Electronic Supplementary Information for

Large-scale QM/MM free energy simulations of enzyme catalysis reveal the influence of charge transfer

Heather J. Kulik[1]

[1]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

Contents

Table S1. Residue protonation states in COMT. If the residue is neutral, no value is indicated.

GLY2		ILE49		ALA96		TRP143		GLU190	-1
ASP3	-1	VAL50		ALA97		LYS144	1	CYS191	
THR4		ASP51	-1	ILE98		ASP145	-1	THR192	
LYS5	1	ALA52		THR99		ARG146	1	HID193	
GLU6	-1	VAL53		GLN100		TYR147		TYR194	
GLN7		ILE54		ARG101	1	LEU148		GLN195	
ARG8	1	GLN55		MET102		PRO149		SER196	
ILE9		GLU56	-1	VAL103		ASP150	-1	PHE197	
LEU10		HIE57		ASP104	-1	THR151		LEU198	
ASN11		GLN58		PHE105		LEU152		GLU199	-1
HID12		PRO59		ALA106		LEU153		TYR200	
VAL13		SER60		GLY107		LEU154		ARG201	1
LEU14		VAL61		VAL108		GLU155	-1	GLU202	-1
GLN15		LEU62		LYS109	1	GLU156	-1	VAL203	
HIE16		LEU63		ASP110	-1	CYS157		VAL204	
ALA17		GLU64	-1	LYS111	1	GLY158		ASP205	-1
GLU18	-1	LEU65		VAL112		LEU159		GLY206	
PRO19		GLY66		THR113		LEU160		LEU207	
GLY20		ALA67		LEU114		ARG161	1	GLU208	-1
ASN21		TYR68		VAL115		LYS162	1	LYS209	1
ALA22		CYS69		VAL116		GLY163		ALA210	
GLN23		GLY70		GLY117		THR164		ILE211	
SER24		TYR71		ALA118		VAL165		TYR212	
VAL25		SER72		SER119		LEU166		LYS213	1
LEU26		ALA73		GLN120		LEU167		GLY214	
GLU27	-1	VAL74		ASP121	-1	ALA168		PRO215	
ALA28		ARG75	1	ILE122		ASP169	-1	MG215	2

ILE29		MET76		ILE123		ASN170		SAM216	1
ASP30	-1	ALA77		PRO124		VAL171		CAT217	-1
THR31		ARG78	1	GLN125		ILE172			
TYR32		LEU79		LEU126		CYS173			
CYS33		LEU80		LYS127	1	PRO174			
GLU34	-1	SER81		LYS128	1	GLY175			
GLN35		PRO82		LYS129	1	ALA176			
LYS36	1	GLY83		TYR130		PRO177			
GLU37	-1	ALA84		ASP131	-1	ASP178	-1		
TRP38		ARG85	1	VAL132		PHE179			
ALA39		LEU86		ASP133	-1	LEU180			
MET40		ILE87		THR134		ALA181			
ASN41		THR88		LEU135		HIE182			
VAL42		ILE89		ASP136	-1	VAL183			
GLY43		GLU90	-1	MET137		ARG184	1		
ASP44	-1	ILE91		VAL138		GLY185			
LYS45	1	ASN92		PHE139		SER186			
LYS46	1	PRO93		LEU140		SER187			
GLY47		ASP94	-1	ASP141	-1	CYS188			
LYS48	1	CYS95		HIE142		PHE189			

Table S2. Umbrella sampling window targets (r_0), where r_0 is a target linear combination of distances (i.e., SAM S-C minus SAM C-O(catecholate), in Å) and Amber force constants (k). The force constants used here are consistent with the Amber convention $k(r\text{-}r_0)^2$ that omits the factor of 1/2 and are in units of kcal/mol·Å^2.

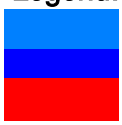
r_0 (Å)	k (kcal/mol·Å^2)
-1.15	10
-0.80	20
-0.40	20
-0.20	40
0.00	240
0.10	40
0.15	240

0.30	240
0.45	120
0.50	80
0.60	40
0.90	40
1.10	40
1.55	10

Table S3. QM region size choices in QM/MM free energy calculations. H2O1 refers to a Mg^{2+}-coordinating water molecule obtained originally from the crystal structure, 3BWM. Residue numbers are from original protein structure, and the total number of atoms is the number in the #at. column plus the # link column. Residue count excludes substrates or water or Mg^{2+}. The charge is the sum of individual residue charges, following color code legend.

Region	# at.	# res.	# link	charge	Residue list						
M1	64	0	0	2	Mg2+	SAM	CAT				
M2	105	3	4	0	D141	D169	N170	Mg2+	SAM	CAT	H2O1
M3	158	7	12	-2	V42	G66	E90	D141	D169	N170	E199
					Mg2+	SAM	CAT	H2O1			
M4	307	16	18	-3	M40	N41	V42	E64	A67	Y68	Y71
					S72	E90	I91	S119	D141	H142	D169
					N170	E199	Mg2+	SAM	CAT	H2O1	
M5	518	28	26	-1	W38	M40	N41	V42	E64	G66	A67
					Y68	Y71	S72	I89	E90	I91	N92
					C95	A118	S119	Q120	F139	D141	H142
					W143	K144	R146	D169	N170	L198	E199
					Mg2+	SAM	CAT	H2O1			

Legend:



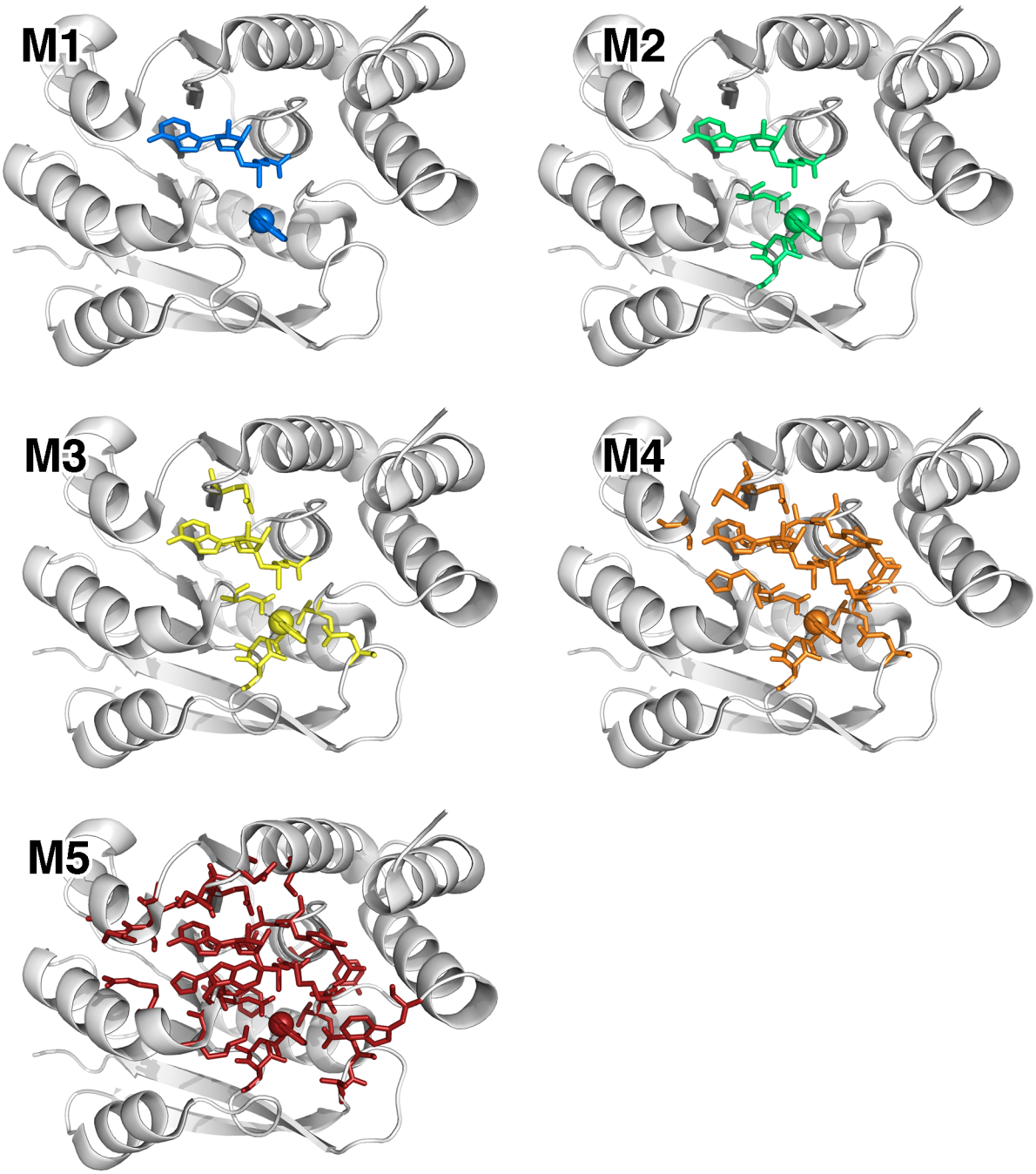


Figure S1. Separate figures of each QM region: **M1**, **M2**, **M3, M4**, and **M5** as in the main text.

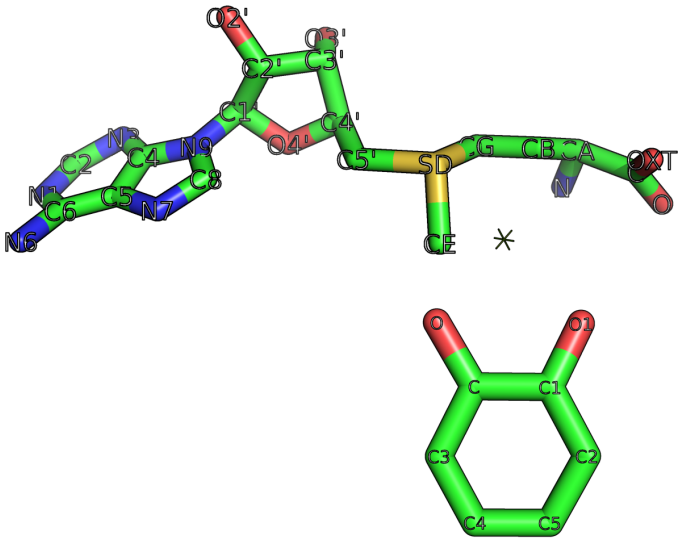


Figure S2. Labels of SAM and CAT for the closest atom analysis except in i) the CATO- corresponds to the CAT O atom and CATOH corresponds to CAT O1 and ii) the oxygen atom on the carboxylate in SAM called "O" in the figure is instead OXT2 in Table S2.

Table S4. Summary of representative reactant, transition state, and product closest heavy atom distances (C A, in Å) between residues and substrates along with average quantities (av(C A)). The R A is the residue atom closest to the substrate and the S A is the substrate atom that is closest, with atom labels for substrates, as indicated in Figure S1.

	Reactant			Transition State			Product			
Res.	C A	S A	R A	C A	S A	R A	C A	S A	R A	av(C A)
TRP38	4.68	CATC2	CD1	4.25	CATC5	CD1	3.36	CATC5	CD1	4.10
MET40	3.13	CATO-	CB	3.32	SAMSD	CE	3.38	CATC	CG	3.28
ASN41	2.73	SAMOXT	N	2.74	SAMOXT	ND2	2.65	SAMOXT	ND2	2.71
VAL42	2.85	SAMOXT2	N	2.79	SAMOXT2	N	2.87	SAMOXT2	N	2.84
GLU64	5.21	SAMN	OE2	5.23	SAMN	OE1	4.42	SAMCA	OE2	4.95
GLY66	2.53	SAMO3	C	3.47	SAMO3	O	3.42	SAMCG	O	3.14
ALA67	2.80	SAMO3	N	4.34	SAMO3	C	4.62	SAMO3	C	3.92
TYR68	2.99	SAMO3	N	2.80	SAMO3	CA	2.92	SAMO3	CA	2.90
TYR71	2.79	SAMOXT	CE2	2.99	SAMOXT	CD2	2.93	SAMOXT	CD2	2.90
SER72	3.77	SAMN	CB	3.10	SAMN	OG	3.08	SAMOXT	N	3.32
ILE89	4.01	SAMN3	O	3.52	SAMN3	O	3.61	SAMN3	O	3.71
GLU90	2.45	SAMO3	OE1	2.46	SAMO3	OE1	2.61	SAMO3	OE1	2.51
ILE91	2.89	SAMO2	N	3.20	SAMN1	CD1	3.30	SAMN3	N	3.13
ASN92	2.96	SAMO2	N	2.89	SAMO2	ND2	2.70	SAMO2	ND2	2.85
CYS95	9.89	SAMO2	C	5.68	SAMO2	CB	6.35	SAMO2	CB	7.31
ALA118	5.30	SAMC2	N	5.18	SAMC2	N	5.40	SAMN1	N	5.29
SER119	6.53	SAMC2	N	4.72	SAMN1	CB	4.67	SAMN1	CB	5.31
GLN120	9.59	SAMC2	N	7.19	SAMN1	N	7.13	SAMN1	N	7.97

PHE139	4.64	SAMN	CE1	3.46	SAMN	CE2	4.86	SAMN	CZ	4.32
ASP141	2.01	MG	OD1	2.16	MG	OD1	2.12	MG	OD1	2.10
HIS142	2.77	SAMO4	CB	3.19	SAMO4	CA	3.25	SAMO4	CA	3.07
TRP143	3.43	SAMN6	CB	3.32	SAMN7	CB	3.10	SAMN7	CB	3.28
LYS144	4.49	CATO-	NZ	3.67	CATO-	NZ	3.81	CATO-	NZ	3.99
ARG146	8.78	SAMC2	N	6.50	SAMN6	CB	6.88	SAMN6	CB	7.39
ASP169	1.98	MG	OD2	2.06	MG	OD2	1.98	MG	OD2	2.01
ASN170	2.08	MG	OD1	2.30	MG	OD1	2.07	MG	OD1	2.15
LEU198	8.16	CATOH	CD1	6.95	CATOH	CD1	7.06	CATOH	CD1	7.39
GLU199	2.60	CATOH	OE1	2.37	CATOH	OE1	2.44	CATOH	OE1	2.47

Table S5. Comparison of formal charge (FC), median by-residue partial charge, difference of the two quantities, standard deviation of the partial charge, 1st quartile and 3rd quartile values and their differences, minimum and maximum values and their differences all in e. Quantities include link atoms for relevant residues.

Res.	FC	med.	diff.	std. dev.	Q1	Q3	min.	max.	Q1-Q3	max-min
TRP38	0	-0.01	-0.01	0.014	-0.02	0.00	0.03	-0.11	-0.02	0.14
MET40	0	0.08	0.08	0.032	0.06	0.10	0.22	-0.07	-0.04	0.28
ASN41	0	-0.09	-0.09	0.060	-0.13	-0.06	0.06	-0.40	-0.07	0.46
VAL42	0	-0.14	-0.14	0.039	-0.16	-0.12	-0.02	-0.30	-0.05	0.28
GLU64	-1	-0.76	0.24	0.108	-0.80	-0.70	-0.43	-0.90	-0.10	0.48
GLY66	0	0.09	0.09	0.027	0.07	0.11	0.18	-0.06	-0.04	0.24
ALA67	0	-0.05	-0.05	0.040	-0.08	-0.03	0.16	-0.21	-0.04	0.37
TYR68	0	-0.13	-0.13	0.040	-0.15	-0.10	0.02	-0.38	-0.04	0.40
TYR71	0	0.06	0.06	0.038	0.03	0.08	0.19	-0.09	-0.05	0.28
SER72	0	-0.22	-0.22	0.062	-0.25	-0.18	0.02	-0.36	-0.08	0.39
ILE89	0	0.12	0.12	0.041	0.09	0.15	0.22	-0.01	-0.06	0.23
GLU90	-1	-0.70	0.30	0.045	-0.73	-0.67	-0.53	-0.87	-0.06	0.34
ILE91	0	-0.08	-0.08	0.041	-0.11	-0.05	0.06	-0.23	-0.06	0.29
ASN92	0	-0.12	-0.12	0.035	-0.14	-0.10	0.00	-0.28	-0.05	0.28
CYS95	0	-0.02	-0.02	0.017	-0.04	-0.01	0.00	-0.07	-0.03	0.08
ALA118	0	0.07	0.07	0.035	0.05	0.09	0.22	-0.07	-0.05	0.29
SER119	0	0.02	0.02	0.060	-0.02	0.06	0.20	-0.18	-0.08	0.39
GLN120	0	-0.08	-0.08	0.034	-0.16	-0.08	0.03	-0.25	-0.08	0.28
PHE139	0	-0.02	-0.02	0.026	-0.04	-0.01	0.08	-0.12	-0.03	0.21
ASP141	-1	-0.50	0.50	0.030	-0.52	-0.48	-0.41	-0.64	-0.04	0.23
HIS142	0	0.10	0.10	0.051	0.08	0.13	0.33	-0.04	-0.06	0.37
TRP143	0	-0.02	-0.02	0.040	-0.05	0.00	0.10	-0.20	-0.05	0.30
LYS144	+1	0.78	-0.22	0.036	0.76	0.81	0.92	0.62	-0.05	0.29
ARG146	+1	0.99	-0.01	0.017	0.98	1.00	1.04	0.90	-0.02	0.14
ASP169	-1	-0.55	0.45	0.027	-0.57	-0.54	-0.46	-0.68	-0.04	0.22
ASN170	0	-0.07	-0.07	0.031	-0.09	-0.05	0.08	-0.23	-0.04	0.30

LEU198	0	0.15	0.15	0.042	0.11	0.17	0.26	-0.01	-0.06	0.27
GLU199	-1	-0.96	0.04	0.109	-1.01	-0.89	-0.52	-1.16	-0.12	0.64
WAT296	0	0.03	0.03	0.019	0.02	0.04	0.14	-0.07	-0.02	0.22

Table S6. Comparison of formal charge (FC), median by-residue partial charge, difference of the two quantities, standard deviation of the partial charge, 1st quartile and 3rd quartile values and their differences, minimum and maximum values and their differences all in e. Quantities **do not** include link atoms.

Res.	FC	med.	std. dev.	Q1	Q3	max.	min.	Q1-Q3	max-min
TRP38	0	-0.53	0.029	-0.55	-0.51	-0.40	-0.64	-0.04	0.23
MET40	0	-0.39	0.035	-0.42	-0.37	-0.25	-0.54	-0.05	0.29
ASN41	0	-0.09	0.060	-0.13	-0.06	0.06	-0.40	-0.07	0.46
VAL42	0	-0.18	0.039	-0.21	-0.16	0.00	-0.36	-0.05	0.36
GLU64	-1	-1.28	0.100	-1.31	-1.23	-0.98	-1.43	-0.09	0.46
GLY66	0	-0.40	0.028	-0.41	-0.38	-0.30	-0.52	-0.04	0.22
ALA67	0	-0.05	0.040	-0.08	-0.03	0.16	-0.21	-0.04	0.37
TYR68	0	-0.12	0.044	-0.15	-0.10	0.06	-0.38	-0.05	0.44
TYR71	0	-0.43	0.040	-0.46	-0.40	-0.27	-0.58	-0.05	0.30
SER72	0	-0.28	0.072	-0.31	-0.22	0.02	-0.43	-0.09	0.46
ILE89	0	-0.38	0.048	-0.41	-0.34	-0.27	-0.51	-0.08	0.24
GLU90	-1	-0.70	0.045	-0.73	-0.67	-0.53	-0.87	-0.06	0.34
ILE91	0	-0.08	0.041	-0.11	-0.05	0.06	-0.23	-0.06	0.29
ASN92	0	-0.26	0.039	-0.29	-0.23	-0.08	-0.41	-0.05	0.33
CYS95	0	-0.53	0.025	-0.55	-0.51	-0.42	-0.62	-0.03	0.20
ALA118	0	-0.41	0.037	-0.44	-0.39	-0.27	-0.58	-0.05	0.31
SER119	0	0.02	0.060	-0.02	0.06	0.20	-0.18	-0.08	0.39
GLN120	0	-0.09	0.041	-0.12	-0.06	0.06	-0.26	-0.06	0.33
PHE139	0	-0.58	0.028	-0.60	-0.56	-0.48	-0.69	-0.04	0.21
ASP141	-1	-0.99	0.034	-1.01	-0.96	-0.88	-1.13	-0.05	0.25
HIS142	0	0.10	0.051	0.08	0.13	0.33	-0.04	-0.06	0.37
TRP143	0	-0.02	0.040	-0.05	0.00	0.10	-0.20	-0.05	0.30
LYS144	+1	0.76	0.042	0.73	0.78	0.91	0.57	-0.05	0.34
ARG146	+1	0.52	0.031	0.50	0.54	0.63	0.38	-0.04	0.25
ASP169	-1	-1.04	0.026	-1.06	-1.02	-0.95	-1.17	-0.03	0.22
ASN170	0	-0.09	0.037	-0.12	-0.07	0.08	-0.30	-0.05	0.38
LEU198	0	-0.33	0.046	-0.37	-0.30	-0.21	-0.50	-0.06	0.29
GLU199	-1	-0.98	0.111	-1.04	-0.91	-0.56	-1.22	-0.13	0.66
WAT296	0	0.03	0.019	0.02	0.04	0.14	-0.07	-0.02	0.22

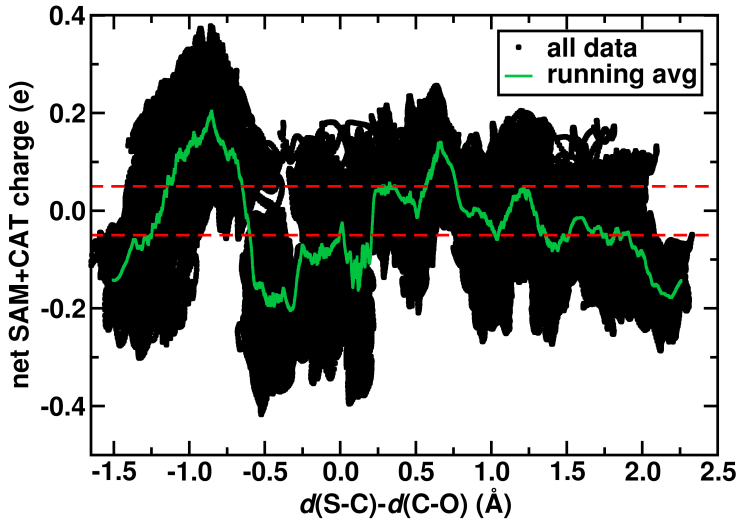


Figure S3. Variation in summed SAM and catecholate charges indicated over the methyl transfer reaction coordinate. Each discrete point is indicated as a black circle, and a 1000-length running average is indicated in green. The red bars represent a +/- 0.05 e range.

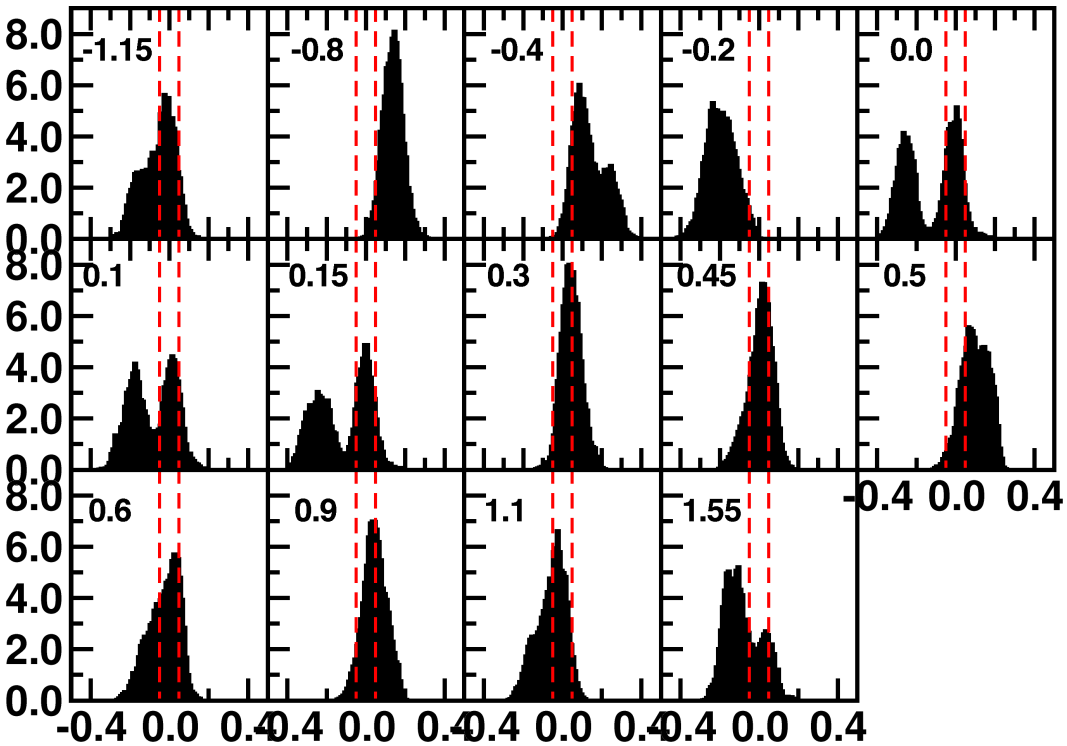


Figure S4. Histogram of summed SAM and catecholate charges for each window target in the free energy simulation. The red bars represent +/- 0.05 e range and each bin is 0.01 e wide. The x- and y-axis range on each plot is the same and the window target position is indicated in inset for each plot.

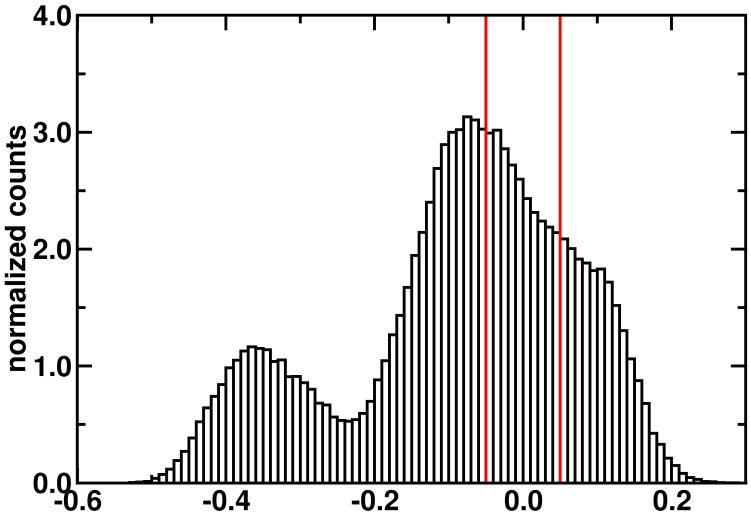


Figure S5. Normalized histogram of partial charge deviation from formal charge (in e) summed over SAM, CAT, Mg^{2+}, active site water, and 14 residues with a minimum closest heavy atom distance within 3.0 Å of the substrate (i.e., Asn41, Val42, Gly66, Ala67, Tyr68, Tyr71, Glu90, Ile91, Asn92, Asp141, His142, Asp169, Asn170, and Glu199) across 200 ps of aggregated M5 free energy simulation. The red bars represent +/- 0.05 e range, which contains 26% of the data, and each bin is 0.01 e wide. The total region summed over should have a formal charge of -2, which has been subtracted.

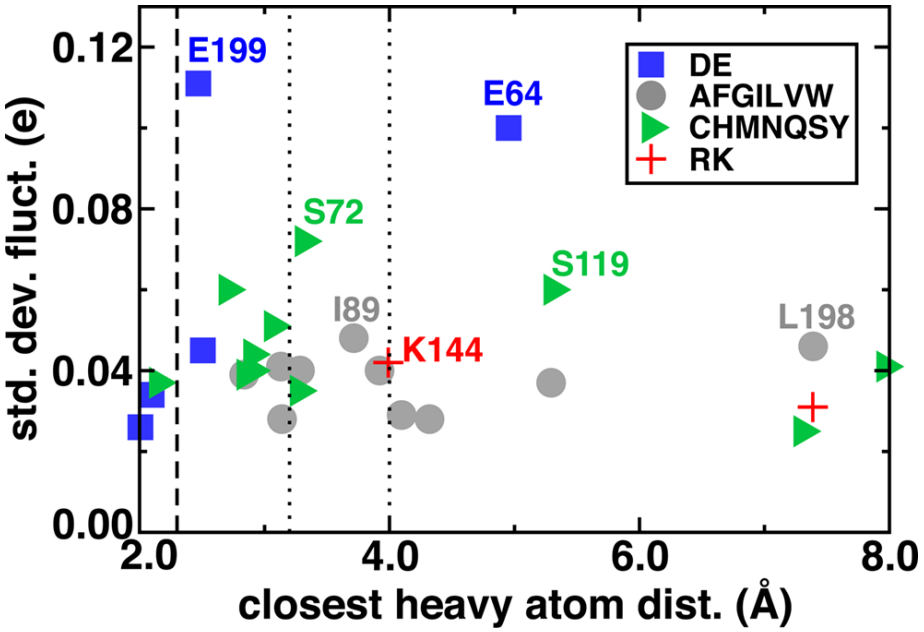


Figure S6. Standard deviations of by-residue summed partial charges **without** link atoms (in e) sampled for each residue grouped by negatively charged (blue squares), nonpolar (gray circles), polar (green triangles), and positively charged (red crosses), with residue types indicated in legend plotted against the average closest heavy atom distance between the residue and SAM, CAT, or Mg^{2+}. Vertical lines indicate approximate covalent bond (dashed), moderate hydrogen bond (dotted), and weak hydrogen bond (dotted) distance cutoffs.

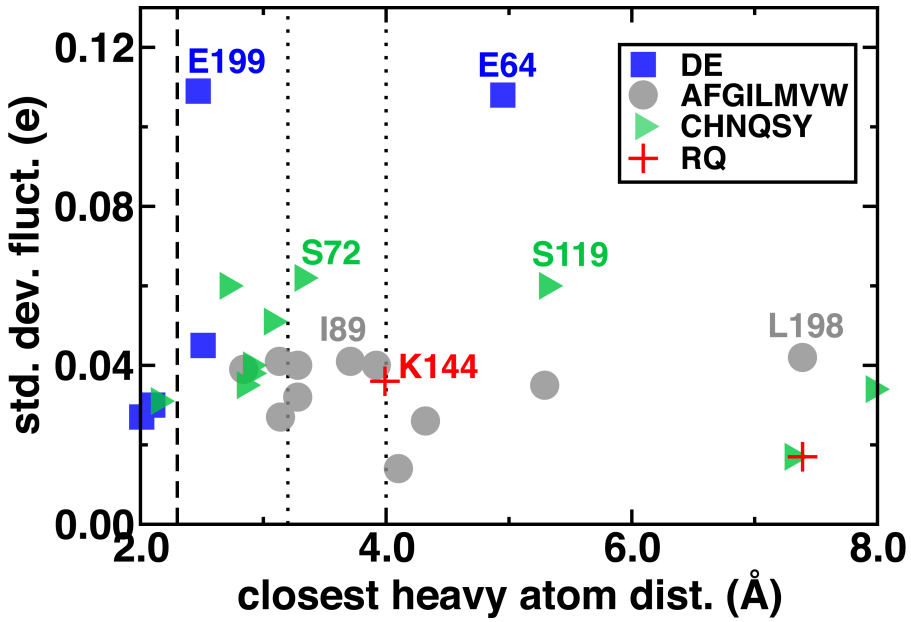


Figure S7. Standard deviations of by-residue summed partial charges **including** link atoms (in e) sampled for each residue grouped by negatively charged (blue squares), nonpolar (gray circles), polar (green triangles), and positively charged (red crosses), with residue types indicated in legend plotted against the average closest heavy atom distance between the residue and SAM, CAT, or Mg^{2+}. Vertical lines indicate approximate covalent bond (dashed), moderate hydrogen bond (dotted), and weak hydrogen bond (dotted) distance cutoffs.

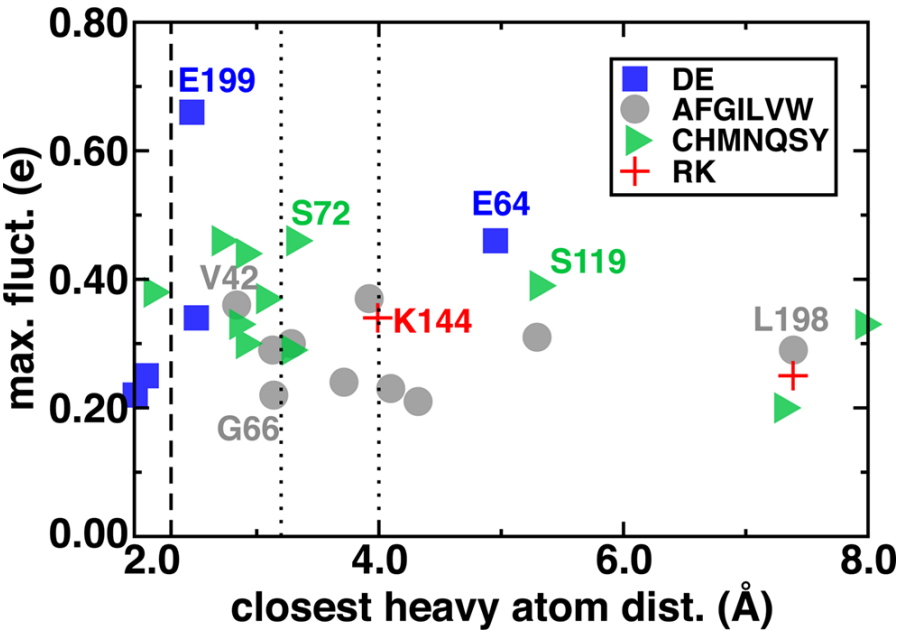


Figure S8. Range of by-residue summed partial charges **without** link atoms (max. fluct., e) sampled for each residue grouped by negatively charged (blue squares), nonpolar (gray circles), polar (green triangles), and positively charged (red crosses), with residue types indicated in legend plotted against the average closest heavy atom distance between the residue and SAM, CAT, or Mg^{2+}. Vertical lines indicate approximate covalent bond (dashed), moderate hydrogen bond (dotted), and weak hydrogen bond (dotted) distance cutoffs.

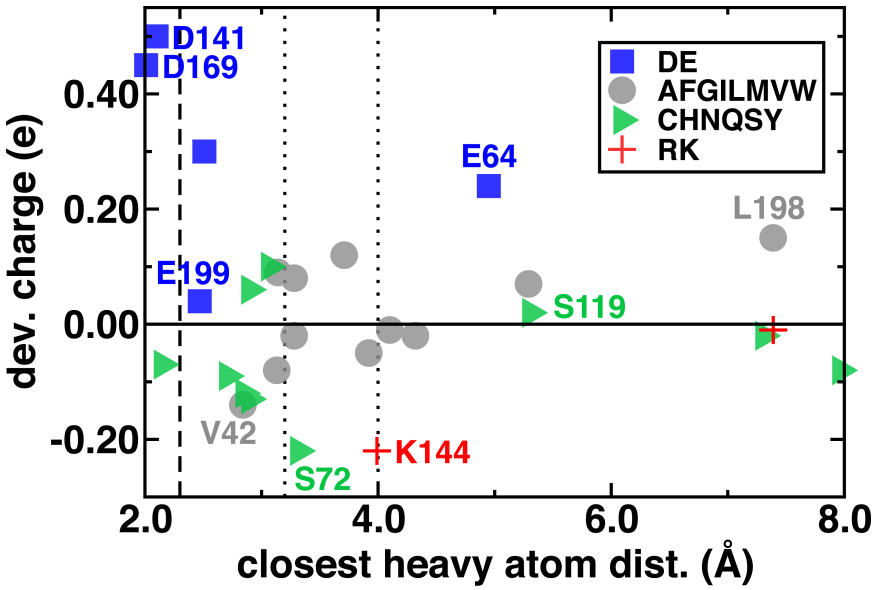


Figure S9. Deviation from formal charge of by-residue summed partial charges including link atoms (in e) sampled for each residue grouped by negatively charged (blue squares), nonpolar (gray circles), polar (green triangles), and positively charged (red crosses), with residue types indicated in legend plotted against the average closest heavy atom distance between the residue and SAM, CAT, or Mg^{2+}. Vertical lines indicate approximate covalent bond (dashed), moderate hydrogen bond (dotted), and weak hydrogen bond (dotted) distance cutoffs.

Text S1. Explanation of shortened Mg^{2+}-water bond.
Evaluation of median water partial charge and its fluctuations (i.e., 0.03 ± 0.02 e std. dev., -0.07 to 0.14 e range) confirms that the water molecule carries a slightly more negative charge than would be expected based on an equal division of the +1 excess charge transferred from the coordination sphere to Mg^{2+} (see Figure 2). This water molecule's proximity to the highly fluctuating V42 as well as a nearby M40 could explain this unexpected behavior for water.

Table S7. Summary of model-dependent barrier height and reaction energy (kcal/mol) along with representative Δ values and associated C-O and S-C bond lengths all in Å.

			Reactants							Transition State							Products						
				C-O			S-C				C-O			S-C				C-O			S-C		
	$\Delta G^{\ddagger}$	ΔG_r	Δ	min	avg	max	min	avg	max	Δ	min	avg	max	min	avg	max	Δ	min	avg	max	min	avg	max
M1	27.9	11.3	-1.37	3.13	3.17	3.26	1.76	1.80	1.89	0.37	1.86	2.02	2.24	2.24	2.39	2.61	1.58	1.47	1.53	1.55	3.05	3.09	3.13
M2	25.7	9.2	-1.32	3.11	3.20	3.34	1.78	1.87	2.02	0.35	1.89	2.01	2.14	2.24	2.37	2.50	1.62	1.50	1.53	1.57	3.13	3.16	3.19
M3	12.7	-16.2	-1.27	3.05	3.16	3.26	1.78	1.88	1.99	0.19	1.96	2.09	2.24	2.16	2.28	2.43	1.92	1.42	1.49	1.58	3.34	3.41	3.51
M4	15.1	-13.9	-1.22	3.01	3.11	3.24	1.78	1.88	2.01	0.27	1.94	2.07	2.24	2.21	2.34	2.51	1.92	1.40	1.48	1.57	3.33	3.40	3.50
M5	19.9	-5.7	-1.17	2.97	3.04	3.14	1.80	1.89	1.96	0.33	1.89	2.06	2.23	2.23	2.40	2.56	1.85	1.43	1.51	1.58	3.29	3.37	3.44

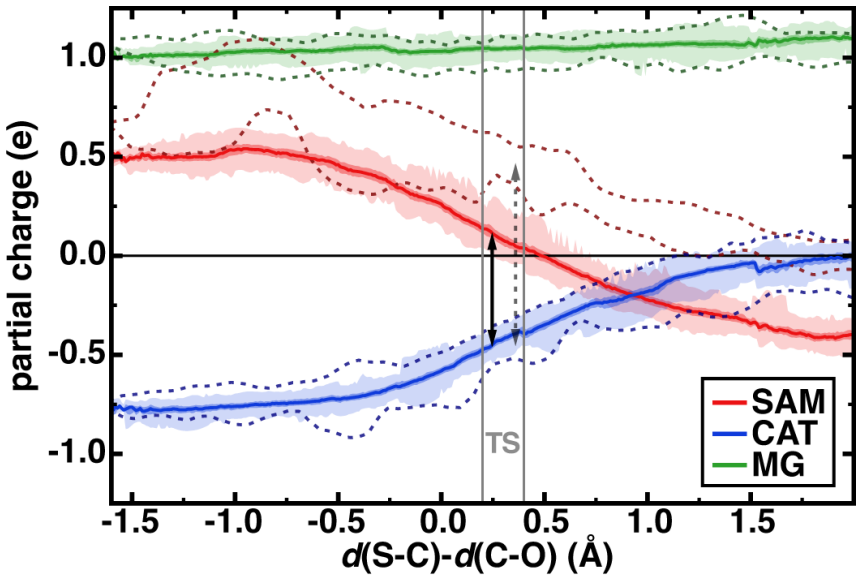


Figure S10. Comparison of M3 and M5 substrate partial charges across the reaction coordinate. The M5 ranges are indicated in dashed lines. For **M3**, sum of charges on SAM (red), CAT (blue), and Mg^{2+} (green) during methyl transfer reaction are shown. Over 0.01 Å bins of simulation data, average charges are indicated by lines, the first to third interquartile range is indicated by the darker shaded region and the full range of charges are indicated by light shading. The methyl charge is partitioned between substrates as described in the main text, and the transition state region is indicated as gray vertical bars with average charge separation shown as a black arrow.

Table S8. Mg^{2+}-O catecholate bond lengths all in Å for M1-M5: anionic Mg^{2+}-O bond in the TS, in the reactants, in the products, neutral Mg^{2+}-O(H) in the transition state and products, shortest anionic Mg^{2+}-O bond in the coordination sphere, and neutral/longest Mg^{2+}-O bond in the coordination sphere, along with percentage of progress of Mg^{2+}-O anionic bond in the TS for main text Figure 5.

Region	Mg-O in TS	Mg-O in R	Mg-O in P	Mg-OH(TS)	MG-OH (P)	Mg-short	Mg-long	
M1	2.04	1.99	2.08	2.11	2.27	1.96	2.09	56%
M2	2.13	2.03	2.17	2.19	2.22	2.03	2.14	74%
M3	2.11	2.03	2.17	2.14	2.13	2.02	2.25	56%
M4	2.08	2.03	2.17	2.12	2.13	2.03	2.25	38%
M5	2.07	2.03	2.16	2.24	2.16	2.02	2.16	30%