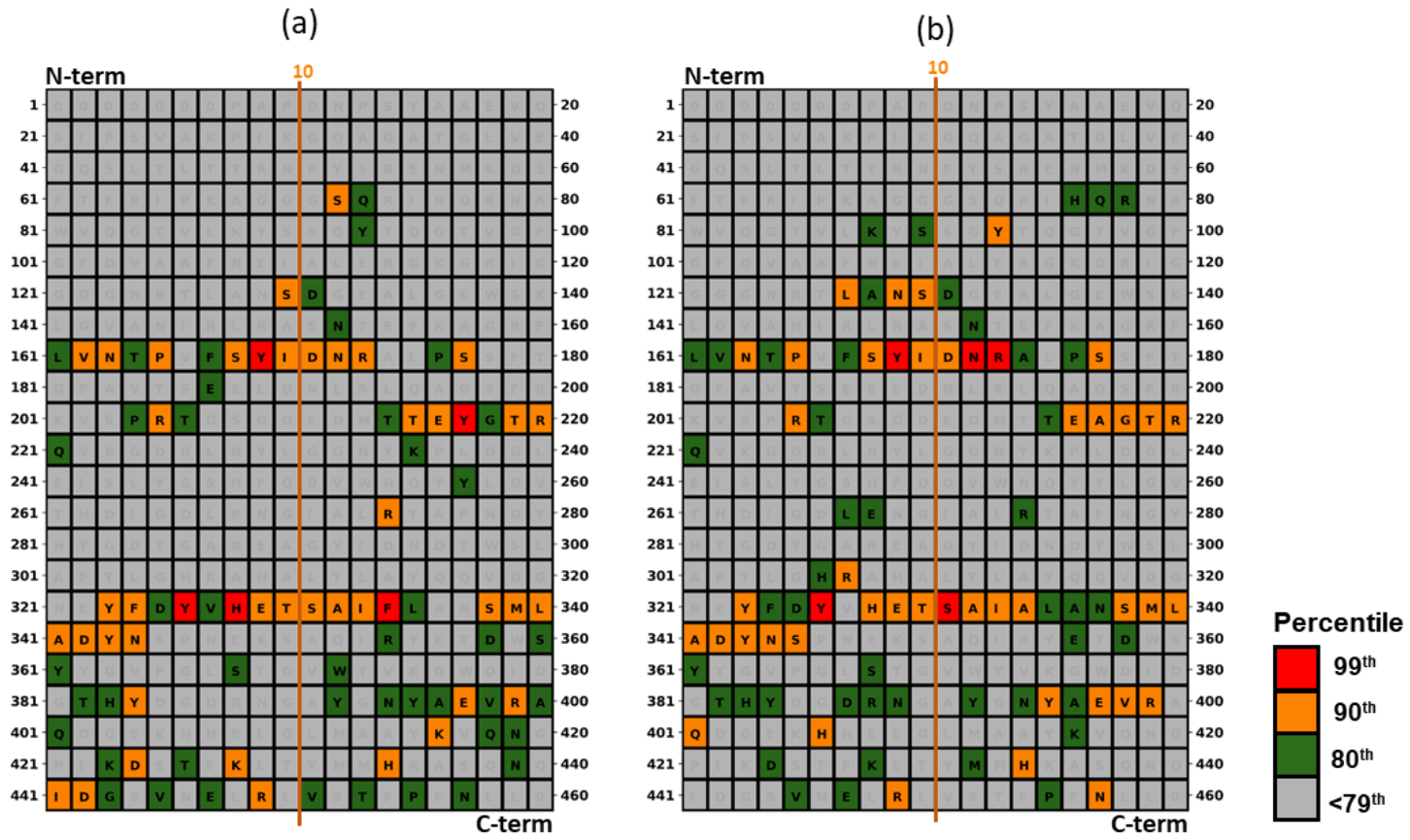


Figure S9. Contact map of Ertapenem with (a) OccD3 and (b) OccD3m

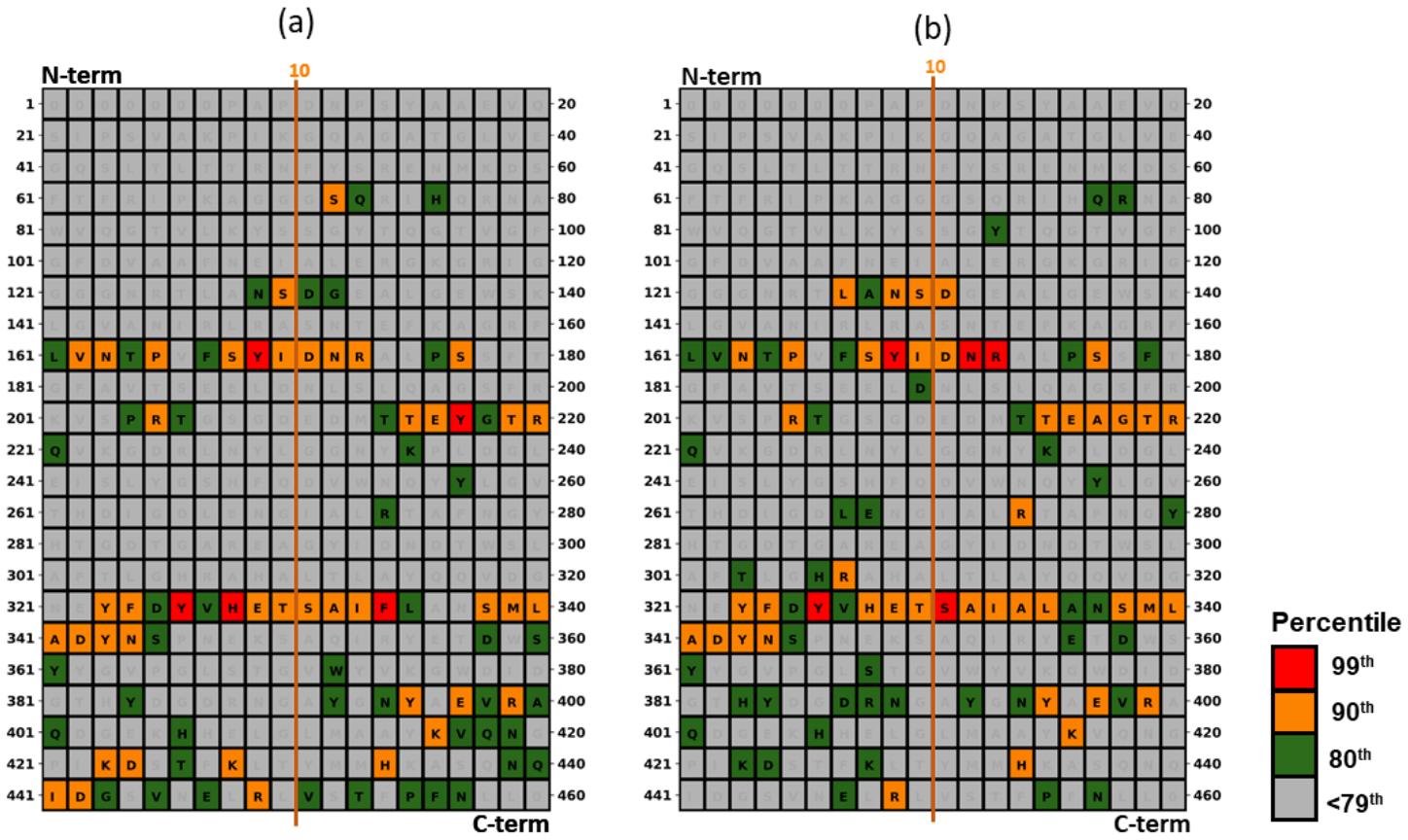


Figure S10. Contact map of Imipenem with (a) OccD3 and (b) OccD3m

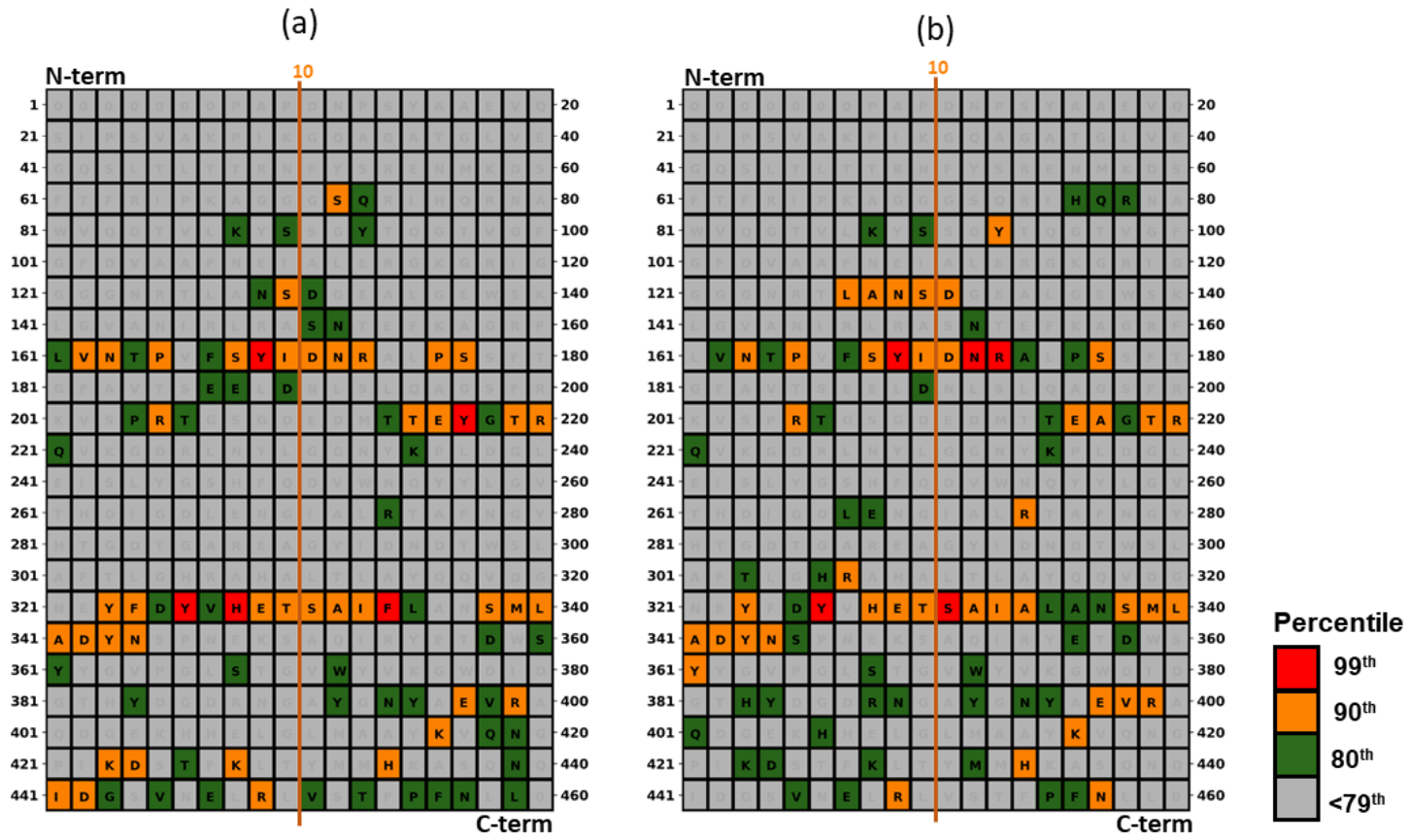


Figure S11. Contact map of Meropenem with (a) OccD3 and (b) OccD3m

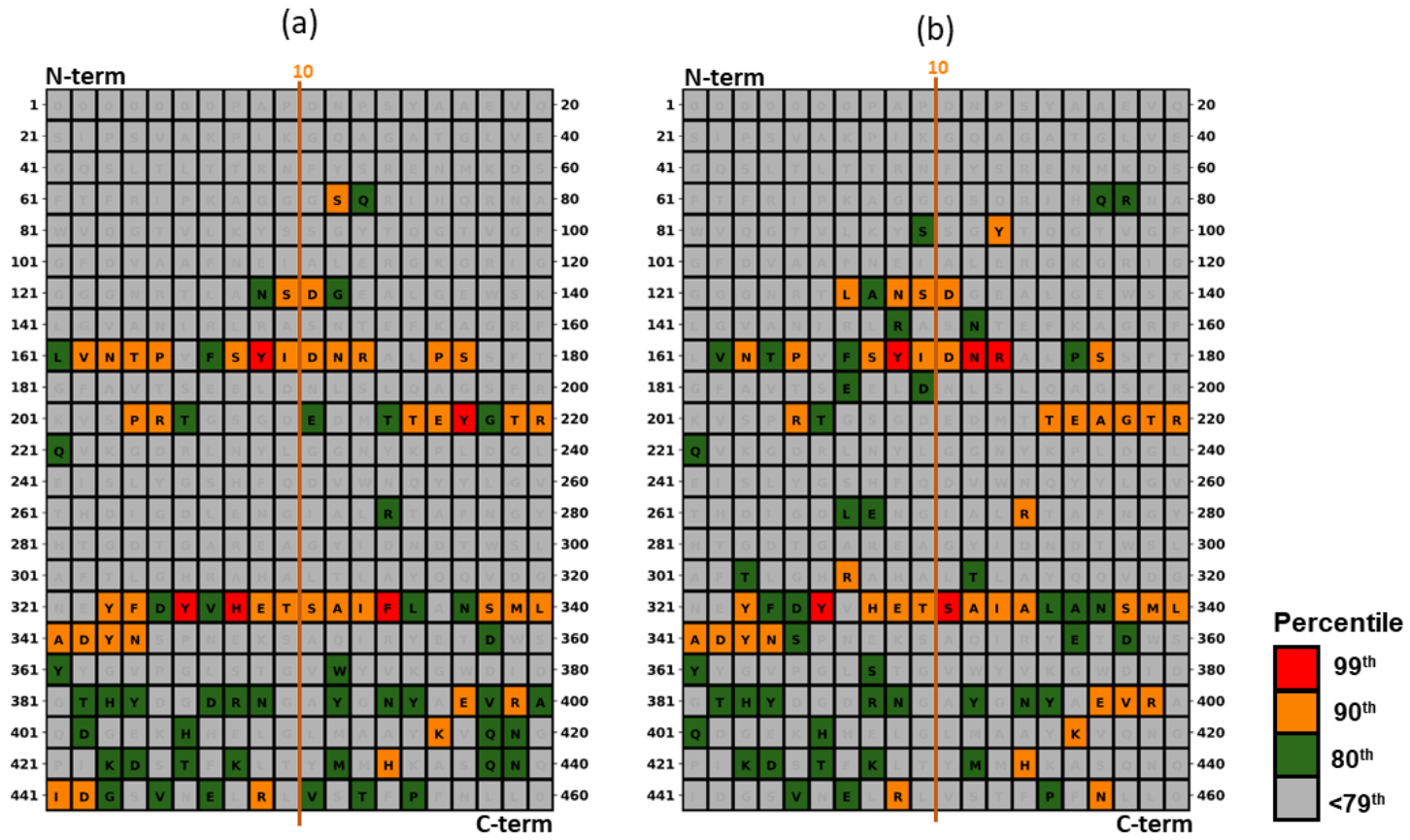


Figure S12. Contact map of Panipenem with (a) OccD3 and (b) OccD3m

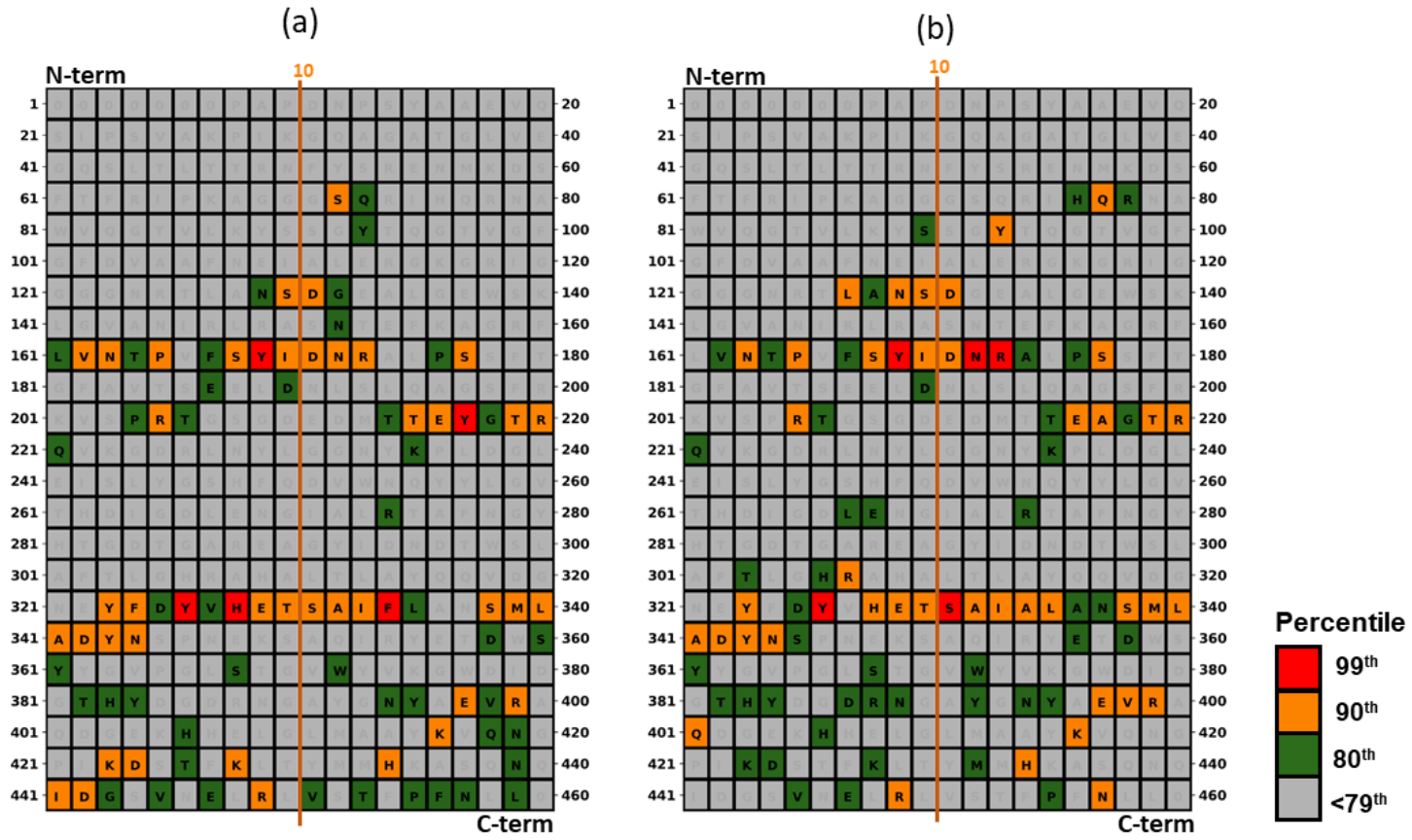


Figure S13. Reserve mapped atomistic structure and orientation of Biapenem in the (a) OccD3 channel and the (b) contacts the molecule makes with the pore-lining residues in the highest energy conformation at $\boldsymbol{s = 5.4}$ nm.

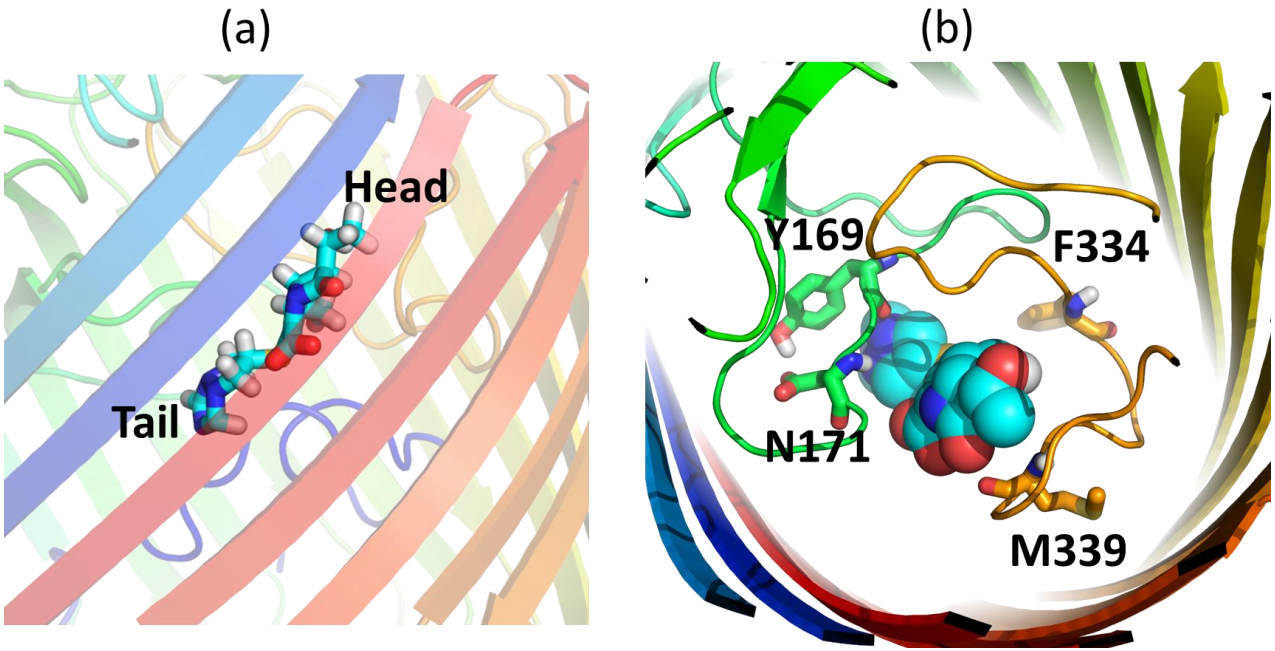


Figure S14. Reserve mapped atomistic structure and orientation of Doripenem in the (a) OccD3 channel and the (b) contacts the molecule makes with the pore-lining residues in the highest energy conformation at $\boldsymbol{s = 6.1}$ nm.

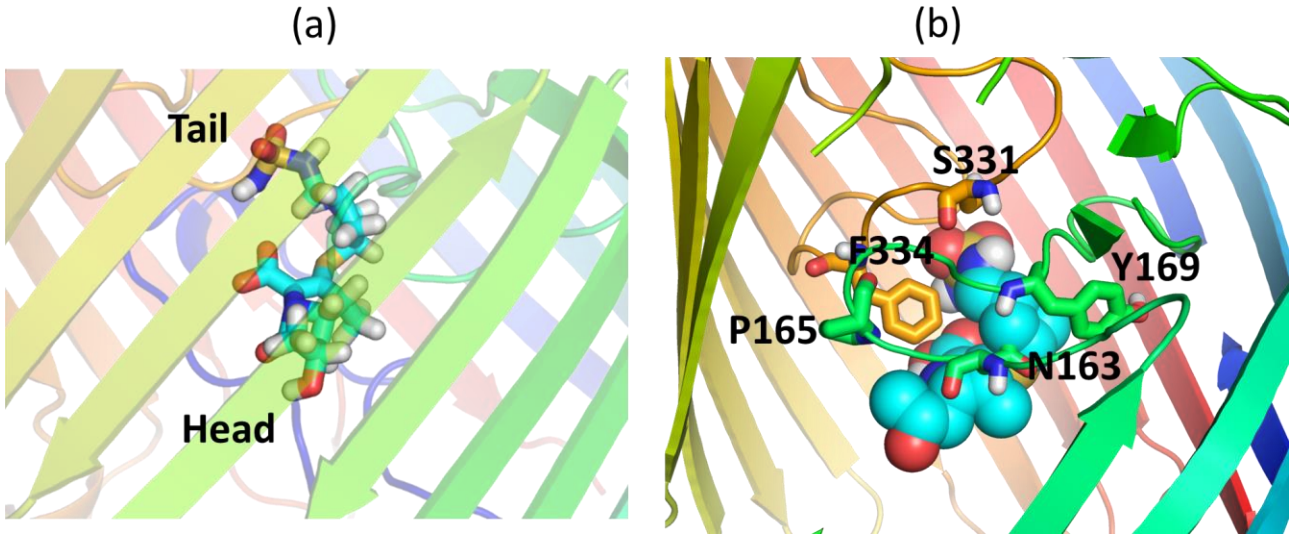


Figure S15. Reserve mapped atomistic structure and orientation of Ertapenem in the (a) OccD3 channel and the (b) contacts the molecule makes with the pore-lining residues in the highest energy conformation at $\boldsymbol{s = 5.7}$ nm.

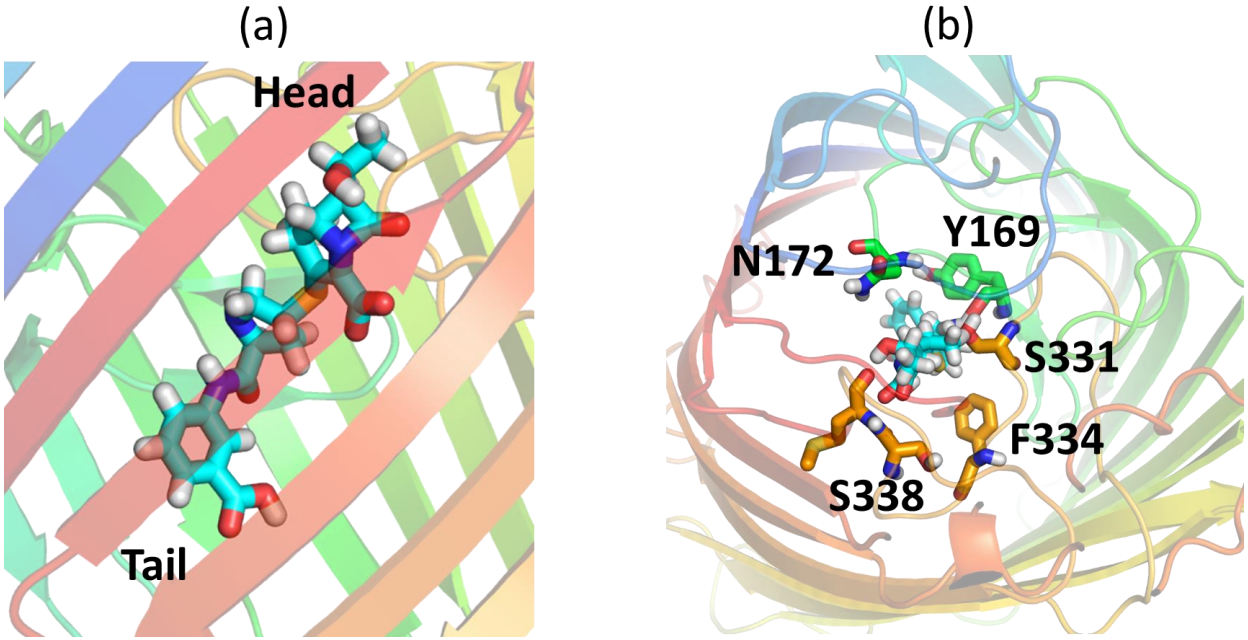


Figure S16. Reserve mapped atomistic structure and orientation of Imipenem in the (a) OccD3 channel and the (b) contacts the molecule makes with the pore-lining residues in the highest energy conformation at $\boldsymbol{s = 5.9}$ nm.

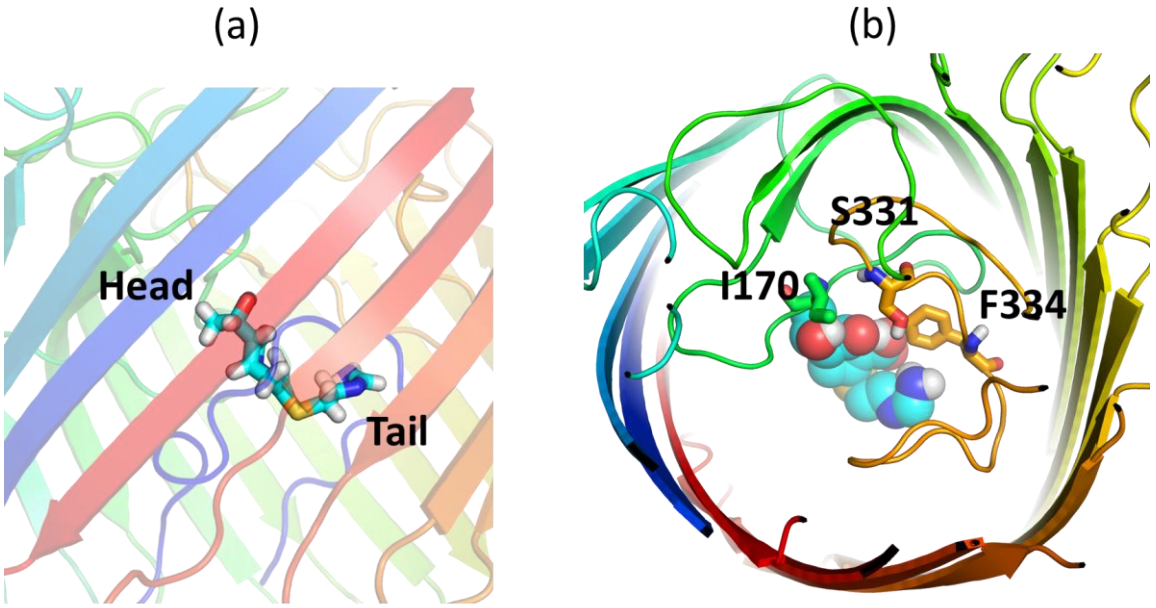


Figure S17. Reserve mapped atomistic structure and orientation of Meropenem in the (a) OccD3 channel and the (b) contacts the molecule makes with the pore-lining residues in the highest energy conformation at $\boldsymbol{s = 5.6}$ nm.

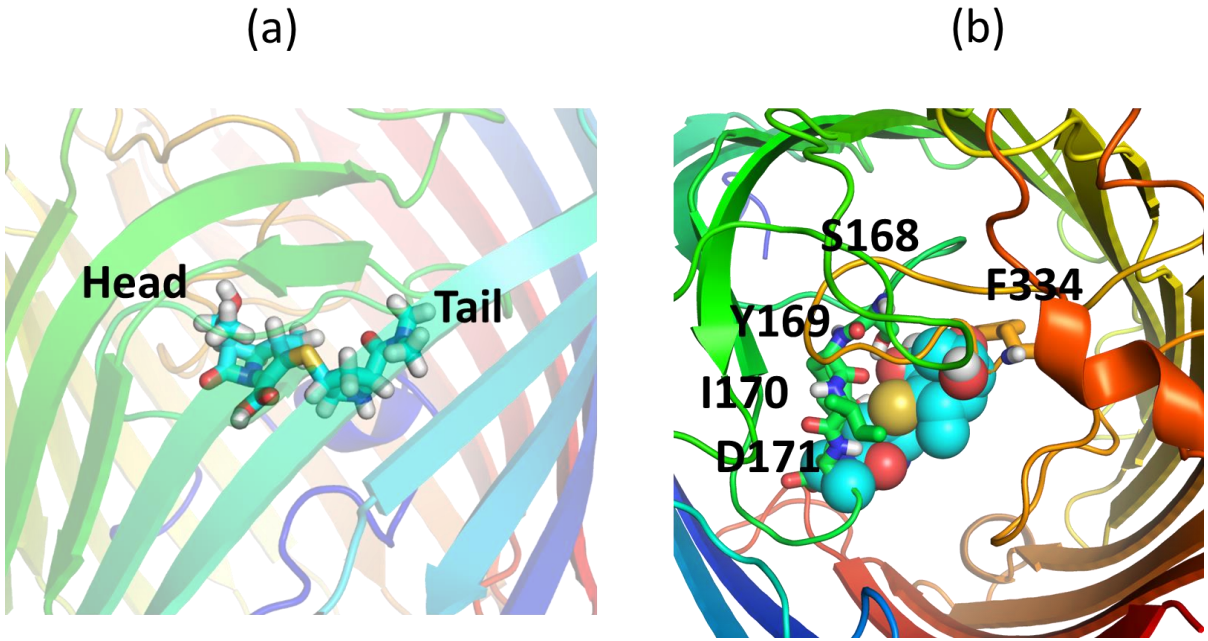


Figure S18. Reserve mapped atomistic structure and orientation of Panipenem in the (a) OccD3 channel and the (b) contacts the molecule makes with the pore-lining residues in the highest energy conformation at $\boldsymbol{s = 6.0}$ nm.

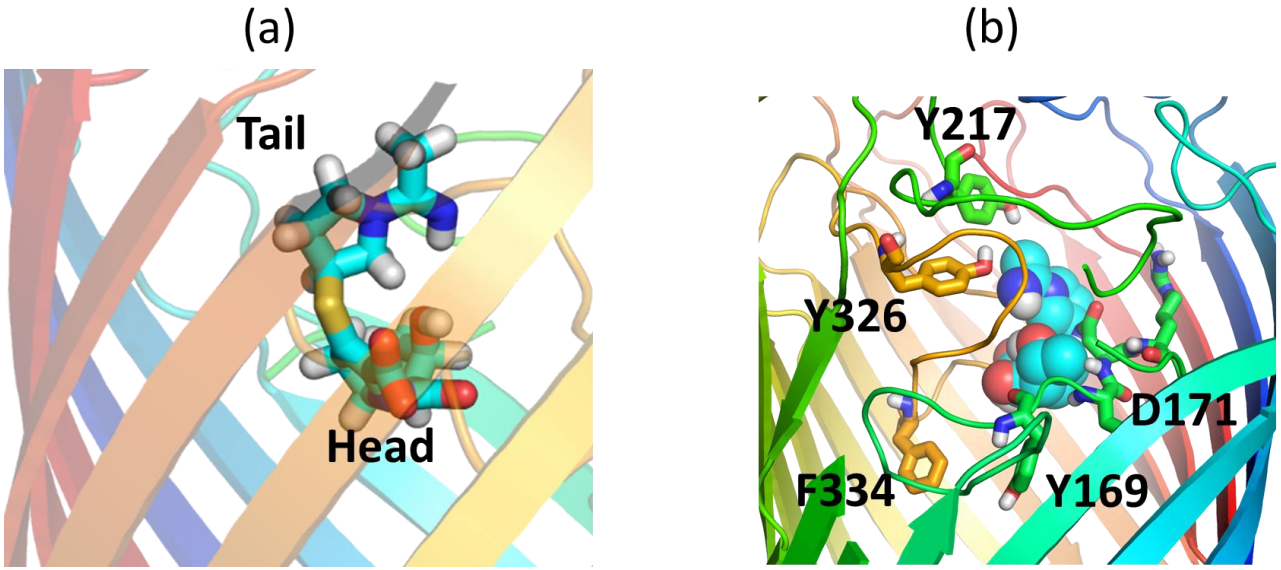


Figure S19. Orientational analysis of carbapenems along the translocation coordinate in OccD3m. The variation in $\mathbf{d}_z$ for (a) biapenem, (b) doripenem, (c) ertapenem, (d) imipenem, (e) meropenem, and (f) panipenem as a function of $\boldsymbol{s}$. The mean $\mathbf{d}_z$ valued are denoted by the yellow dots, and the standard deviation are shown by the black bars. The $\boldsymbol{s}$ coordinate is subdivided into outer membrane region (green), constriction region (white) and periplasmic region (blue).