

Electronic Supporting information (SI)

Mechanism of the highly effective peptide bond hydrolysis by MOF-808 catalyst under biologically relevant conditions

Dragan Conic,^a Kristine Pierloot,^a Tatjana N. Parac-Vogt, *^b Jeremy N. Harvey

^a Division of Quantum Chemistry and Physical Chemistry, Department of Chemistry, KU Leuven, Celestijnenlaan 200F, 3001 Leuven, Belgium

^b Laboratory of Bioinorganic Chemistry, Department of Chemistry, KU Leuven, Celestijnenlaan 200F, 3001 Leuven, Belgium

Table of Contents

S1. An accuracy test for reaction models with conserved (con) and non-conserved (non_con) number of charged particles.....	S2
S2. Analysis of the spurious imaginary frequencies.....	S2-S5
S3. MOF-808 ligand decoration under the physiological condition.....	S6-S7
S4. Experimental details for the observed pH drop.....	S7
S5. Binding of the GlyGly dipeptide to the MOF-808.....	S8
S6. Reaction mechanism of the MOF-808 catalyzed hydrolysis of the GlyGly dipeptide bond.....	S9

S1. An accuracy test for the reaction models with conserved (con) and non-conserved (non_con) number of charged particles

Free energies of dissociation for small reactants used throughout the study, namely, formic acid and water molecule, were calculated using both conserved (con) and non-conserved (non-con) reaction models, in each case with the indicated degree of microsolvation. Results were compared with the corresponding experimental values to test the accuracy of both models. The level of theory is the same as in the main text, i.e. B3LYP-D3BJ/def2-TZVP(SMD)//BP86-D3BJ/def2-SVP(SMD).

Table S1. Error estimation for the con and the non-con reaction types.

		ΔG^0 (kcal/mol)		
		Calc.	Expt.	Err.
1	$\text{HCOOH} + 6 \text{H}_2\text{O} \rightarrow (\text{HCOO}^-) \cdot (\text{H}_2\text{O})_2 + (\text{H}_3\text{O}^+) \cdot (\text{H}_2\text{O})_3$	12.7	5.1	7.6
2	$6 \text{H}_2\text{O} \rightarrow (\text{H}_3\text{O}^+) \cdot (\text{H}_2\text{O})_3 + (\text{OH}^-) \cdot (\text{H}_2\text{O})$	31.7	19.1	12.6
-1 + 2	$(\text{HCOO}^-) \cdot (\text{H}_2\text{O})_2 \rightarrow \text{HCOOH} + (\text{OH}^-) \cdot (\text{H}_2\text{O})$	19.0	14.0	5.0

As shown in Table S1, reactions with non-conserved number of charged species on both reactant and product side give bigger errors comparing to the reaction model where the number of charged species is conserved. The errors are even more pronounced when charged particles are smaller in size, such as naked or insufficiently solvated OH^- and H^+ , for which continuum models fail to predict solvation accurately.¹

S2. Analysis of the spurious imaginary frequencies

Frequency analysis done with Orca 4.1.0 software^{2,3} delivered a few (usually artifactual) imaginary frequencies in six types of cases. As mentioned in the main text, similar observations have been made elsewhere⁴ and seem to indicate some minor numerical instabilities in the continuum module during frequency runs. After testing several computational protocols (e.g. computation of the vibrational frequencies at the same level of theory but with a different code, optimization with a different code, change in level of theory), we concluded that the best combination of accuracy and efficiency was obtained by using the protocol described in the main text computational details, with corrective action taken for individual species displaying spurious imaginary frequencies. This corrective action is described here.

First, in the frequency calculations involving MOF species, in all cases (except a few cases discussed below), exactly six imaginary soft modes ($< \sim 20 \text{ cm}^{-1}$) occurred, all of which corresponded to torsional motions of the atoms in the phenyl groups within the benzoate ligands representing the BTC linkers of the MOF. Given the fixed atom positions of the C_6H_5 groups in benzoate linkers during the geometry optimizations, these imaginary modes were ascribed to be a consequence of the constrained calculations and are thereby not artifacts. Since these frequencies cancel out in calculation of the relative free energies, they were raised to 100 cm^{-1} for the thermal contributions computations.

Second, two imaginary frequencies occurred for the Orca 4.1.0 optimized structure of the zwitterionic form of the GlyGly dipeptide and one for the Gly amino acid, both related to the rotation of the terminal NH_3^+ ammonium group. This issue is most likely related to the use of the CPCM solvation module, with similar artefacts having been noted by other users and discussed on the Orca portal.⁴ Our comparative analysis of the GlyGly dipeptide molecule with the corresponding Gaussian 16⁵ protocol converged to almost identical structure to the one obtained with the Orca software, but without imaginary frequencies (see tables S2 and S3). Additionally, a previous study did not report imaginary frequencies for an almost identical geometry of the same molecule, in line with our Gaussian 16 results.⁶ Since Gibbs free energy contributions of the Orca and Gaussian calculations did not differ significantly for the GlyGly dipeptide (by about only 1 kcal/mol), Orca frequencies were used for assessing the thermal contribution for both the GlyGly dipeptide and the Gly amino acid. The same type of spurious imaginary frequencies occurred with positively charged MOF species with GlyGly bound in its zwitterionic form in cases where the ammonium group was free to rotate or was only loosely hydrogen bonded. However, for these structures, it was obvious that whatever the precise protocol used to compute the free energy for these molecules, they would end up being characterized by substantially higher free energies compared to the free energies of the corresponding neutral binding forms. Thus, the precise treatment of the artifactual modes was considered less important, and the reported free energies for these species were based on statistical mechanics computations in which these imaginary modes were simply raised to 100 cm^{-1} for the calculations of the thermal contributions.

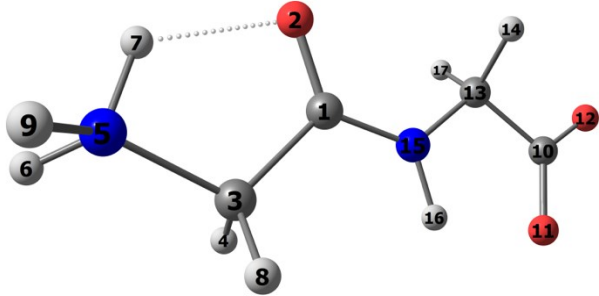
Third, a similar observation was made in the case of the deprotonated GlyGly dipeptide (anionic form with a deprotonated -NH_2 N-terminus, microsolvated at the carboxylate end with one explicit water molecule) in case of Orca optimization and frequency calculation, where one imaginary frequency associated with a torsional mode near the N-terminus occurred. In contrast, Gaussian 16 frequencies showed no imaginary frequencies for almost the same structure. Here, this apparently spurious imaginary frequency, as well as the rest of the frequencies below 100 cm^{-1} , were raised to this threshold and used for evaluation of the thermal contributions.

Fourth, the frequency calculation for the hydronium ion including explicit solvation by three water molecules, $(\text{H}_3\text{O}^+)(\text{H}_2\text{O})_3$, was characterized by four imaginary frequencies, three of which related to soft rotational modes of each water molecule hydrogen bonded to the hydronium ion, and one related to a collective soft motion of the whole cluster. Similar to the GlyGly case, Gaussian 16 calculation of the same species converged to almost identical structure, showing no imaginary frequencies. In this case, a procedure of using Gaussian 16 frequencies for calculation of the thermal contributions was applied. A very similar type of imaginary mode for a water molecule occurred in the case of $\text{Zr}_6(\mu_3\text{-O})_5(\mu_3\text{-OH})_3(\text{BzO})_6(\text{HCOO})_5(\text{H}_2\text{O})_2$, but for this case, it was obvious that whatever procedure was used to compute the free energy for this species, it would lie considerably higher in free energy than competing possibilities, and accordingly, the free energy was computed using the simplest protocol available, i.e. the mode was raised to 100 cm^{-1} for the thermal contribution calculations.

The fifth case that needs to be noted here is the exception to the rule stated in point (1) above. We noted that all MOF structures except one returned six artefactual imaginary frequencies associated with the benzoate groups, which are clearly due to the use of constrained atom positions for these groups during optimization. However, for the chelating form of the all-formate case, only four such imaginary modes were obtained. Inspection of the calculated potential energy and free energy contributions for this species made it clear that whatever protocol was used for free energy calculation, it would lie significantly higher in free energy than competing forms. Accordingly, the simple approach whereby all the frequencies below 100 cm^{-1} were raised to this threshold for the thermochemical analysis.

The sixth case is that of the isomeric structure of TS_1 in which the terminal amine group acts as the general base. This is the structure shown in Figure 4 in the main text. The frequency analysis returned twenty-four imaginary frequencies for this structure, some of which are very large (they fall in the range of $i20 - i2305\text{ cm}^{-1}$) and do not correspond to any of the categories 1-5 above. These effects again appear to be due to numerical instabilities in the continuum model, as mentioned above.⁴ In this case, vibrational frequencies were recomputed at the same optimized structure and with the same DFT functional and basis set, but in this case using the Gaussian 16 code. Many of the imaginary frequencies were absent in this alternative computation, with only one large imaginary frequency mode, associated with the transition state motion and thirteen smaller imaginary frequencies distributed roughly uniformly in the range $i10 - i120\text{ cm}^{-1}$, mainly due to factor 1) mentioned above, plus the fact that the frequencies have been obtained with Gaussian at a stationary point optimized with Orca (albeit with almost identical computational settings). Using the Gaussian 16-derived frequencies to compute thermal properties with our otherwise standard protocol, we arrived at the estimated relative free energy for this TS isomer shown in Fig. 4. This TS still lies higher in free energy than the main TS_1 isomer in which COO^- plays the role of general base.

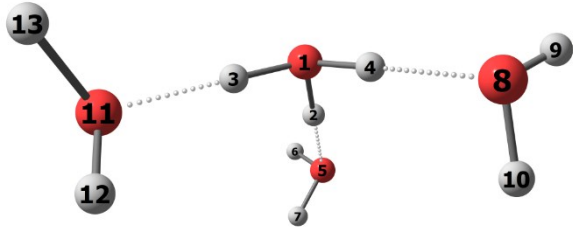
Table S2. Optimized structural parameters for GlyGly



Distance designation	R_{G16} (Å)	R_{Orca} (Å)	$\Delta R_{\text{Orca-G16}}$ (Å)
R(1,2)	1.254	1.253	0.000
R(1,3)	1.535	1.535	0.000
R(1,15)	1.334	1.332	-0.001
R(3,4)	1.106	1.106	-0.001
R(3,5)	1.487	1.486	-0.001
R(3,8)	1.104	1.104	-0.001
R(5,6)	1.034	1.032	-0.001
R(5,7)	1.058	1.059	0.001
R(5,9)	1.035	1.034	-0.002
R(10,11)	1.272	1.271	-0.001
R(10,12)	1.263	1.262	-0.001
R(10,13)	1.554	1.553	-0.001
R(13,14)	1.112	1.111	-0.001
R(13,15)	1.451	1.450	-0.001
R(13,17)	1.112	1.111	-0.001
R(15,16)	1.036	1.036	0.000
ΔG contribution to the electronic energy from the Orca and G16 output respectively (kcal/mol)		60.8	59.8

Table S3. Normal modes of GlyGly from optimization and frequency calculations in Orca and G16

Orca frequencies enumeration	Orca frequencies (cm ⁻¹)	Corresponding G16 frequencies (cm ⁻¹)
1	-195	265
2	-75	61
3	32	26
4	90	104
5	129	78
6	159	172
7	243	240
8	302	306
9	367	366
10	512	527
11	543	544
12	555	563
13	648	650
14	680	696
15	702	710
16	845	879
17	867	867
18	927	932
19	953	960
20	1012	1021
21	1029	1058
22	1039	1052
23	1107	1111
24	1145	1145
25	1161	1174
26	1234	1268
27	1248	1255
28	1282	1307
29	1323	1348
30	1352	1338
31	1372	1394
32	1387	1396
33	1474	1503
34	1497	1493
35	1511	1535
36	1650	1616
37	1687	1675
38	2969	2972
39	3012	3055
40	3015	3018
41	3043	3041
42	3112	3118
43	3316	3319
44	3387	3378
45	3448	3441

Table S4. Optimized structural parameters for $(\text{H}_3\text{O}^+)(\text{H}_2\text{O})_3$


Distance designation	R_i^{G16} (Å)	R_i^{Orca} (Å)	$\Delta R_i^{\text{Orca}-\text{G16}}$ (Å)
R(1,2)	1.046	1.048	0.002
R(1,3)	1.046	1.047	0.001
R(1,4)	1.046	1.047	0.001
R(2,5)	1.510	1.498	-0.012
R(3,11)	1.510	1.499	-0.011
R(4,8)	1.510	1.500	-0.009
R(5,6)	0.981	0.980	-0.001
R(5,7)	0.981	0.980	-0.001
R(8,9)	0.981	0.980	-0.001
R(8,10)	0.981	0.980	-0.001
R(11,12)	0.980	0.980	-0.001
R(11,13)	0.981	0.980	-0.001
ΔG contribution to the electronic energy from the Orca and G16 output respectively (kcal/mol)			44.2 43.9

Table S5. Normal modes of $(\text{H}_3\text{O}^+)(\text{H}_2\text{O})_3$ from optimization and frequency calculations in Orca and G16

Orca frequencies enumeration	Orca frequencies (cm^{-1})	Corresponding G16 frequencies (cm^{-1})
1	-296	166
2	-291	112
3	-264	134
4	-14	62
5	35	65
6	39	81
7	305	326
8	313	332
9	318	327
10	333	406
11	337	395
12	348	371
13	398	447
14	401	467
15	435	487
16	824	864
17	976	1001
18	984	1019
19	1349	1404
20	1529	1570
21	1532	1564
22	1541	1559
23	1637	1666
24	1640	1661
25	2360	2392
26	2370	2387
27	2632	2659
28	3647	3641
29	3647	3637
30	3647	3640
31	3718	3713
32	3718	3709
33	3718	3711

S3. MOF-808 ligand decoration under the physiological conditions

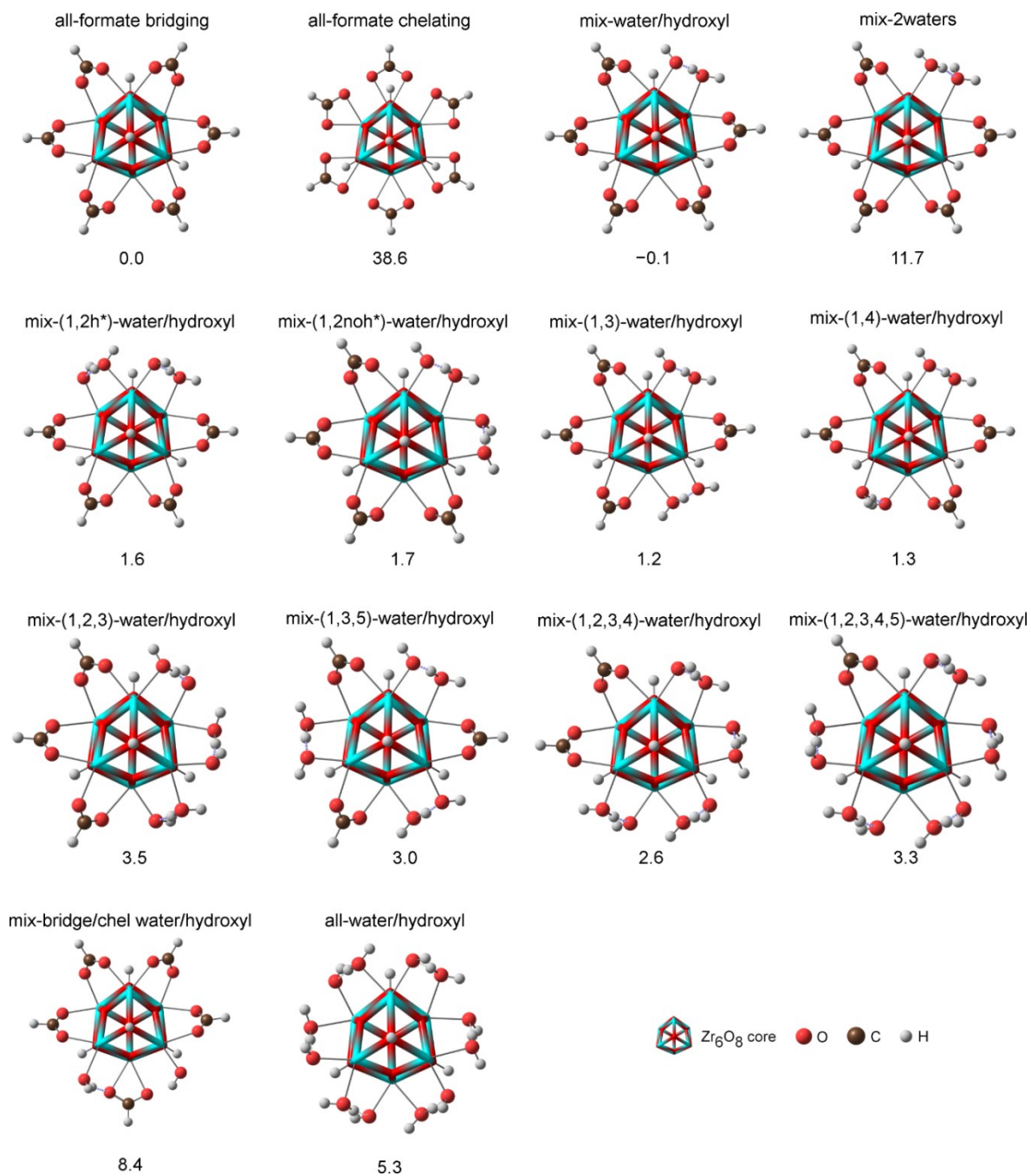


Figure S1. MOF-808 ligand decoration types and their relative free energies to the all-formate bridging form (kcal/mol)

Table S6. Free energies of MOF-808 species relative to the free energy of the all-formate bridging form (kcal/mol)

Group	Chemical formula	Designation	ΔG_{con}^0	$\Delta G_{\text{con}}'$	$\Delta G_{\text{non-con}}^0$
I	$\text{Zr}_6(\mu_3\text{-O})_4(\mu_3\text{-OH})_4(\text{BzO})_6(\text{HCOO})_{6\text{b}}$	all-formate bridging	0	0	0
	$\text{Zr}_6(\mu_3\text{-O})_4(\mu_3\text{-OH})_4(\text{BzO})_6(\text{HCOO})_{6\text{c}}$	all-formate chelating	38.6	38.6	38.6
	$\text{Zr}_6(\mu_3\text{-O})_4(\mu_3\text{-OH})_4(\text{BzO})_6(\text{HCOO})_{5\text{b}}(\text{OH})(\text{H}_2\text{O})$	mix-water/hydroxyl	-0.1	4.2	12.5
	$\text{Zr}_6(\mu_3\text{-O})_4(\mu_3\text{-OH})_4(\text{BzO})_6(\text{HCOO})_{4\text{b}}(\text{OH})_2(\text{H}_2\text{O})_2$	mix-(1,2h)- water/hydroxyl	1.6	10.3	26.8
	$\text{Zr}_6(\mu_3\text{-O})_4(\mu_3\text{-OH})_4(\text{BzO})_6(\text{HCOO})_{4\text{b}}(\text{OH})_2(\text{H}_2\text{O})_2$	mix-(1,2noh)- water/hydroxyl	1.7	10.3	26.8
II	$\text{Zr}_6(\mu_3\text{-O})_4(\mu_3\text{-OH})_4(\text{BzO})_6(\text{HCOO})_{4\text{b}}(\text{OH})_2(\text{H}_2\text{O})_2$	mix-(1,3)- water/hydroxyl	1.2	9.8	26.3
	$\text{Zr}_6(\mu_3\text{-O})_4(\mu_3\text{-OH})_4(\text{BzO})_6(\text{HCOO})_{4\text{b}}(\text{OH})_2(\text{H}_2\text{O})_2$	mix-(1,4)- water/hydroxyl	1.3	9.9	26.4
	$\text{Zr}_6(\mu_3\text{-O})_4(\mu_3\text{-OH})_4(\text{BzO})_6(\text{HCOO})_{3\text{b}}(\text{OH})_3(\text{H}_2\text{O})_3$	mix -(1,2,3)- water/hydroxyl	3.5	16.4	41.2
	$\text{Zr}_6(\mu_3\text{-O})_4(\mu_3\text{-OH})_4(\text{BzO})_6(\text{HCOO})_{3\text{b}}(\text{OH})_3(\text{H}_2\text{O})_3$	mix -(1,3,5)- water/hydroxyl	3.0	15.9	40.7
	$\text{Zr}_6(\mu_3\text{-O})_4(\mu_3\text{-OH})_4(\text{BzO})_6(\text{HCOO})_{2\text{b}}(\text{OH})_4(\text{H}_2\text{O})_4$	mix -(1,2,3,4)- water/hydroxyl	2.6	19.8	52.8
III	$\text{Zr}_6(\mu_3\text{-O})_4(\mu_3\text{-OH})_4(\text{BzO})_6(\text{HCOO})_{2\text{b}}(\text{OH})_4(\text{H}_2\text{O})_4$	mix -(1,2,3,4,5)- water/hydroxyl	3.3	24.8	66.1
	$\text{Zr}_6(\mu_3\text{-O})_4(\mu_3\text{-OH})_4(\text{BzO})_6(\text{OH})_6(\text{H}_2\text{O})_6$	all-water/hydroxyl	5.3	31.1	80.6
IV	$\text{Zr}_6(\mu_3\text{-O})_5(\mu_3\text{-OH})_3(\text{BzO})_6(\text{HCOO})_{5\text{b}}(\text{H}_2\text{O})_2$	mix-2waters	11.7	16.0	24.3
V	$\text{Zr}_6(\mu_3\text{-O})_4(\mu_3\text{-OH})_4(\text{BzO})_6(\text{HCOO})_{4\text{b}}(\text{OH})(\text{HCOO})_{\text{c}}(\text{H}_2\text{O})$	mix-bridge/chel water/hydroxyl	8.4	12.7	17.8

* Subscript "b" designates bridging binding mode

** Subscript "c" designates chelating binding mode

S3. Experimental details for the MOF preparation and its effect on pH of liquid phase

MOF-808 was prepared using the same procedure as previously reported.⁷ Briefly, 3.75 mmol (0.786g) of 1,3,5-benzenetricarboxylic acid and 3.75 mmol (1.209g) of $\text{ZrOCl}_2 \cdot 8\text{H}_2\text{O}$ were mixed in DMF (150 mL) and formic acid (150 mL). The mixture was placed in a closed 1 L Schott DURAN bottle and heated at 130°C for 48 h. The white powder obtained from this synthesis was washed with DMF 2 times a day for 3 days, followed by ethanol washing 2 times a day for 3 days. The washed as-synthesized sample was collected by centrifugation and heated at 150°C for 20h to reach an activated MOF-808 sample.

2.0 μmol (3.57 mg) of the activated MOF-808 sample was mixed with 1 mL D_2O , resulting in the decrease of pH from 7.03 to 3.84.

S4. Binding of the GlyGly dipeptide to the MOF-808

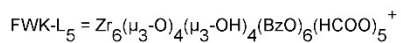
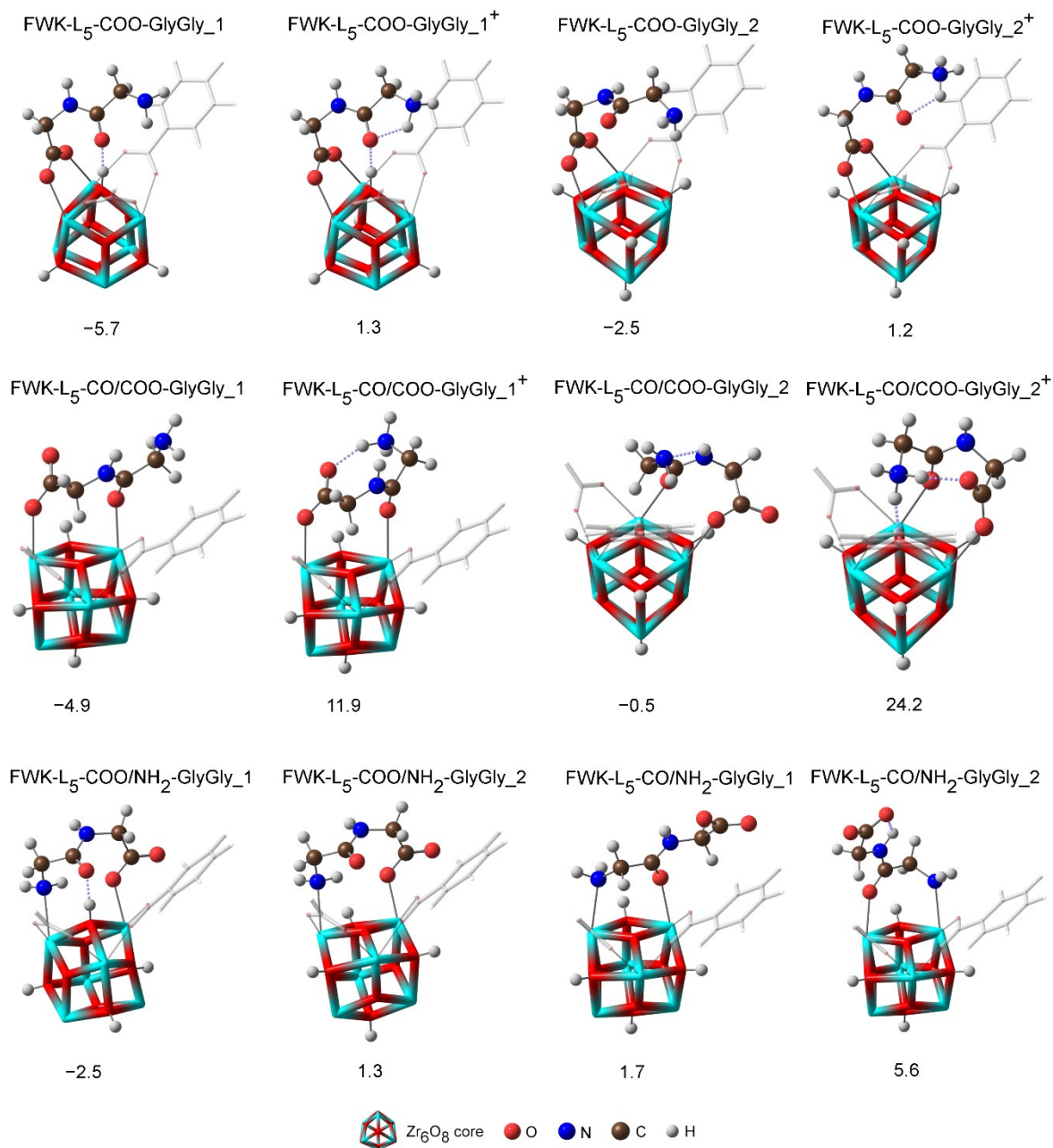


Figure S2. GlyGly to MOF-808 binding modes with the corresponding relative free energies of binding (kcal/mol)

S5. Reaction mechanism of the MOF-808 catalyzed hydrolysis of the GlyGly dipeptide bond

Table S7. Free energies of reaction pathways where terminal COO⁻ and terminal NH₂ group act as a general base, calculated relative to the most stable binding complex (MSBC) (kcal/mol)

Chemical formula	COO ⁻ as a general base	NH ₂ as a general base
MSBC + H ₂ O	0	0
Reactant complex	13.9	14.6
Transition state 1	30.6	34.8 (40.9 [§])
Intermediate 1	29.8	29.7
	low barrier proton shuffle	
Intermediate -1	31.3	
Transition state -1	32.3	
Product complex	15.5	
Final products	-9.3	-9.3

[§]Second lowest free energy structure

Notes and references

- 1 J. Ho, M. L. Coote, *Theor. Chem. Acc.*, **2010**, 125, 3–21.
- 2 F. Neese, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **2012**, 2, 73–78.
- 3 F. Neese, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **2018**, 8, e1327. doi: 10.1002/wcms.1327.
- 4 Orcaforum.kofo.mpg.de. **2020**. ORCA Forum - Portal. [online] Available at: <<https://orcaforum.kofo.mpg.de>> [Accessed 26 May 2020].
- 5 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, Gaussian 16 (Revision A.03), Gaussian, Inc., Wallingford CT **2016**.
- 6 H. G. T. Ly, T. Mihaylov, G. Absillis, K. Pierloot, T. N. Parac-Vogt, *Inorg. Chem.*, **2015**, 54, 11477–11492.
- 7 H. G. T. Ly, G. Fu, A. Kondinski, B. Bueken, D. De Vos, T. N. Parac-Vogt, *J. Am. Chem. Soc.*, **2018**, 140, 6325–6335.