Supporting Information: Quantifying Selective Metabolite Transport for the Bacterial Microcompartment from Haliangium ochraceum with Molecular Dynamics Simulations

Neetu S. Yadav,[†] Saad Raza,[†] Yali Wang,[‡] Joel F. Landa,[¶] Eric L. Hegg,[§,¶] Robert P. Hausinger,[‡,§] and Josh V. Vermaas[*,†,§]

[†]*MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing MI, 48824, USA*

[‡]*Department of Microbiology, Genetics, and Immunology, Michigan State University, East Lansing, MI, 48824, USA*

[¶]*Cell and Molecular Biology Program and Molecular Plant Sciences Program, Michigan State University, East Lansing, MI, 48824, USA*

[§]*Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, USA*

E-mail: vermaasj@msu.edu

Phone: +1 (517) 884-6937

Simulation Time for each of the molecules

REUS simulations were run for each molecule for 128 replicas.

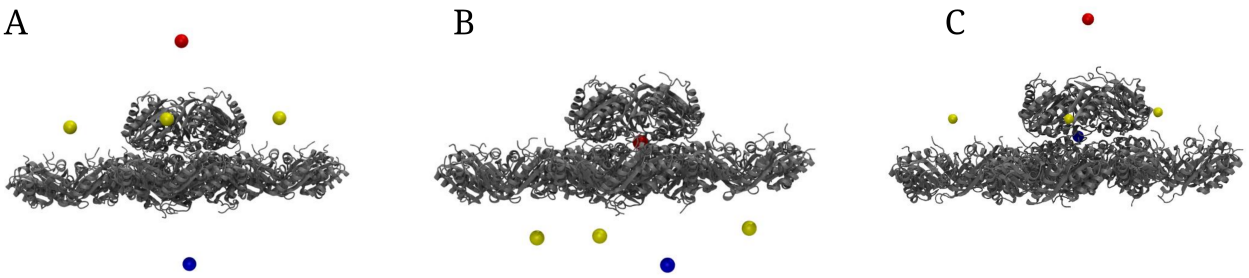

Figure S1: Steered molecular dynamics simulations were performed to pull metabolites using a force constant of 5 kcal/(mol·Å^2) along the Z-axis. Panel A- illustrates the initial configuration, showing the placement of metabolites: two were positioned on each side of the BMC-T unit, and three were placed above the hexameric unit. Panel B- represents the first pulling step. Here, the metabolites positioned above the hexamer were moved from +30 Å to –30 Å along the Z-axis. Simultaneously, the metabolites near the trimer were pulled from +60 Å toward the center of the trimeric pore, while the metabolites located below the structure were kept fixed. Panel C- depicts the second step, where the hexamer-associated metabolites were pulled back from –30 Å to +30 Å, and the trimer-associated metabolites were moved upward along the Z-axis.

Table S1: Simulation time for the selected metabolites. The aggregate time indicates the time for each replica.

Compound	Time (ns)	Aggregate time (μs)
2-aminophenol	80	10.24
2-aminomuconic acid-6-semialdehyde	80	10.24
2-aminomuconate-6-semialdehyde	80	10.24
Picolinic acid	80	10.24
Picolinate	80	10.24
Ammonium	80	10.24
Nitrite	80	10.24
Dithionite	150	19.2

Accessing the PMF Convergence

To obtain a better estimate of the convergence of the PMF, the data were divided into various time chunks. Figures S2 and S3 plot the free energies the molecules, in different time slabs.

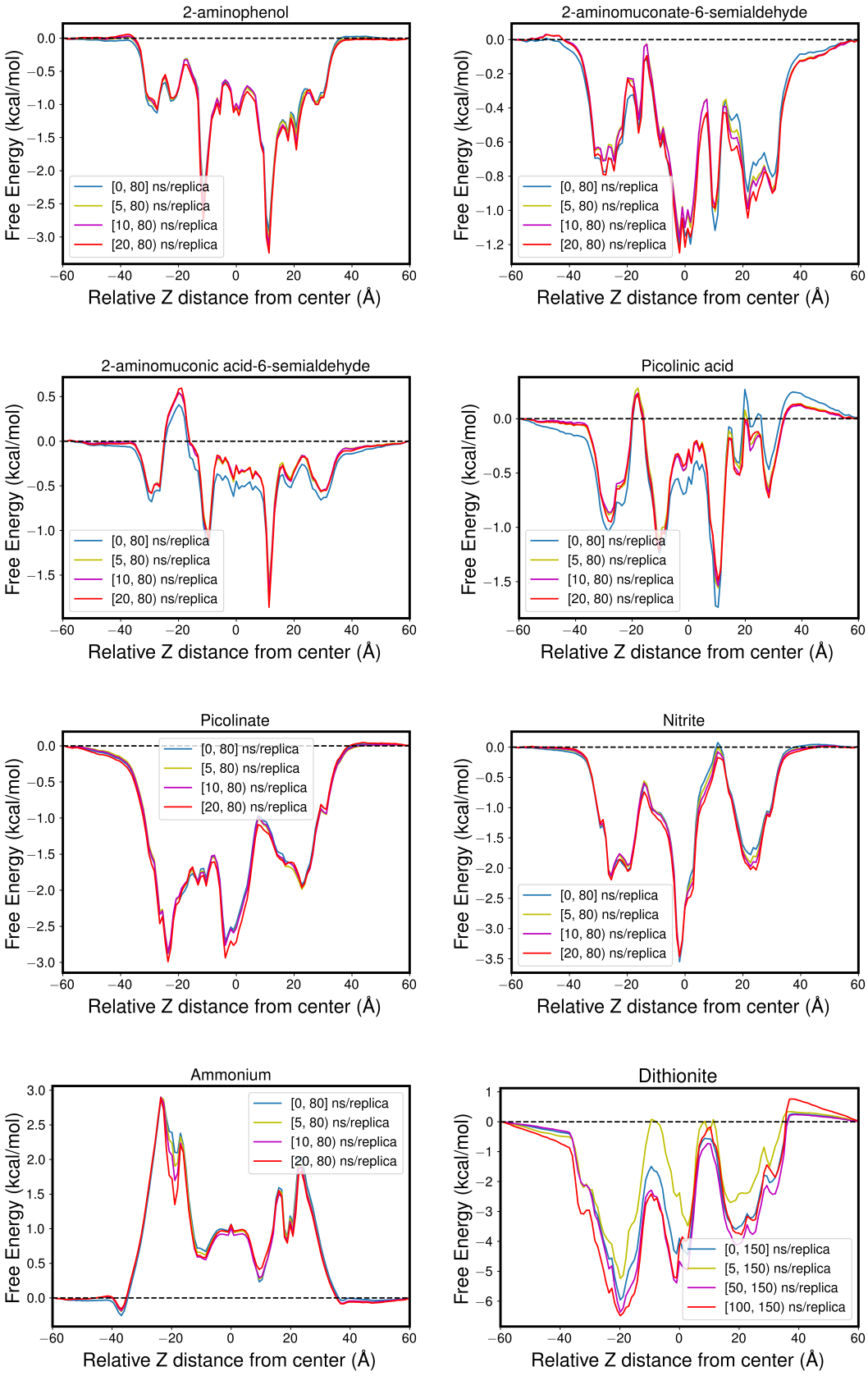

Figure S2: PMF convergence for metabolites across the stacked trimers. The color code for the data is mentioned in the plot. The name of the molecule is mentioned on the top.

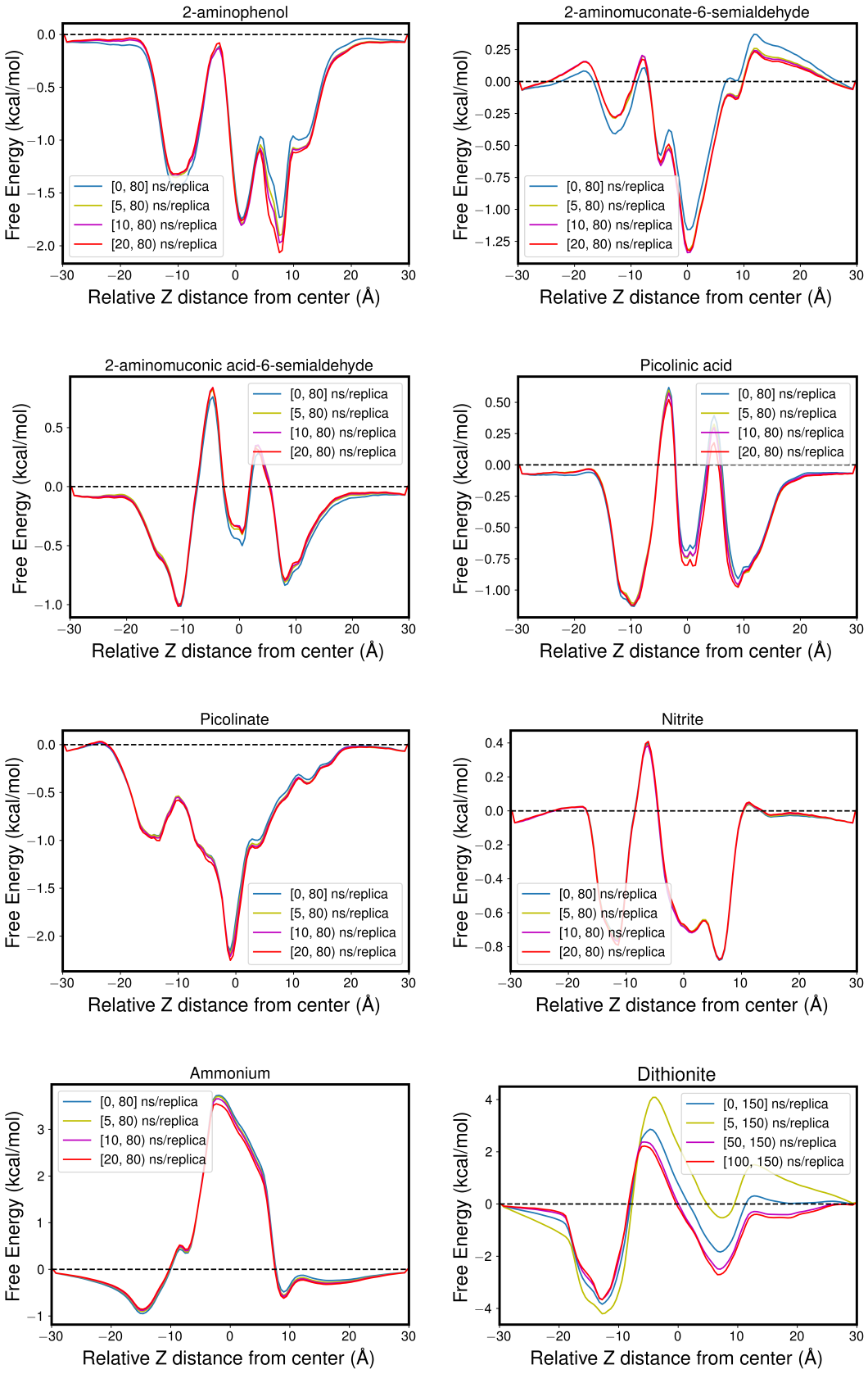

Figure S3: PMF convergence for metabolites across hexamers. The color code for the data is mentioned in the plot. The name of the molecule is mentioned on the top.

Table S2: Water interaction energy for dithionite from QM and MM before and after optimization.

Interactions		Interaction Energy (kcal/mol)		
Ligand Atom	Water	QM	MM Diff CGenFF	MM Diff Optimized
O1	LN1	-21.557	3.692	-0.475
O1	LN2	-21.472	3.473	-0.771
O2	LN1	-21.557	3.692	-0.475
O2	LN2	-21.472	3.474	-0.770
O3	LN1	-21.480	3.481	-0.763
O3	LN2	-21.561	3.702	-0.463
O4	LN1	-21.479	3.481	-0.763
O4	LN2	-21.561	3.702	-0.463
RMSE			3.588	0.635

Table S3: Water interactions energy for nitrite from QM and MM before and after optimization.

Interactions		Interaction Energy (kcal/mol)		
Ligand Atom	Water	QM	MM Diff CGenFF	MM Diff Optimized
O1	LN1	-14.957	-10.991	0.252
O1	LN2	-14.957	-10.991	0.252
O2	LN1	-14.954	7.513	0.249
O2	LN2	-14.954	7.513	0.249
RMSE			9.414	0.250

Extended Method Section

Parameter Optimization

Initial structures of 2-aminomuconic acid-6-semialdehyde, dithionite, nitrite and picolinic acid were downloaded from the PubChem database. All of the compounds were run through CGenFF to obtain initial parameters. Charges and some dihedral parameter penalty scores were substantially higher, and therefore these parameters were refined through the standard

Table S4: Water interactions energy for 2-aminomuconic acid-6-semialdehyde from QM and MM before and after optimization.

		Interaction Energy (kcal/mol)		
Interactions		QM	MM Diff	MM Diff
Ligand Atom	Water		CGenFF	Optimized
O1	LN1	-4.005	-0.309	-0.889
O1	LN2	-4.481	-0.329	-0.112
O1	LP2	-5.249	1.154	0.784
O2	LN1	-3.342	-1.150	-0.763
O2	LN2	-2.999	-1.059	-0.404
O2	LP1	-3.769	-1.374	-0.856
O2	LP2	-3.407	-1.118	-0.298
O3	LN1	-5.010	-1.171	-0.601
O3	LN2	-4.961	-1.197	-0.617
O3	LP1	-6.691	-0.684	0.687
O3	LP2	-6.933	-0.049	0.599
H3	90	-6.123	0.962	0.486
H5	90	-1.547	-0.174	0.505
H6	90	-4.357	1.005	0.500
H11	0	-7.939	1.418	1.132
H11	90	-7.331	1.018	0.758
H12	0	-7.569	0.479	0.336
H12	90	-7.569	0.363	0.209
RMSE			0.938	0.6373

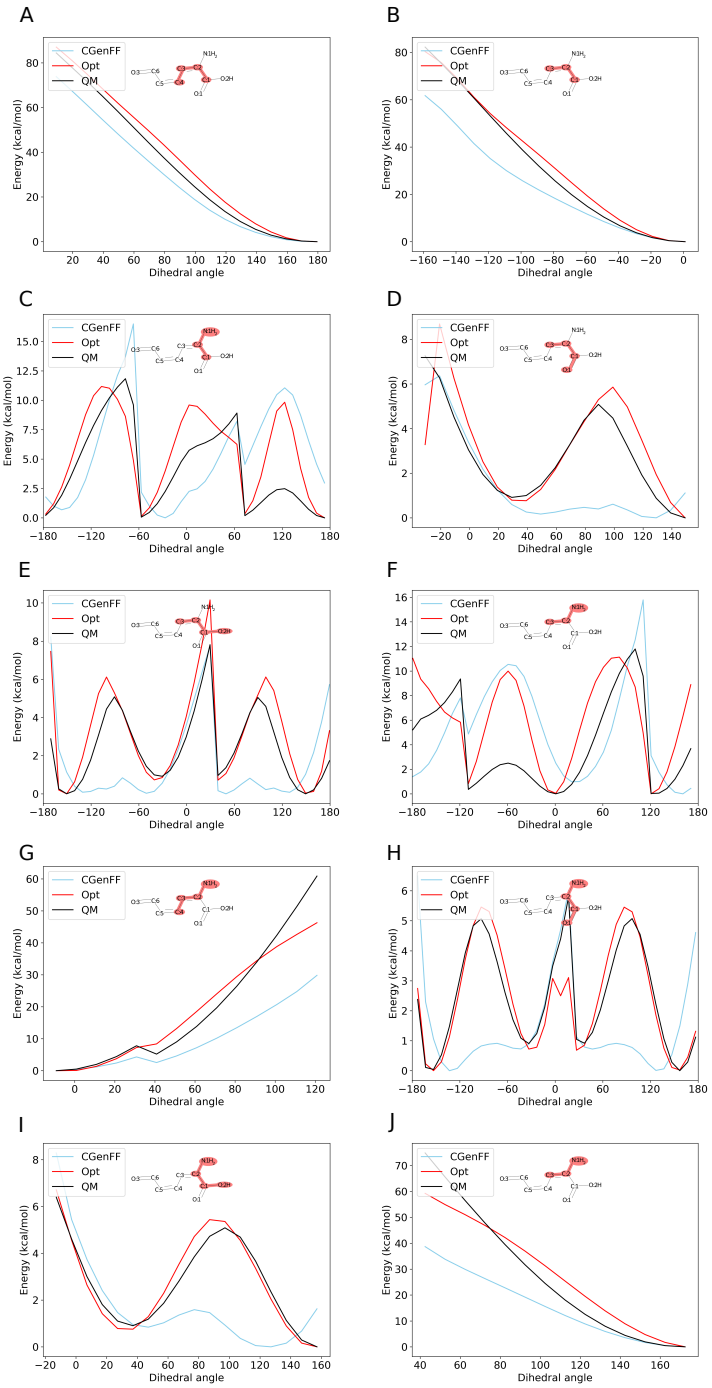

Figure S4: Dihedral free energy scans for 2-aminomuconic acid-6-semialdehyde done by QM calculation (black), through CGenff parameters (light blue) and optimized parameters (red). The exact dihedral being optimized is highlighted in red within the molecular structure shown as a plot inset.

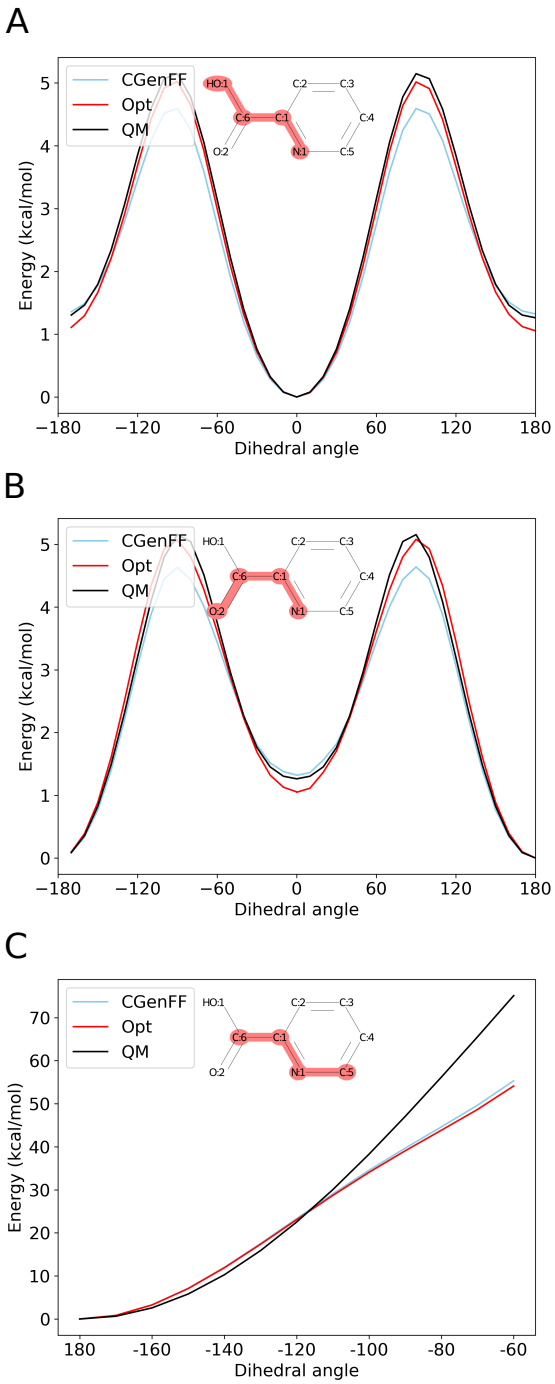

Figure S5: Dihedral free energy scans for picolinic acid done by QM calculation (black), through CGenff parameters (light blue) and optimized parameters (red). The exact dihedral being optimized is highlighted in red within the molecular structure shown as a plot inset.

Table S5: Water interactions energy for picolinic acid from QM and MM before and after optimization.

Interactions		Interaction Energy (kcal/mol)		
Ligand Atom	Water	QM	MM Diff CGenFF	MM Diff Optimized
O1	LP1	-0.956	-0.519	-0.277
O1	LP2	-0.956	-0.519	-0.277
O2	LN1	-4.978	-0.287	-0.311
O2	LN2	-4.978	-0.287	-0.311
O2	LP1	-6.652	-0.199	-0.237
N	LN1	-4.582	1.693	-0.084
N	LN2	-4.582	1.693	-0.084
H2	90	-1.183	0.655	0.354
H3	90	-3.053	1.469	0.472
H4	90	-2.958	0.925	0.276
H5	90	-2.040	-0.277	0.276
H6	90	-7.601	-0.308	-0.154
RMSE			0.916	0.280

optimization protocol for CHARMM. Partial charges optimization protocol was to map the water interaction energy between QM and MM calculation to adjust the partial charges in MM parameters.

Multiple water interaction site geometries were generated by FFparam to cover the water interaction energy. Water molecule is oriented in 4 different orientations around oxygen atom by FFparam, two oriented towards the oxygen lone pair (LP1 and LP2), and two oriented linearly towards the oxygen atom (LN1 and LN2). LP1 and LP2 are oriented with respect to the two different lone pairs on oxygen atom. LN1 and LN2 represent the 180° flip orientation of the second hydrogen atom in the water. For hydrogen atom the water molecule is placed in a 90° orientation. Water conformation with steric clash were avoided for charge optimization. Water interaction energies in QM were calculated at Hartee-Fock level of theory. These are interactions were compared with MM energy value to optimize the partial charges. Standard partial charges of +0.419, +0.21, +0.15, +0.115 and +0.09 were used for hydroxyl, terminal alkene, non terminal alkene, aromatics and alkane hydrogens respectively for these molecules

to maintain compatibility with existing forcefield parameters. The charges on non-hydrogen atoms were redistributed to perform the optimization.

Dihedral potential energy scans are the next step in fitting the forcefield parameters through QM derived potential energy surface (PES). PES is done by scanning the optimized geometry in QM and repeating the scans while changing the dihedral angle. MM energies and QM scan energies overlap is induced by optimizing the dihedral bonded parameters. We modify the force constant, phase and multiplicity for the cosine terms to optimize the dihedral bonded parameters. Initial fitting was done with LSFITPAR with variations of multiplicities to find the least root mean square error (RMSE). Manual tweaking of parameters was done to further minimize the RMSE value (Figure S4, S5)

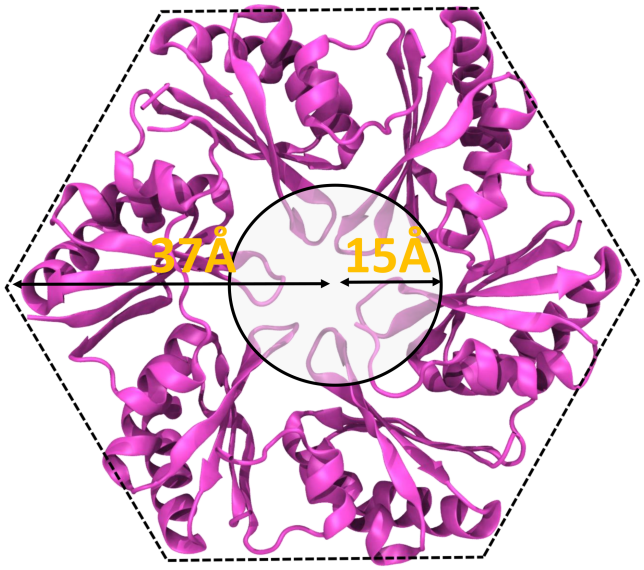

Figure S6: Pictorial representation of the BMC-H shell, highlighting the restrained region with a radius of 15 Å applied to the metabolite.This restraint was designed to ensure accurate computation of metabolite permeability through the pore region.

NrfA Plasmid

For the NrfA studies, NrfA from *Geobacter lovleyi* was inserted via Gibson assembly onto a pBAD202/D-TOPO vector (Invitrogen) with the addition of a SpyC001, and a PelB periplasmic localization signal sequence on the N-terminus. Strep-Tag II was added onto the C terminus for purification. SpyC-NrfA was then induced with 0.02% arabinose and expressed in *Shewanella oneidensis* at 30°C for 16 hours. Periplasmic contents were released via sucrose osmotic shock and purified on a Strep-Tactin XT 4Flow column from IBA Lifesciences. We have added the plasmid map for this construct and amino acid sequence of SPyC001-NrfA with NrfA highlighted in Figure S7.

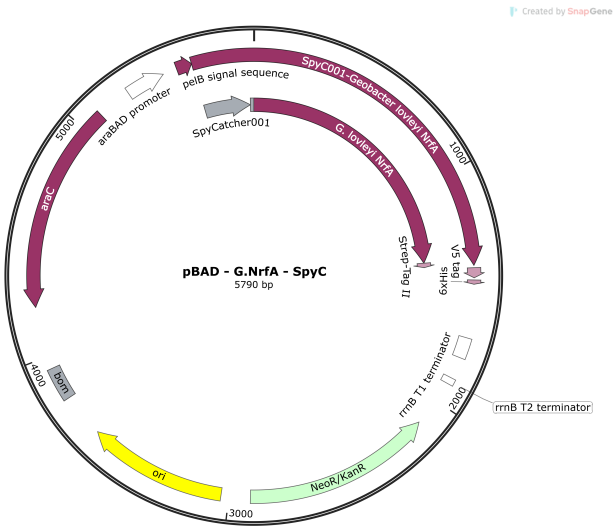

Figure S7: Plasmid construct encoding NrfA gene.

NrfA Amino Acid Sequence:

DSATHIKFSKRDEDGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVE
TAAPDGYEVATAITFTVNEQGQVTVNGGSGGSAPPKAEQAKIAEIPDGTIDPAVWG
KNYPEEYQTWKDTALPTPEGKSKYKKGNDGGKVYDKLSEYPFIALLFNGWGFGIE
YNEPRGHVYMMKDQKEIDPSRLKGGGACLTCKTPYAPQLAQKQGVTYFSQSYAD
AVNQIPKEHQEMGVACIDCHNNKDMGLKISRGFTLVKALDKMGVDQTKLTNQDKR

SLVCAQCHVTYTIPKDANMKSQDVFFPWDESKWGKISIENIIKKMRSDKSYGEWTQ
AVTGFKMAYIRHPEFEMYSNQSVHWMAGVSCADCHMPYTKVGSKKISDHRIMSPL
KNDFKGCKQCHSESSEWLKNQVITIQDRAASQYIRSGYALATVAKLFEMTHKQQAA
GKQIDQKMYDQAKFYYEEGFYRNLFFGAENSIGFHNPTEAMRILGDATMYAGKAD
GLLRQALTKAGVDVPVKIDLELSKYTNNRGAKKLMFKPEQELKDPYGPQKWSHPQ
FEK

T1-AmnA and AmnB Plasmids

For the AmnAB studies, two plasmids were created by Gibson assembly. Plasmid H-T1-AmnA-T2T3P has the gene encoding AmnA inserted into the gene encoding BMC T1 so the fusion protein positions AmnA on a loop of T1 located inside the shell. This plasmid also carries genes encoding BMC components H, T2, T3, and P. Plasmid AmnB contains the gene encoding AmnB.The two plasmids were co-expressed in *E. coli* cells to produce BMC-encapsulated AmnAB.

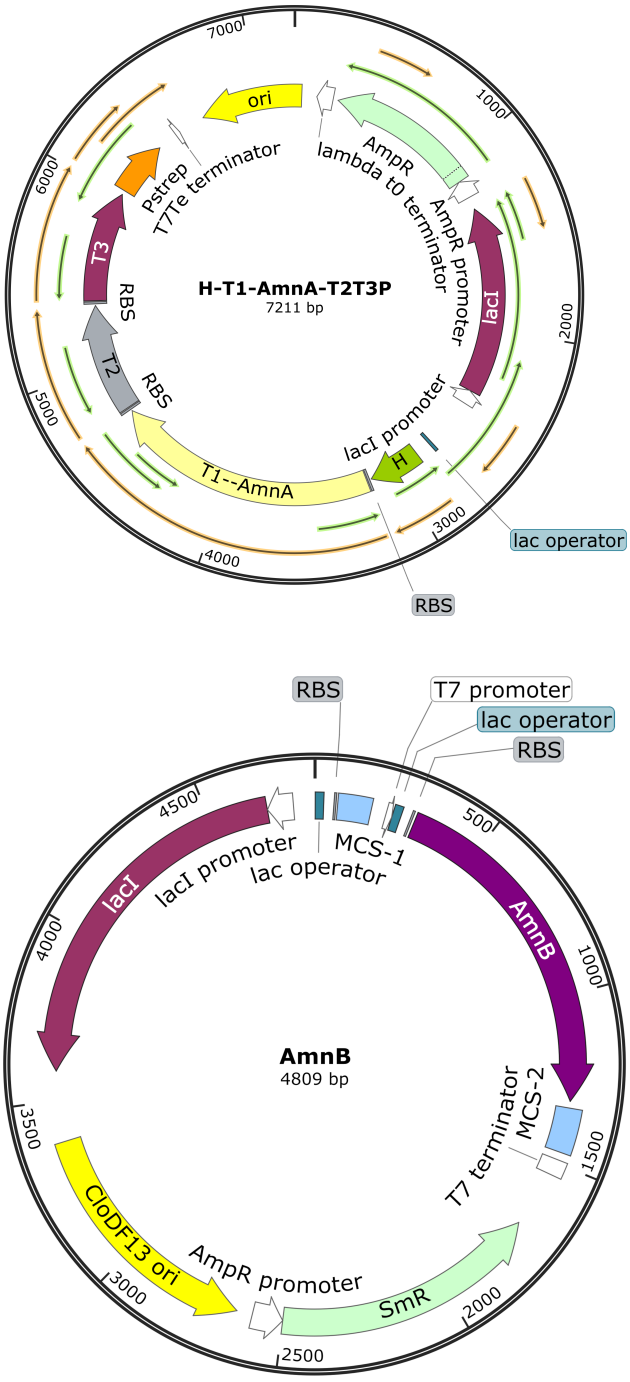

Figure S8: Plasmids encoding T1-AmnA, AmnB and BMC components.

T1-AmnA and AmnB Sequence

The protein sequences of the T1-AmnA fusion protein (with AmnA highlighted in yellow) and AmnB are shown below.

T1-AmnA:

```
MDHAPERFDATPPAGEPDRPALGVLELTSIARGITVADAALKRAPSLLLM    50

SRPVSSGKHLLMMRGQVAEVEESMIAAREIAGAGGGSGGSMADRTGIVAG   100

ALLPGMPHLLAEHPAPSWSALAGAARDVGARLRRLEPDVVLLLSTQWFTV   150

LGHQFQCDPNPRGEHVDENWYAYDYGLLDYDLRFDVDFTERWADRVQAGG   200

MQARRTRYDGFPIDTGTIVTSALLDPDRRLRWAQVSCNLYADADTLADVG   250

RAGAAAARDAGLRAAVVVVTGMSSGLIQQWIEPGQDRIGEPGHDQWNTRV   300

LDLLTAGKVDEVLAVREDFARQAQADSQFRALAFAAGAEATTGPAHLHAY   350

GPIWGTGAAVLSWNLPDHPGGSGGGALLDELELPYAHEQLWRFLDAPVVA   400

DAWEEDTESVIIVETATVCAAIDSADAALKTAPVVLRDMRLAIGIAGKAF   450

FTLTGELADVEAAAEVVRERCGARLLELACIARPVDELRGRLFFGGSGGH   500

HHHHH   505
```

AmnB:

```
MSRVFRVPLAPTRGGATTAPRTPAPAPAPTRPGIVAGCLSPHPPHLIYGE   50

NPPQNEPRSTGGWETLRWAYERLRARIRDVHKPDVLIVHAPHWITMVGHH  100

VNCVPNPRGLSVEPIFPHLFRYRYDFRTDVELGEAIAEEASGLGLVTRTL  150

RDPRVRVDYATIGALHLANPAWDIPVVSLSANNNPYFYSDASLTEMEVLG  200

EATRLAVEATGRRAVLLASNSLSHLHWHEEPELPEDMEREHPYNNHQYRW  250

DMKLLEAIRRGPTAPLRDLIPEHIEATASETKAGSLTWMLAAMGWPKVAG  300

DVLGYGTIIGTGNAIVEWLPEGSHHGGSGGSHHHHHH   337
```

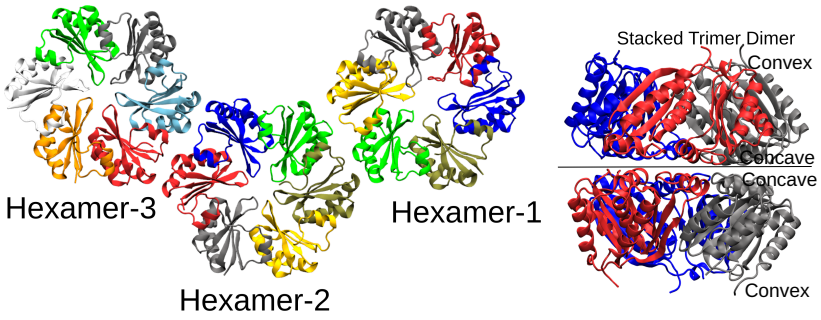

Figure S9: Illustrative representation of the hexamer and trimers. All the chains of hexamers and stacked trimers are colored distinctly for clarity. Both assemblies exhibit concave and convex surfaces, as indicated in the figure for trimer.

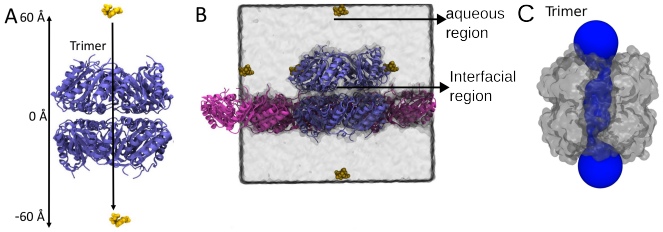

Figure S10: A) Pictorial representation of the BMC-T shell highlighting the stacked trimer dimers. (B) Visualization of the BMC hexamer and the stacked BMC trimer together with the metabolites in the solvated simulation box, providing a sense of scale and illustrating the top, bottom, and interfacial aqueous regions. (C) Illustration of the third aqueous region running through the protein core.

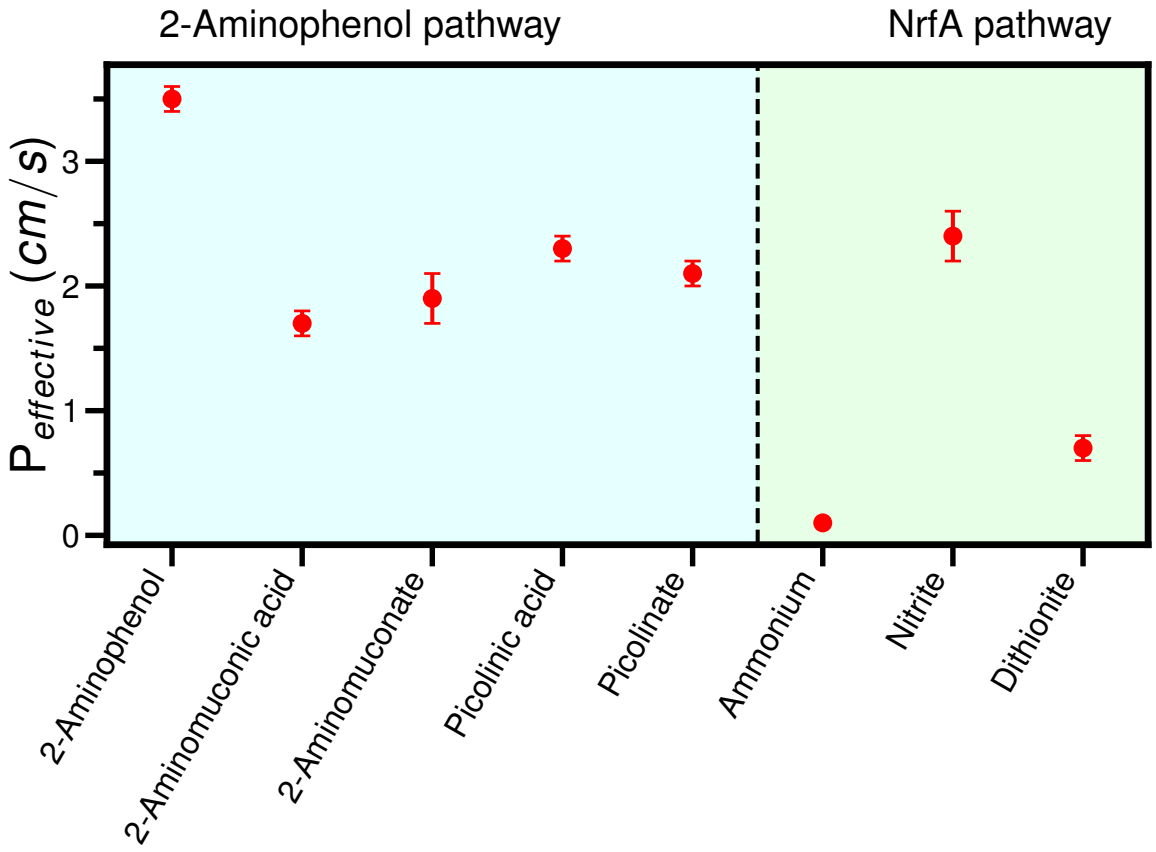

Figure S11: **Effective permeability of metabolites across the hexamer and trimeric pore.**

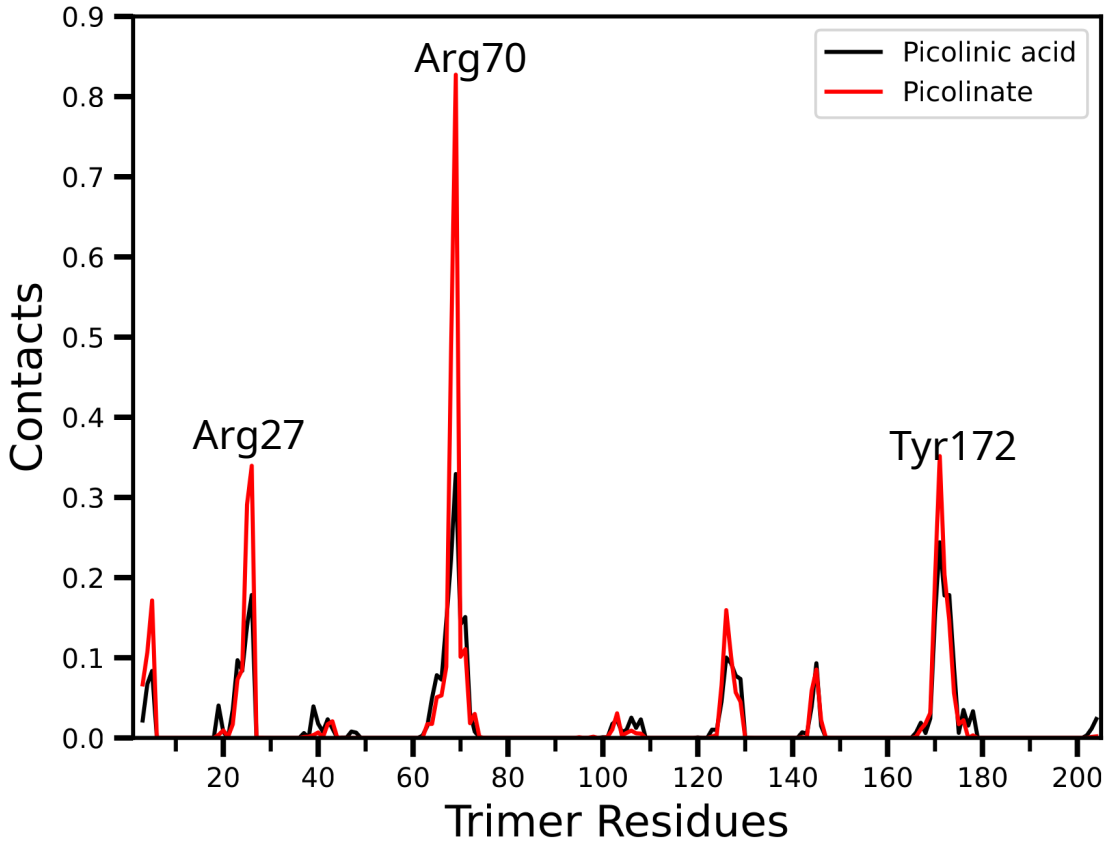

Figure S12: Number of contacts formed between picolinate, picolinic acid and BMC-T^{SD} residues. BMC-T forms a stacked trimer (Fig. S9) composed of three identical protomer subunits. For clarity, all contacts are mapped onto a single representative monomer. Contact frequencies were normalized across all trajectory frames and symmetrically equivalent monomeric unit.

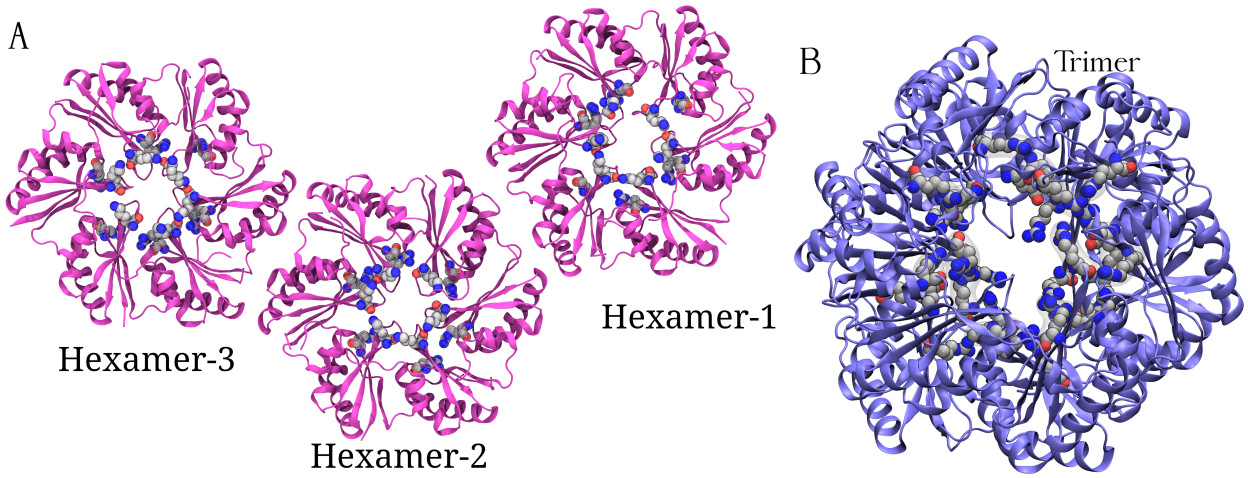

Figure S13: Basic amino acids lining the (A) hexameric and (B) stacked trimeric pores.

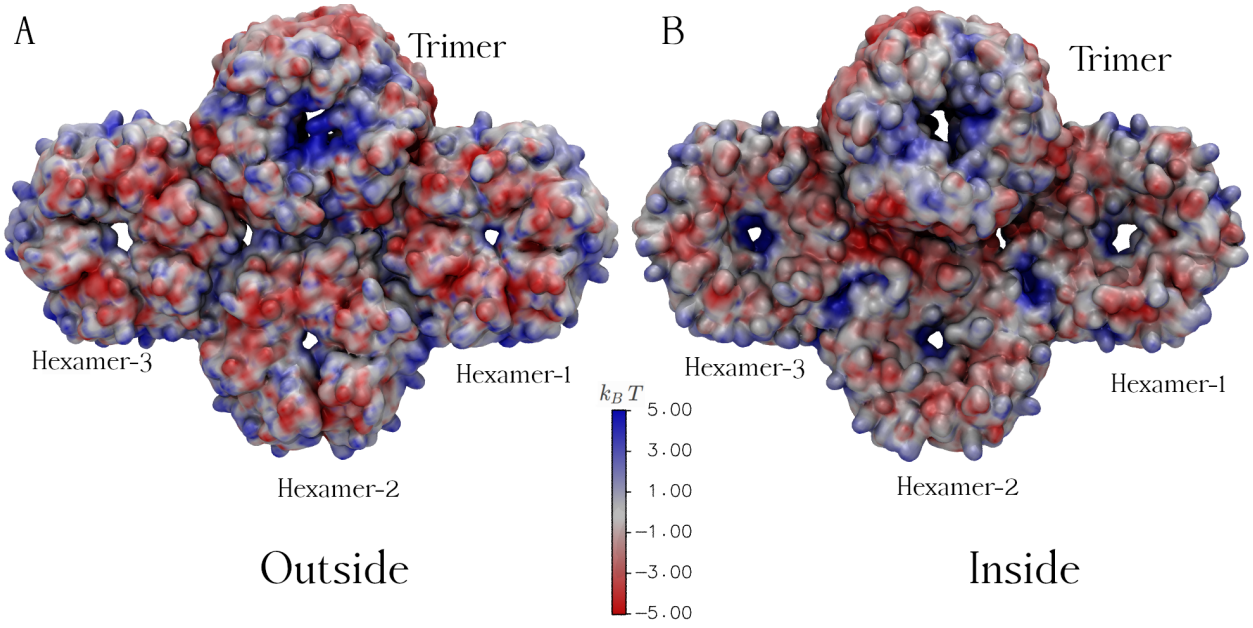

Figure S14: Electrostatics of HO shell. A, Surface as seen from the outside colored by electrostatic potential (red, negative; blue, positive)B, Same as A, but for the inside view. The scale bar below indicates the range from -5 to +5 kβT.

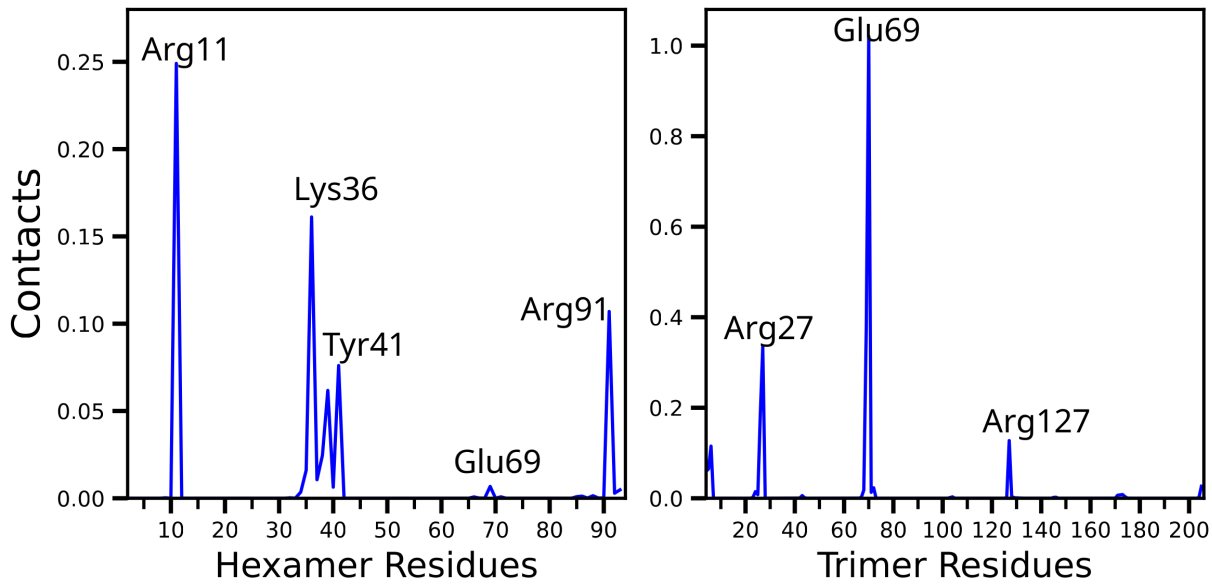

Figure S15: Number of contacts formed between dithionite and BMC-H and BMC-T residues. The BMC shell contains three hexamers, each composed of six identical protomer subunits. (Fig. S9), while the trimer is composed of three repeating monomeric units For clarity, all contacts are projected onto a single representative monomer. Contact frequencies were normalized over all trajectory frames and symmetrically equivalent monomeric units.

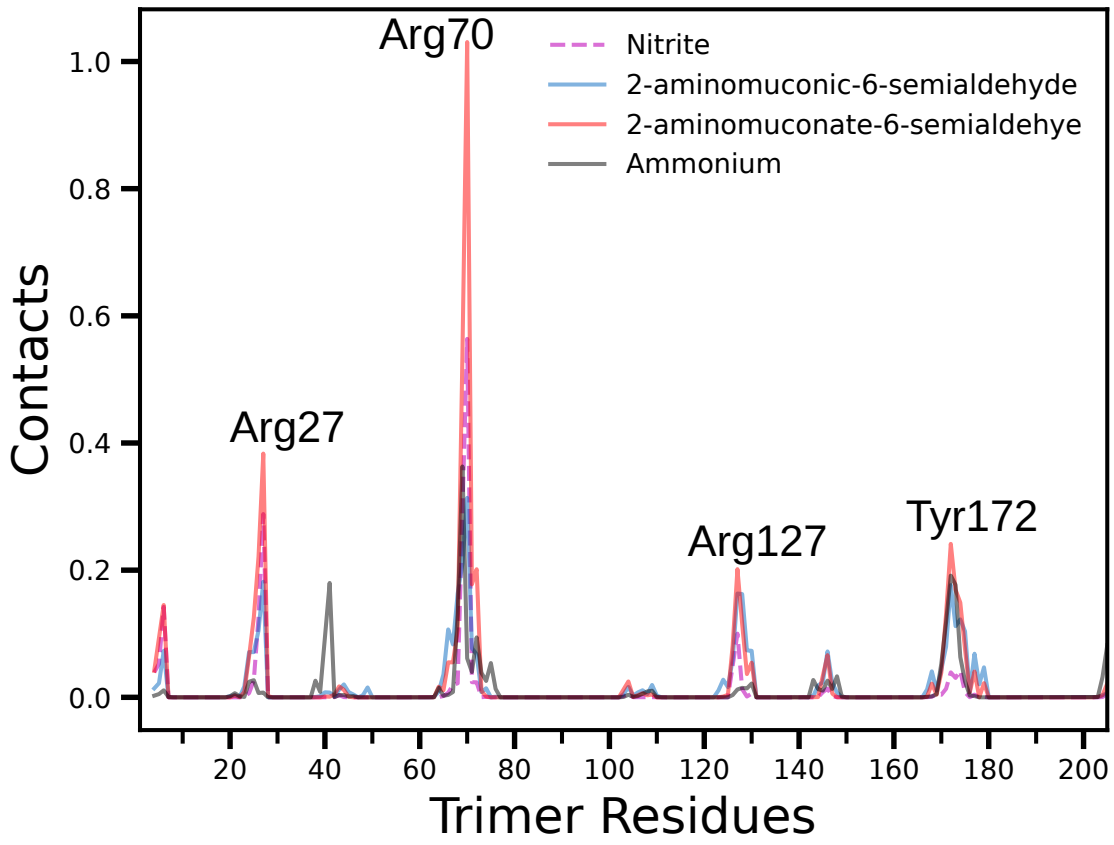

Figure S16: Number of contacts formed between nitrite, ammonium, 2-aminomuconic acid-6-semialdehyde and 2-aminomuconate-6-semialdehyde with BMC-T residues. BMC-T forms a stacked trimer (Fig. S9) composed of three identical protomer subunits. For clarity, all contacts are mapped onto a single representative monomer. Contact frequencies were normalized across all trajectory frames and symmetrically equivalent monomeric unit.

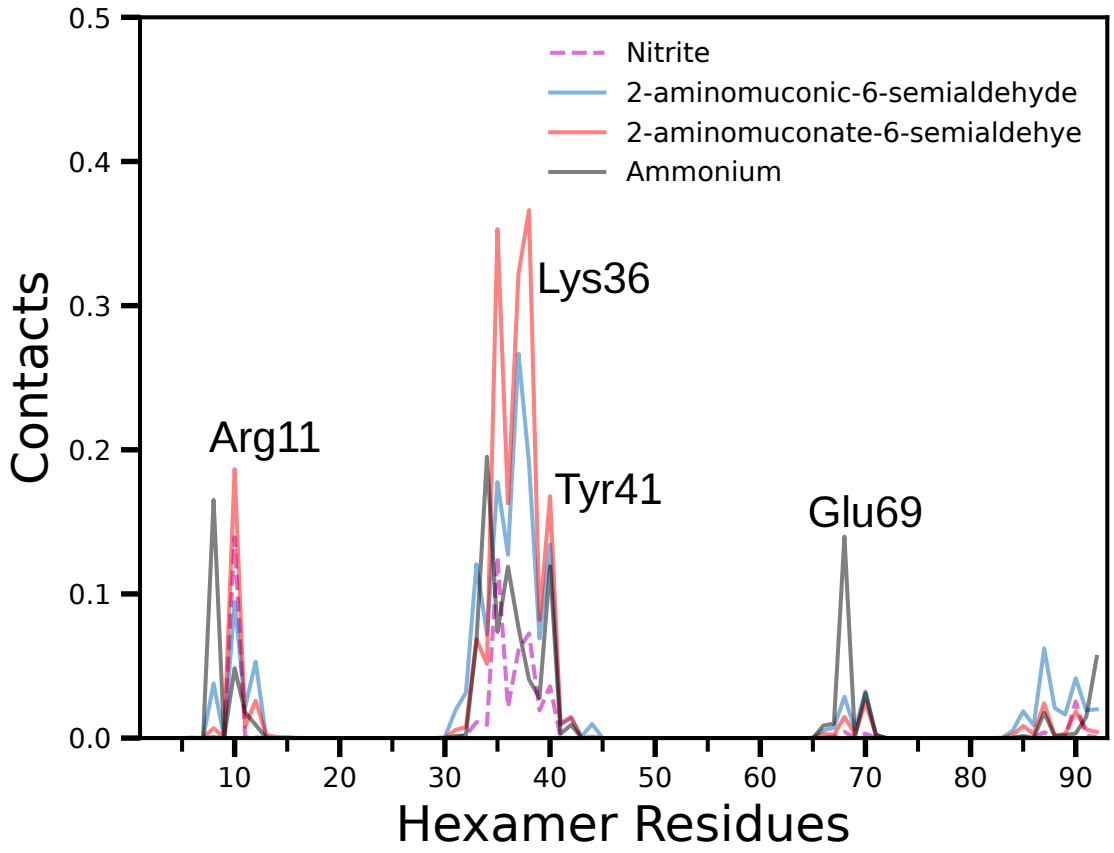

Figure S17: Number of contacts formed between nitrite, ammonium, 2-aminomuconic acid-6-semialdehyde and 2-aminomuconate-6-semialdehyde with BMC-H. The BMC shell contains three hexamers, each composed of six hree identical protomer subunits. (Fig. S9). For clarity, all contacts are projected onto a single representative monomer. Contact frequencies were normalized over all trajectory frames and symmetrically equivalent monomeric units.