Efficient automatic construction of atom-economical QM regions with point-charge variation analysis

Felix Brandt, Christoph R. Jacob[1]

Technische Universität Braunschweig, Institute of Physical and Theoretical Chemistry,
Gaußstraße 17, 38106 Braunschweig, Germany

Supporting Information

Date: May 3, 2023

[1]E-Mail: c.jacob@tu-braunschweig.de

List of Figures

List of Tables

S1 QM region construction with PCVA

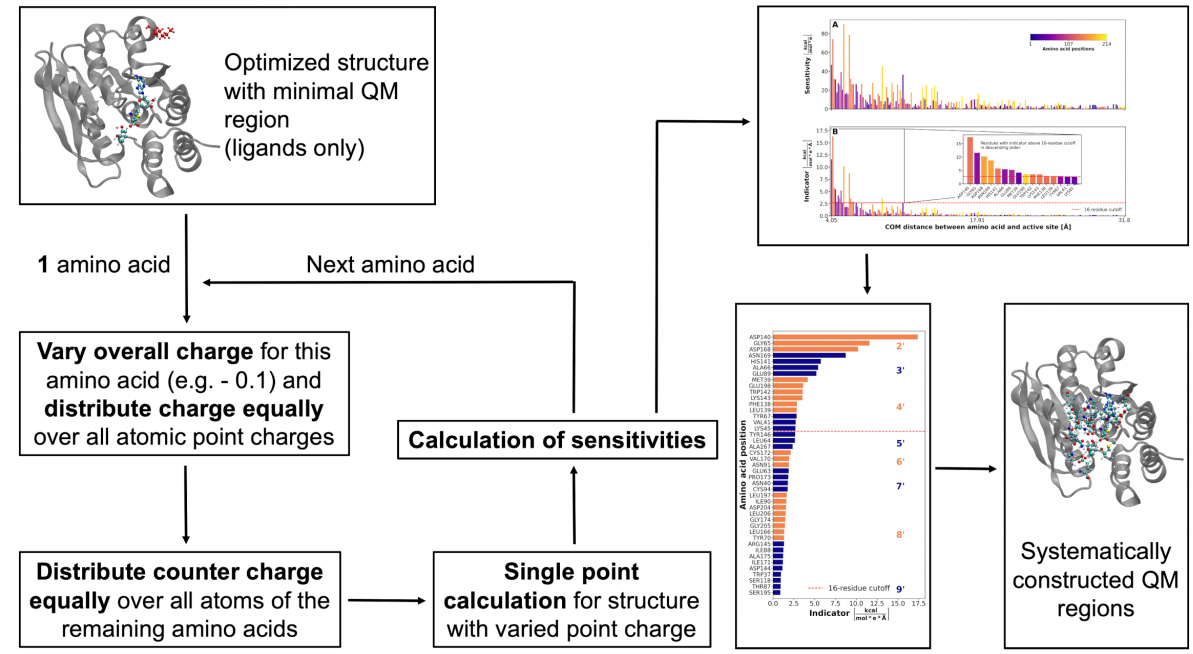

Figure S1: Schematic representation of the workflow for systematic QM region construction based on a single amino acid point charge variation analysis. Reproduced from *J. Chem. Theory Comput.* **18**, 2584-2596 (2022), DOI: 10.1021/acs.jctc.1c01093.

S2 Simplified PCVA schemes for COMT

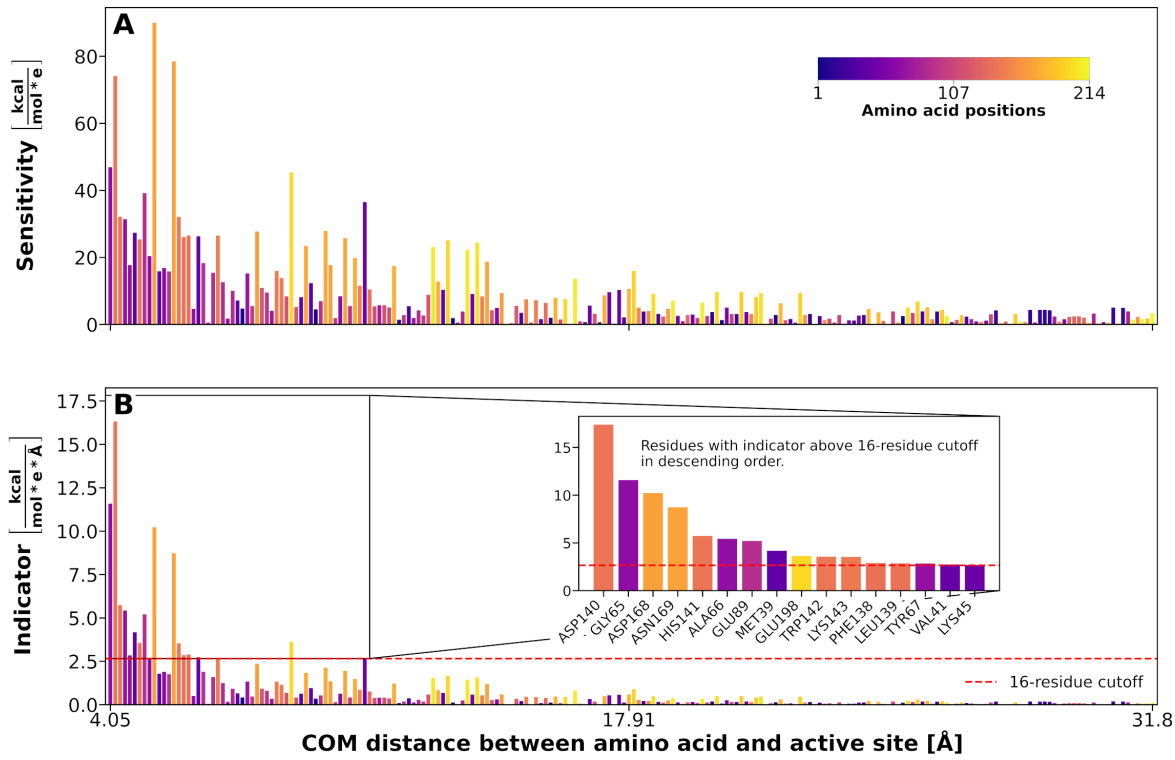

Figure S2: *standard* PCVA for COMT reactant energy. A: QM reactant energy sensitivities of all COMT amino acids sorted by ascending residue-active site distance. B: Corresponding QM region indicators for all amino acids. The 16 residues with the highest indicators are represented in the additional plot sorted by descending indicators.

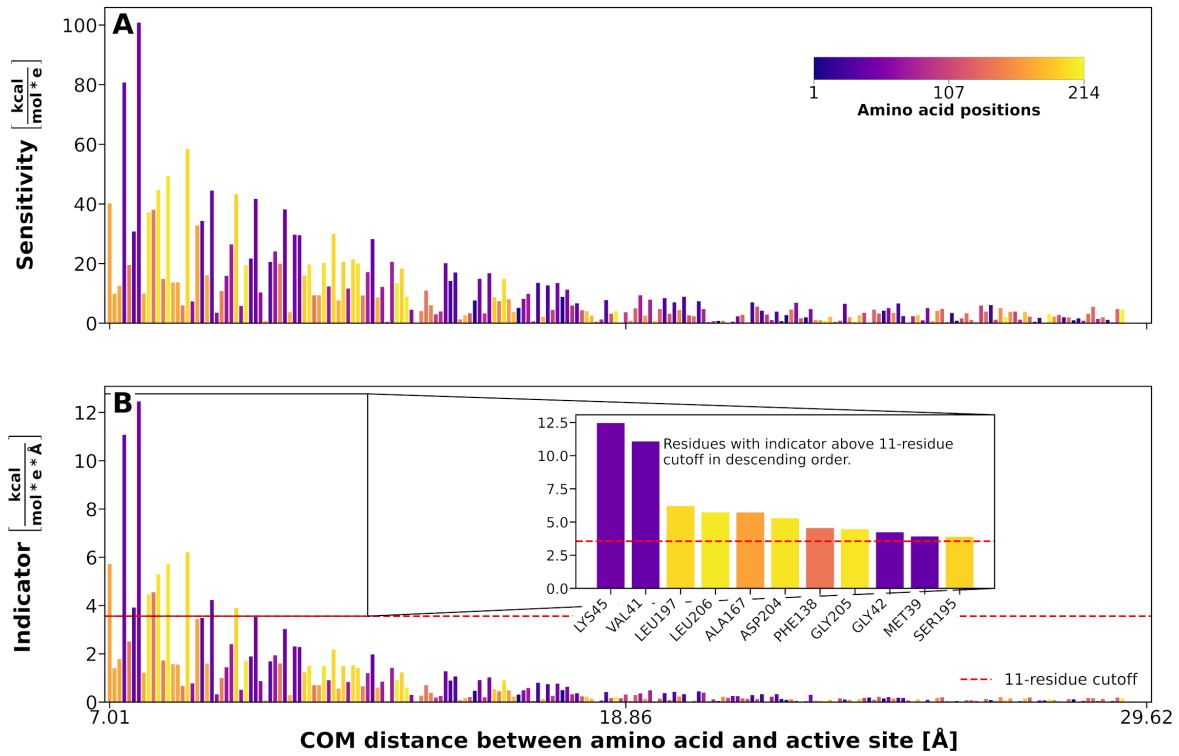

Figure S3: *no lig* PCVA for COMT reactant energy. A: QM reactant energy sensitivities of all COMT amino acids sorted by ascending residue-active site distance. B: Corresponding QM region indicators for all amino acids. The 11 residues with the highest indicators are represented in the additional plot sorted by descending indicators.

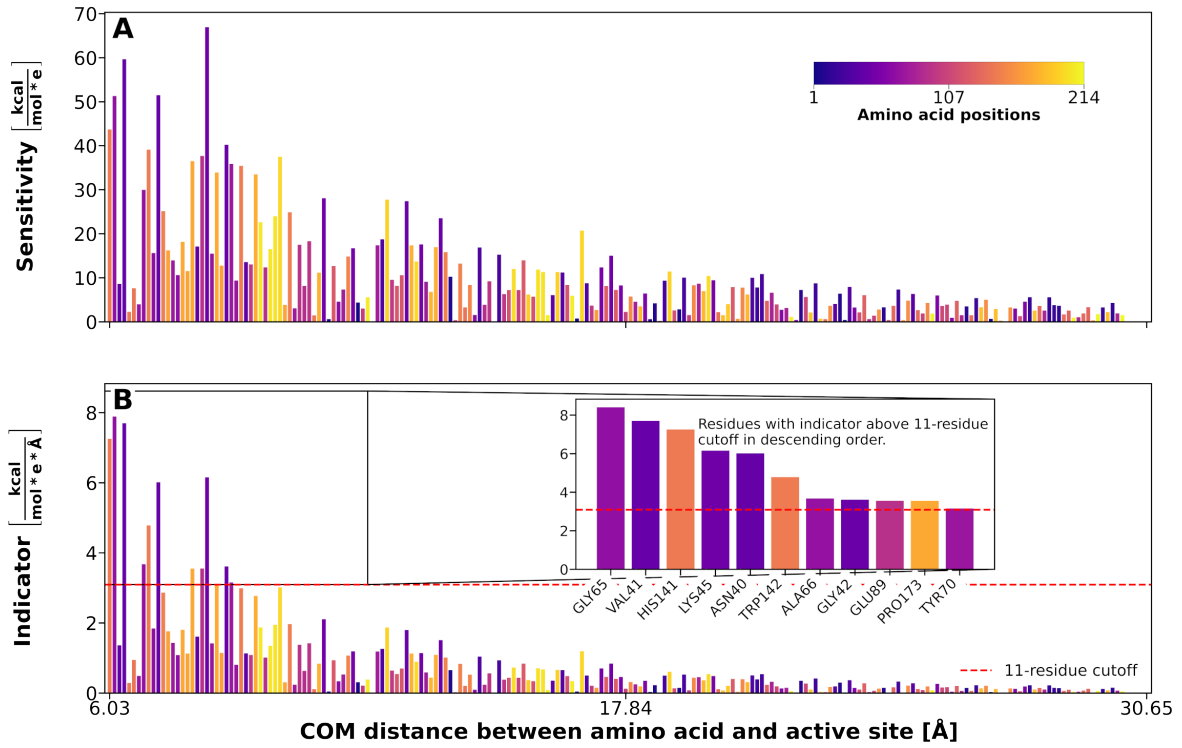

Figure S4: *active* PCVA for COMT reactant energy. A: QM reactant energy sensitivities of all COMT amino acids sorted by ascending residue-active site distance. B: Corresponding QM region indicators for all amino acids. The 11 residues with the highest indicators are represented in the additional plot sorted by descending indicators.

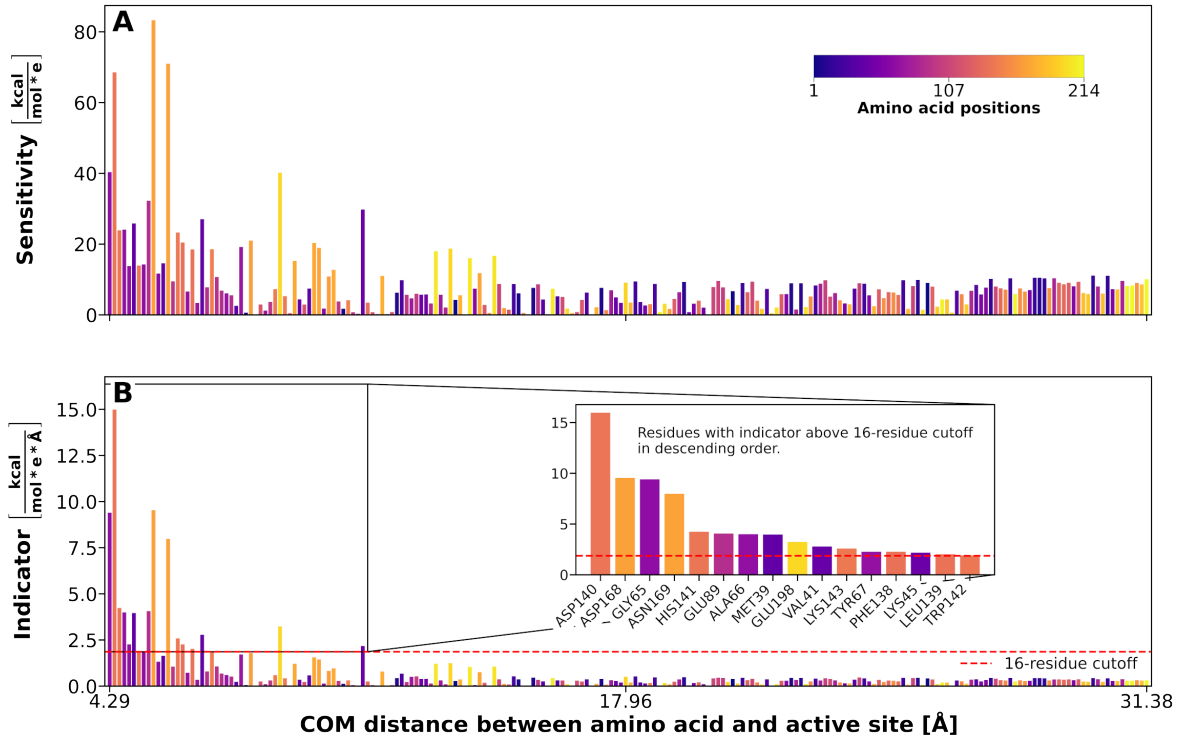

Figure S5: *vac fix* PCVA for COMT reactant energy. A: QM reactant energy sensitivities of all COMT amino acids sorted by ascending residue-active site distance. B: Corresponding QM region indicators for all amino acids. The 16 residues with the highest indicators are represented in the additional plot sorted by descending indicators.

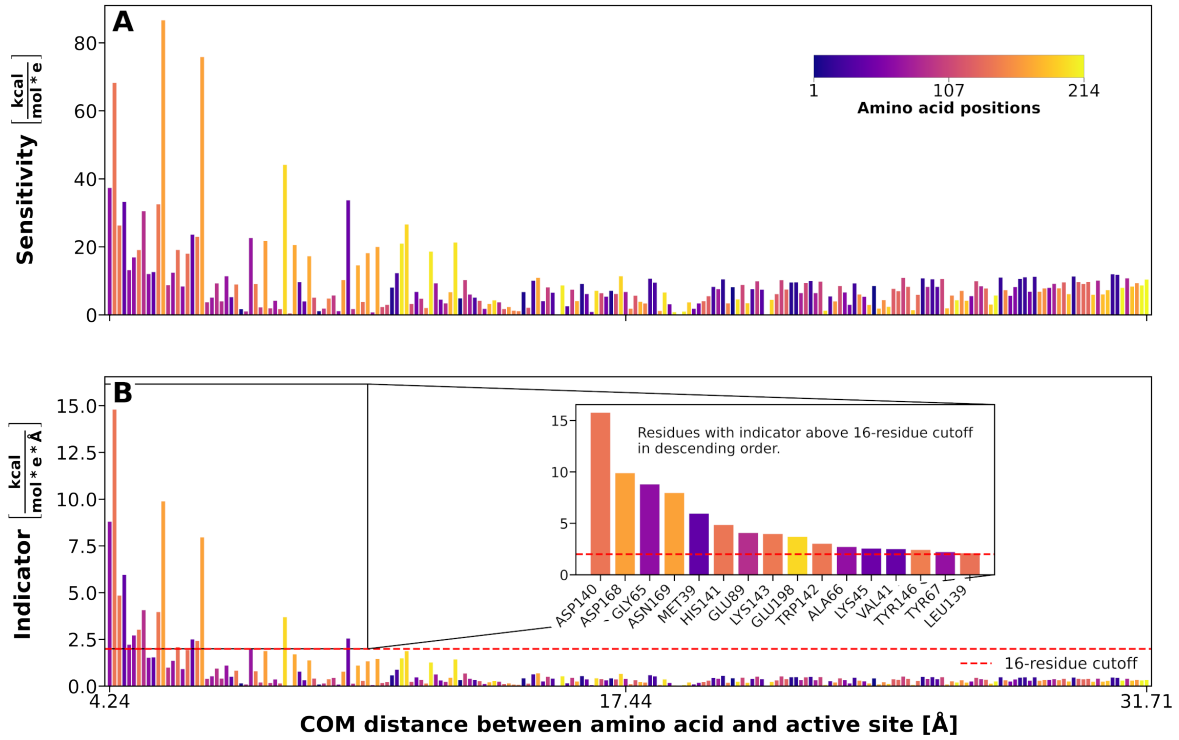

Figure S6: *vac* PCVA for COMT reactant energy. A: QM reactant energy sensitivities of all COMT amino acids sorted by ascending residue-active site distance. B: Corresponding QM region indicators for all amino acids. The 16 residues with the highest indicators are represented in the additional plot sorted by descending indicators.

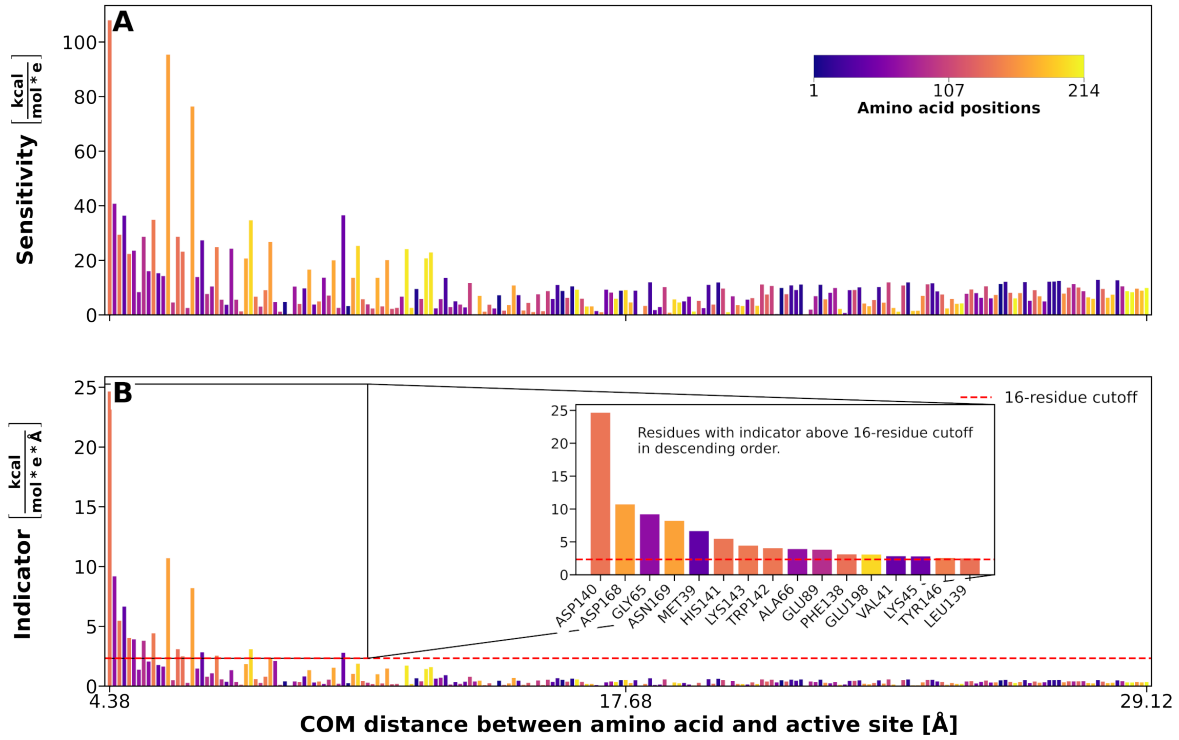

Figure S7: *cryst fix* PCVA for COMT reactant energy. A: QM reactant energy sensitivities of all COMT amino acids sorted by ascending residue-active site distance. B: Corresponding QM region indicators for all amino acids. The 16 residues with the highest indicators are represented in the additional plot sorted by descending indicators.

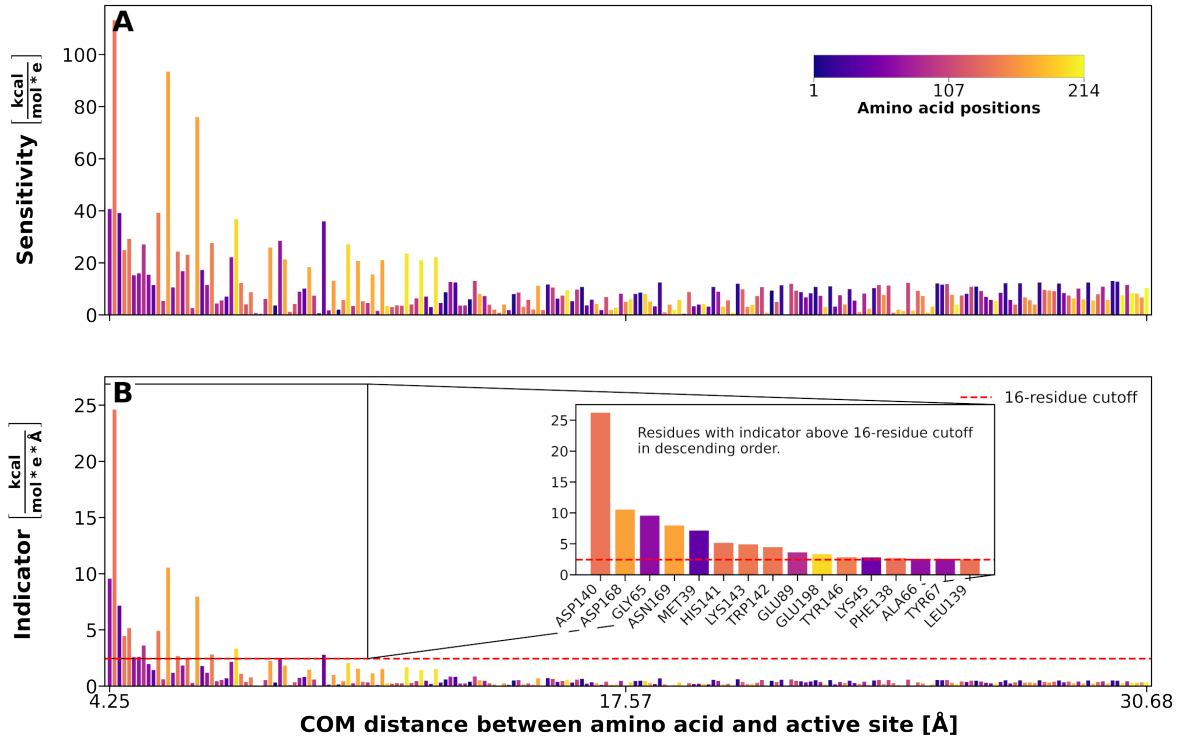

Figure S8: *cryst* PCVA for COMT reactant energy: QM reactant energy sensitivities of all COMT amino acids sorted by ascending residue-active site distance. B: Corresponding QM region indicators for all amino acids. The 16 residues with the highest indicators are represented in the additional plot sorted by descending indicators.

Table S1: Information on the QM regions of increasing size for COMT constructed by PCVA with a variation of −0.5. Residues are ordered according to their position in the indicator ranking. Reproduced from *J. Chem. Theory Comput.* **18**, 2584-2596 (2022), DOI: 10.1021/acs.jctc.1c01093.

Region	Charge	Residues	Link atoms	Included residues
1	+1	0	0	MG SAM CAT
2'	0	3	6	D140 G65 D168 *
3'	-1	7	8	D140 G65 D168 N169 H141 A66 E89 *
4'	-1	13	12	D140 G65 D168 N169 H141 A66 E89 M39 E198 W142 K143 F138 L139 *
5'	0	19	18	D140 G65 D168 N169 H141 A66 E89 M39 E198 W142 K143 F138 L139 Y67 V41 K45 Y146 L64 A167 *
6'	0	22	22	D140 G65 D168 N169 H141 A66 E89 M39 E198 W142 K143 F138 L139 Y67 V41 K45 Y146 L64 A167 C172 V170 N91 *
7'	-1	26	22	D140 G65 D168 N169 H141 A66 E89 M39 E198 W142 K143 F138 L139 Y67 V41 K45 Y146 L64 A167 C172 V170 N91 E63 P173 N40 C94 *
8'	-2	34	24	D140 G65 D168 N169 H141 A66 E89 M39 E198 W142 K143 F138 L139 Y67 V41 K45 Y146 L64 A167 C172 V170 N91 E63 P173 N40 C94 L197 I90 D204 L206 G174 G205 L166 Y70 *
9'	-2	43	26	D140 G65 D168 N169 H141 A66 E89 M39 E198 W142 K143 F138 L139 Y67 V41 K45 Y146 L64 A167 C172 V170 N91 E63 P173 N40 C94 L197 I90 D204 L206 G174 G205 L166 Y70 R145 I88 A175 I171 D144 W37 S118 T87 S195 *

* + MG SAM CAT H_2O

Table S2: Information on the 16-residue QM regions constructed based on different PCVA schemes. Residues are ordered according to their position in the indicator ranking.

Region	Charge	Residues	Link atoms	Included residues
standard	0	16	16	D140 G65 D168 N169 H141 A66 E89 M39 E198 W142 K143 F138 L139 Y67 V41 K45 *
active	0	16	18	D140 K143 D168 N169 E198 G65 V41 H141 K45 N40 W142 A66 G42 E89 Y70 G174 *
no lig	0	16	20	D140 K143 D168 N169 E198 K45 V41 L197 L206 A167 D204 F138 G205 G42 M39 S195 *
vac	0	16	18	D140 D168 G65 N169 M39 H141 E89 K143 E198 W142 A66 K45 V41 Y146 Y67 L139 *
vac fix	0	16	16	D140 D168 G65 N169 H141 E89 A66 M39 E198 V41 K143 Y67 F138 K45 L139 W142 *
cryst	0	16	16	D140 D168 G65 N169 M39 H141 K143 W142 E89 E198 Y146 K45 F138 A66 Y67 L139 *
cryst fix	0	18	22	D140 D168 G65 N169 M39 H141 K143 W142 A66 E89 F138 E198 V41 K45 Y146 L139 *

* + MG SAM CAT H_2O

Table S3: Comparison of residues included in a 16-residue QM region for different simplified PCVA schemes. Numbers indicate the rank of the amino acid according to the value assigned by the corresponding indicator. The highest-ranked 16 residues are marked in red. Active site residues already included in the minimal QM region for PCVA are indicated by an 0.

Residue	standard	nolig	active	vac fix	vac	cryst fix	cryst
ILE8	155	72	65	62	47	44	52
MET39	8	10	32	8	5	5	5
ASN40	25	13	5	21	22	22	31
VAL41	15	2	2	10	13	13	24
GLY42	53	9	8	99	129	140	206
LYS45	16	1	4	14	12	14	12
LEU62	52	94	106	174	201	177	189
GLU63	23	53	45	24	35	24	33
GLY65	2	57	1	3	3	3	3
ALA66	6	71	7	7	11	9	14
TYR67	14	36	81	12	15	30	15
TYR70	34	22	11	20	17	18	17
VAL73	138	68	70	48	45	41	38
ARG74	186	37	37	113	93	105	85
MET75	79	44	75	208	207	212	195
GLU89	7	191	9	6	7	10	9
MET101	113	91	90	44	42	40	37
ILE122	87	137	84	187	194	197	207
MET136	81	60	131	214	214	199	208
VAL137	48	61	165	145	144	92	108
PHE138	12	7	50	13	18	11	13
LEU139	13	16	114	15	16	16	16
ASP140	1	0	0	1	1	1	1
HIS141	5	25	3	5	6	6	6
TRP142	10	47	6	16	10	8	8
LYS143	11	0	0	11	8	7	7
TYR146	17	63	15	18	14	15	11

THR150	71	164	77	202	202	194	190
THR163	134	140	193	160	169	167	175
LEU165	56	62	207	178	184	165	156
LEU166	33	14	38	33	40	34	36
ALA167	19	5	25	19	19	17	18
ASP168	3	0	0	2	2	2	2
ASN169	4	0	0	4	4	4	4
VAL170	21	24	42	23	29	28	27
PRO173	24	43	10	27	21	21	23
PHE178	45	168	56	92	75	86	68
SER195	43	11	21	47	43	67	92
LEU197	27	3	13	25	20	20	20
GLU198	9	0	0	9	9	12	10
ASP204	29	6	19	29	26	25	28
GLY205	32	8	33	31	30	29	30
LEU206	30	4	20	26	24	23	25
GLU207	47	26	96	64	51	57	57
TYR211	192	182	174	144	152	152	164

S3 PCVA for Triosephosphate Isomerase (TIM)

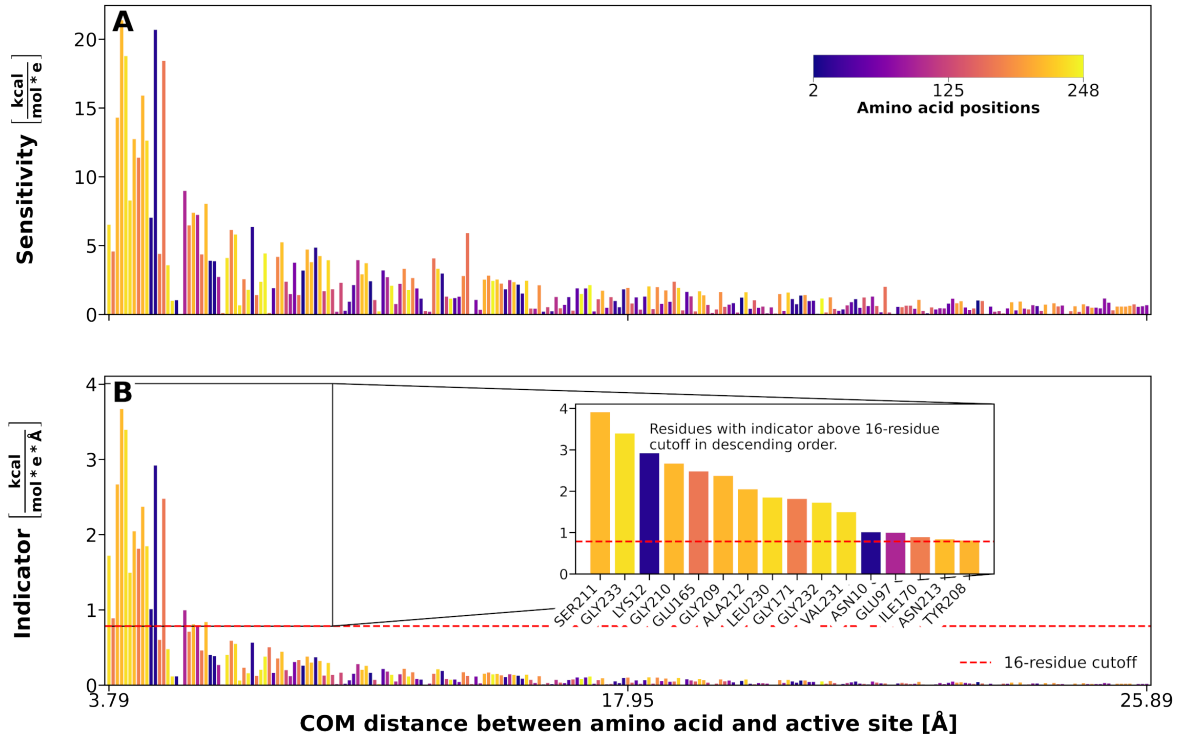

Figure S9: *standard* PCVA for TIM reactant energy. A: QM reactant energy sensitivities of all TIM amino acids sorted by ascending residue-active site distance. B: Corresponding QM region indicators for all amino acids. The 16 residues with the highest indicators are represented in the additional plot sorted by descending indicators.

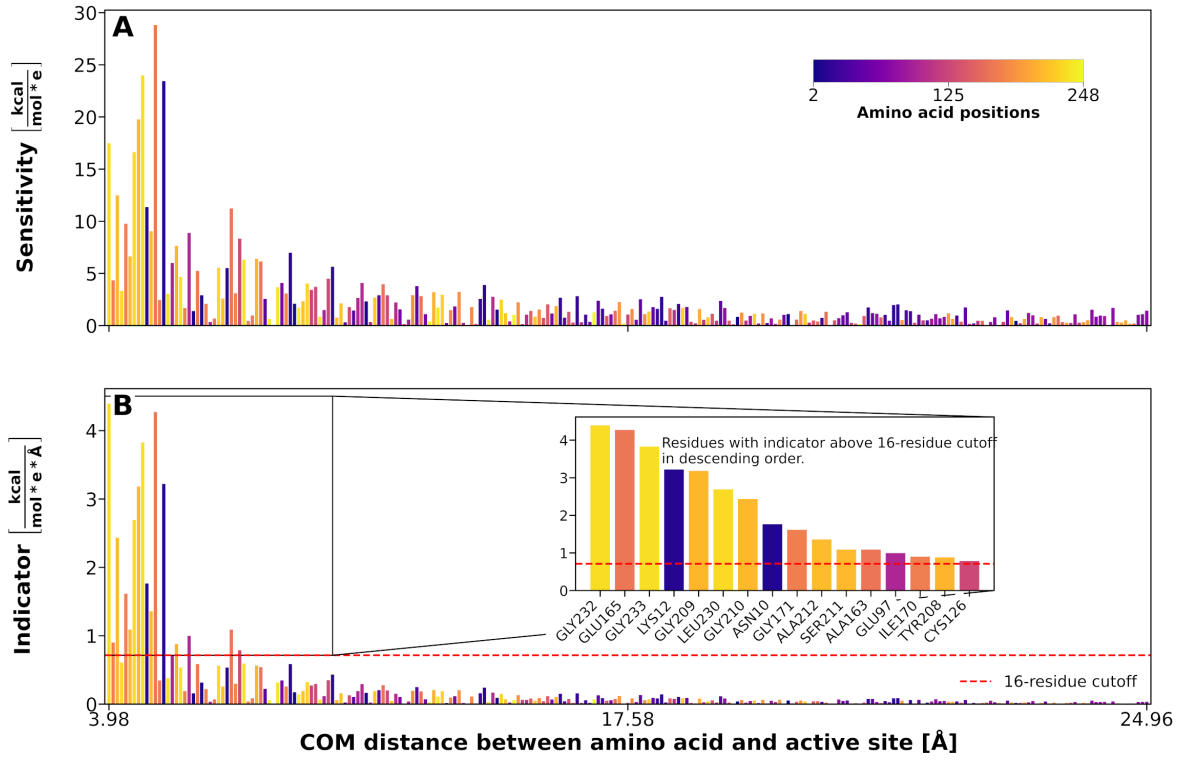

Figure S10: *cryst* PCVA for TIM reactant energy. A: QM reactant energy sensitivities of all TIM amino acids sorted by ascending residue-active site distance. B: Corresponding QM region indicators for all amino acids. The 16 residues with the highest indicators are represented in the additional plot sorted by descending indicators.

Table S4: Information on the TIM distance-based QM regions constructed by radially increasing the QM region size

Region	Charge	Residues	Link atoms	Included residues
1	0	0	0	DHAP/PGH
2	0	2	4	G232 I170 *
3	0	4	6	G232 I170 G210 V231 *
4	-1	12	10	G232 I170 G210 V231 L230 N10 G171 G209 G233 S211 E165 A212 *
5	0	14	12	G232 I170 G210 V231 L230 N10 G171 G209 G233 S211 E165 A212 K12 A169 *
6	-1	20	16	G232 I170 G210 V231 L230 N10 G171 G209 G233 S211 E165 A212 K12 A169 A234 H95 S235 E97 G9 Y208 *
7	-1	27	14	G232 I170 G210 V231 L230 N10 G171 G209 G233 S211 E165 A212 K12 A169 A234 H95 S235 E97 G9 Y208 T172 P166 G8 S96 V167 G173 F229 *
8	-1	33	18	G232 I170 G210 V231 L230 N10 G171 G209 G233 S211 E165 A212 K12 A169 A234 H95 S235 E97 G9 Y208 T172 P166 G8 S96 V167 G173 F229 F11 A163 N213 C126 L236 A175 *
9	0	44	22	G232 I170 G210 V231 L230 N10 G171 G209 G233 S211 E165 A212 K12 A169 A234 H95 S235 E97 G9 Y208 T172 P166 G8 S96 V167 G173 F229 F11 A163 N213 C126 L236 A175 C41 L207 I92 Y164 W168 L174 Q64 F240 K237 N14 V7 *

* + DHAP in reactants or PGH in products, respectively

Table S5: Information on the QM regions of increasing size for TIM constructed by PCVA including atom-economical QM regions for different PCVA schemes and CDA. Residues ordered according to position in indicator ranking. Only 12 residues could be extracted from Bash *et al.* 1991 for CDA,; they are ordered numerically.

Region	Charge	Residues	Link atoms	Included residues
1	0	0	0	DHAP/PGH
2	0	2	4	S211 G233 *
3	1	4	6	S211 G233 K12 G210 *
4	0	12	12	S211 G233 K12 G210 E165 G209 A212 L230 G171 G232 V231 N10 *
5	-1	14	14	S211 G233 K12 G210 E165 G209 A212 L230 G171 G232 V231 N10 E97 I170 *
6	-1	20	20	S211 G233 K12 G210 E165 G209 A212 L230 G171 G232 V231 N10 E97 I170 N213 Y208 H95 G173 A169 A175 *
7	-1	27	28	S211 G233 K12 G210 E165 G209 A212 L230 G171 G232 V231 N10 E97 I170 N213 Y208 H95 G173 A169 A175 N14 F229 A163 A234 P166 C216 L236 *
8	-1	33	28	S211 G233 K12 G210 E165 G209 A212 L230 G171 G232 V231 N10 E97 I170 N213 Y208 H95 G173 A169 A175 N14 F229 A163 A234 P166 C216 L236 G8 F11 I243 L207 L13 A176 *
9	-1	44	30	S211 G233 K12 G210 E165 G209 A212 L230 G171 G232 V231 N10 E97 I170 N213 Y208 H95 G173 A169 A175 N14 F229 A163 A234 P166 C216 L236 G8 F11 I243 L207 L13 A176 Y164 A217 Q64 G214 G228 Y101 S96 V7 F220 L174 A181 *
PCVA	-1	16	14	S211 G233 K12 G210 E165 G209 A212 L230 G171 G232 V231 N10 E97 I170 N213 Y208 *
cryst	-1	16	20	G232 E165 G233 K12 G209 L230 G210 N10 G171 A212 S211 A163 E97 I170 Y208 C126 *
CDA	0	12	20	N10 K12 E77 H95 E97 R98 R99 E104 K112 E129 E165 K237 *

* + DHAP in reactants or PGH in products, respectively

Table S6: Comparison of residues included in a 16-residue (and 12-residue) QM region for standard PCVA, crystal structure PCVA and CDA. Numbers indicate the rank of the amino acid according to the value assigned by the corresponding inclusion scheme. The highest-ranked 16 residues are marked in red. For CDA, the 12 highest-ranked residues are indicated by an x.

Residue	standard	cryst	CDA
ASN10	12	8	x
LYS12	3	4	x
HIS95	17	17	x
GLU97	13	13	x
ARG98	51	52	x
ARG99	127	184	x
GLU104	107	118	x
LYS112	148	177	x
CYS126	49	16	
GLU129	106	71	x
ALA163	23	12	
GLU165	5	2	x
ILE170	14	14	
GLY171	9	9	
TYR208	16	15	
GLY209	6	5	
GLY210	4	7	
SER211	1	11	
ALA212	7	10	
ASN213	15	41	
LEU230	8	6	
VAL231	11	18	
GLY232	10	1	
GLY233	2	3	
LYS237	117	33	x

S4 PCVA for DNA

Table S7: Information on the DNA QM regions constructed by distance and PCVA according to different schemes. Per region the two bases which are added to the QM region are indicated with their abbreviation and position in the system. The number of link atoms corresponds to the number of bases added and the charge of the QM region stays constant at 0.

Scheme	Bases added per QM region									
	1	2	3	4	5	6	7	8	9	10
distance	G7, C14	A6, T15	A8, T13	T5, A16	G9, C12	C4, G17	G10, C11	T3, A18	C2, G19	C1, G20
PCVA QM single	G7, C14	A6, A8	C12, G9	C11, A16	T5, G10	T15, G17	T13, A18	G19, T3	G20, C2	C1, C4
PCVA QM pairs	G7, C14	A6, T15	G9, C12	A8, T13	T5, A16	G10, C11	T3, A18	C4, G17	C2, G19	C1, G20
PCVA RE single	G7, C14	T15, A6	T13, T5	G10, G9	C12, A8	A18, G17	G19, C4	G20, C11	T3, A16	C1, C2
PCVA RE pairs	G7, C14	A6, T15	A8, T13	G9, C12	T5, A16	G10, C11	C4, G17	T3, A18	C2, G19	C1, G20