SUPPORTING INFORMATION FOR

Influence of extreme thermodynamic conditions and pyrite surfaces on peptide synthesis in aqueous media

Eduard Schreiner, Nisanth N. Nair, and Dominik Marx

Lehrstuhl für Theoretische Chemie, Ruhr–Universität Bochum, 44780 Bochum, Germany

I. GENERAL INFORMATION

This computer simulation study is carried out directly at condensed phase conditions (using supercells that are periodically replicated in all three dimensions, thus allowing treatment of bulk water at pre–selected temperature / pressure conditions and a slab representation of the mineral surface according to details given in Sec. II) in the framework of accelerated reactive molecular dynamics simulations (more precisely *ab initio* Car–Parrinello metadynamics as explained in Secs. III and IV). Several critical tests demonstrating the accuracy and reliability of this approach are presented in Sec. V. In addition, selected activation free energies are compared in Sec. VI to experimental values where available.

II. SYSTEM SETUP

Depending on the thermodynamic conditions, different supercell setups and temperatures were used for the simulations of the reactions that yield the peptide synthesis cycle. For simulations at ambient bulk water (ABW) standard conditions, in which $T = 300\,\mathrm{K}$ and $p \approx 0.1\,\mathrm{MPa}$, an orthorhombic supercell with the dimensions $10.8{\times}10.8{\times}9.5\,\text{Å}$ was used, whereas for simulations of hot–pressurized bulk water (HPW) extreme conditions a cubic box having the dimensions $10.8{\times}10.8{\times}10.8\,\text{Å}$ was employed. At HPW conditions, the temperature was increased to 500 K resulting in an expected pressure of about 20 MPa which is roughly estimated based on using the experimental equation of state [1]; note that the PBE functional in particular has been shown to reproduce quite well volume–dependent properties of ice [2]. At both conditions the supercell contained 36 water molecules corresponding to a density of $\approx 1.00\,\mathrm{g/cm^3}$ at ABW conditions and $\approx 0.85\,\mathrm{g/cm^3}$ at HPW conditions [1].

The simulations mimicking interfacial water at a mineral surface were performed using pyrite, FeS_2, in contact with hot–pressurized water (PIW). The pyrite surface is represented by a slab that is periodically repeated in two dimensions (i.e. along the [100] and [010] directions as indicated in SI Figure 1). It exposes its (001) face to establish an interface with water and consists of 9 atomic layers with 24 Fe and 48 S atoms in total. During all slab calculations performed, its undermost three atomic layers were kept fixed at their optimized bulk positions, whereas all other atoms, in particular those that are the adsorption sites for admolecules, were allowed to move without imposing further constraints. This is expected

to yield a reliable representation of the ideal surface which is known to not undergo easily surface reconstructions [3]. In addition to the ideal surface, we use a defective surface in one step of the cycle, step $\mathbf{C}''''$. A sulfur vacancy is selected because is it known to be an ubiquitous point defect on pyrite (001) which is found to increase the retention time of adsorbed amino acids by many orders of magnitude in comparison to the ideal pyrite–water interface, see Ref. [6] for background information and for details of the respective setup.

These PIW simulations were carried out in an orthorhombic box of dimensions $10.8\times10.8\times18.9\,\text{Å}$ containing 36 water molecules [4–6]. The three–dimensional periodic boundary conditions employed in the simulations establish a lamella of water between the upper and lower faces of the pyrite slab, as can clearly be seen in SI Figure 1. At all these simulation conditions, introduction of a glycine or Leuchs anhydride molecule replaces a water molecule thus maintaining the water density approximately. The PIW system was also thermostated to 500 K, which results in an expected pressure of about 20 MPa according to the experimental equation of state [1].

The starting structures for the simulations were obtained either by using the final structure of the previous reaction or by the following equilibration protocol. A pre–equilibrated water system was added to the solute, which was adsorbed at PIW conditions on the surface, and equilibrated at the target temperature for 1 ps, while the atomic positions of the solute as well as those of the pyrite slab were kept fixed. Each remaining nuclear degree of freedom was thermostated separately at the target temperature in order to ensure energy equipartition in the total system even in the presence of stiff vibrational modes. After releasing all of the nuclei with the exception of the undermost three atomic layers of the pyrite slab, all unconstrained degrees of freedom were additionally equilibrated individually for typically 1 to 3 ps prior to the metadynamics simulations, in which one thermostat for all nuclei was used.

III. *AB INITIO* MOLECULAR DYNAMICS

All calculations were performed within spin–restricted Kohn–Sham density functional theory in its plane wave / pseudopotential formulation [7]. The PBE [8] exchange–correlation functional was chosen, and the core electrons were taken into account using Vanderbilt's ultrasoft pseudopotentials [9] containing additional d–projectors in case of sulfur as well as

scalar relativistic corrections and semicore states for iron. This particular approach has been demonstrated to yield an accurate description of the electronic structure of the system of interest [5, 6].

The *ab initio* molecular dynamics (AIMD) simulations were performed using the Car–Parrinello scheme [10] along with efficient Nosé–Hoover chain thermostats [11] for nuclei and electronic orbitals. A molecular dynamics time step of $\Delta t = 0.144\,\mathrm{fs}$ was used for the integration of the Car–Parrinello equations of motion, the fictitious mass for the orbitals was 700 a.u., and as usual the hydrogen masses were substituted by deuterium masses for technical reasons. All simulations have been performed with the `CPMD` simulation package [7, 12].

IV. *AB INITIO* METADYNAMICS

A. Theoretical Background

The metadynamics technique is used to accelerate chemical reactions and to explore their reaction mechanisms by investigating the topology of free energy hypersurfaces obtained by this method. The essence of this technique [13] is a coarse–grained description of the underlying chemistry by choosing a greatly reduced set of collective coordinates $\{S_\alpha(\mathbf{R})\}$, which are generalized coordinates depending in a very general way on the coordinates $\mathbf{R}$ of the nuclei, in conjunction with a non–Markovian (pseudo–) time evolution as reviewed in Ref. [14]. Here, we use the efficient extended Lagrangian formulation [15] of the metadynamics technique in which a set of auxiliary degrees of freedom $\{s_\alpha\}$ associated to the collective coordinates is introduced.

These auxiliary variables $\mathbf{s}$ are coupled to the collective coordinates through a restraining potential $\sum_{\alpha=1}^{N_\mathrm{S}} k_\alpha[S_\alpha(\mathbf{R}) - s_\alpha]^2$, where k_α is the coupling constant and N_S is the number of collective coordinates and thus the dimensionality of the spanned subspace. Within the space spanned by all collective coordinates, a multivariate Gaussian–type potential $V^\mathrm{MTD}(t, \mathbf{s})$ is incremented slowly on discrete points along the trajectory of the auxiliary variables as time progresses. Here, this potential is defined as

$$V^\mathrm{MTD}(t, \mathbf{s}) = \sum_{t_i<t} H(t_i) \exp\left\{-\frac{[\mathbf{s}(t) - \mathbf{s}(t_i)]^2}{2\left[\mathbf{w}(t_i)\delta s(t_i)\right]^2}\right\}, \tag{4.1}$$

where the sum runs over all the previous metadynamics steps Δt^{MTD}, $2\Delta t^{\mathrm{MTD}}$, ..., t_i, ... with corresponding Gaussian centers $\mathbf{s}(t_i)$ where $\Delta t^{\mathrm{MTD}} \gg \Delta t$. The parameters $H(t_i)$ and $\delta s(t_i)$ determine the height and the width of the Gaussian dropped at a time t_i, respectively. To sample the available collective coordinate space efficiently, $H(t_i)$ and $\delta s(t_i)$ are adapted during the simulation according to the depth and width of the well and the progress of filling. Nonspherical free energy basins are filled efficiently by the usage of anisotropic scaling factors $\mathbf{w} = \{w_\alpha\}$ that are determined as explained in Sec. V B. The metadynamics time step Δt^{MTD} is an integer multiple of molecular dynamics time step Δt and is calculated adaptively according to the diffusion of the collective coordinates as described in Sec. V B.

The accumulated history–dependent potential $V^{\mathrm{MTD}}(t, \mathbf{s})$ slowly fills the underlying free energy surface and drives the system to other minima along the minimum free energy pathway. Due to this feature, the metadynamics technique has the capability to explore surprisingly new structures and reaction pathways (see e.g. Ref. [6]). Moreover, if all proper collective coordinates are included, i.e. if the subspace is sufficient to host the true (high–dimensional) reaction coordinate, it provides an unbiased estimate of the free energy if carefully filled with Gaussians. Utilizing the useful relationship that the negative of the accumulated biasing potential is a quantitative estimate of the (Helmholtz) free energy

$$F(\mathbf{s}) = -\lim_{t\to\infty} V^{\mathrm{MTD}}(t, \mathbf{s}) + \text{constant}, \tag{4.2}$$

in the subspace spanned by $\{s_\alpha\}$, free energy differences and free energy barriers can be obtained directly from metadynamics simulations for the explored reaction mechanism.

Such NVT free energy surfaces were determined separately for each reaction step, **A**, **B**, ..., that contributes to the full synthesis cycle. This procedure is the analog of the standard approach used in static quantum–chemical investigations of chemical reaction sequences in the gas phase (or equivalently in continuum solvents) where (i) total energies of reactants, transition states, and products are computed separately for all individual reaction steps followed by (ii) estimating the relative free energies by invoking the harmonic approximation to compute harmonic frequencies used to correct for finite–temperature and pressure effects. Here, we generalize this well–established approach to dynamical simulations that do include finite–temperature and pressure as well as solvent effects explicitly when computing relative free energies and activation barriers. In addition, the changing number of particles along the cycle due to addition or removal of the small gaseous molecules COS, SH_2, and

CO_2 has been approximately compensated by the removal and addition, respectively, of solvation water molecules according to the protocol outlined in Sec. II in order to keep the number density roughly constant.

B. Collective Coordinates

In the presented work three different types of collective coordinates $\{S_\alpha(\mathbf{R})\}$ are used for metadynamics simulations. The simplest one used is just the distance

$$d(\mathrm{A}-\mathrm{B}) = |\mathbf{R}_\mathrm{A} - \mathbf{R}_\mathrm{B}| \tag{4.3}$$

between two chosen atoms A and B.

The second type is the coordination number [16] between an atom A with respect to a set of other atoms B, $c(\mathrm{A}-\mathrm{B})$, consisting of the same species, which is defined as

$$c(\mathrm{A}-\mathrm{B}) = \sum_{I\in\mathrm{B}} \frac{1-(R_{\mathrm{A}I}/R^0_{\mathrm{AB}})^6}{1-(R_{\mathrm{A}I}/R^0_{\mathrm{AB}})^{12}} \ . \tag{4.4}$$

Here $R_{\mathrm{A}I}$ is the distance between atom A and an atom I and R^0_{AB} is a fixed cutoff parameter that characterizes the typical bond distance between A and B. For each pair of atoms belonging to A and B, function $c(\mathrm{A}-\mathrm{B})$ is nearly unity when the actual bond distance $R_{\mathrm{A}I} < R^0_{\mathrm{AB}}$, and approaches zero rapidly when $R_{\mathrm{A}I} > R^0_{\mathrm{AB}}$.

The third type of collective coordinate is a total coordination number $c^{\mathrm{tot}}(\mathrm{A}-\mathrm{B})$ between two sets of atoms A and B, defined as

$$c^{\mathrm{tot}}(\mathrm{A}-\mathrm{B}) = \sum_{J\in\mathrm{A}}\sum_{I\in\mathrm{B}} \frac{1-(R_{IJ}/R^0_{\mathrm{AB}})^6}{1-(R_{IJ}/R^0_{\mathrm{AB}})^{12}} \ . \tag{4.5}$$

The cutoff distances R^0_{AB} in equations (4.4) and (4.5) for different pairs of species A and B used in the work are given in SI Table I. The particular sets of collective coordinates utilized in the metadynamics simulations of the individual reaction steps are listed in SI Table II.

V. ASSESSMENT: TESTS AND ACCURACY

A. Electronic Structure: Density Functional and Pseudopotentials

Benchmark calculations presented in Ref. [5] show that for the employed ultrasoft pseudopotentials a plane wave cutoff of 25 Ry is sufficient to obtain convergence in the unit cell

parameter and the critical S–S distance of bulk pyrite. Additionally, the electronic structure calculations [6] on bulk pyrite with the same setup found a bulk band gap of 0.87 eV, which is very close to the experimentally measured value [17] of 0.95 eV. The size of the supercell is chosen in such a way that all of the important environmental effects are included, for instance the number of solvation shells which can form within the size of the supercell, and the integral over the Brillouin zone is close to being converged at the Γ–point. An excellent agreement was observed in the adsorption energies in benchmark calculations [6] of an isolated water molecule on the ideal pyrite (001) surface for the molecular and dissociative adsorption modes using the PBE exchange–correlation functional and Vanderbilt ultrasoft pseudopotentials as in the present work, in comparison to those reported in Ref. [18] using a larger supercell, the BLYP functional and Troullier–Martins normconserving pseudopotentials instead. This furthermore confirms the reliability of the current approach (see also the Supplementary Information to Ref. [6]) including the choice of density functional, pseudopotentials, plane wave cutoff, and supercell size for the present study.

B. Metadynamics: Parameters and Simulation Protocol

Due to the finite size of the Gaussians added during metadynamics sampling, the filled potential can be locally rough, and thus the error in the free energy estimate depends on both the height and the width of the Gaussian. Our previous experience [6] shows that the best way to balance the efficiency and accuracy is to first fill the free energy well by using large Gaussian heights of $\approx 2-4\,k_{\mathrm{B}}T$ while keeping an appropriately narrow Gaussian width. Once a barrier crossing event is observed, the free energy estimate is refined subsequently by restarting the simulation sometime before this event using comparatively smaller Gaussian heights of $\leq 1\,k_{\mathrm{B}}T$. The error of the free energy estimates obtained by this procedure have been shown to amount to about one $k_{\mathrm{B}}T$ (see the reference [77] in Ref. [6]). Since smaller biasing potentials are used during the refinement, the system has more time to relax but the simulation time increases. The latter effect is not necessarily a disadvantage because it helps, together with any abnormal convergence behavior in the barrier height and mechanism during this procedure, to detect if any slow modes are missing in the set of collective coordinates, and thus if the subspace as determined by their very definition can in fact accommodate the true reaction coordinate.

The width dimensionless parameter δs is set to 0.05, which is $\approx 1/4^{\text{th}}$ of the width of the fluctuation of a collective coordinate type with the smallest amplitude oscillation in our simulations. The anisotropic scaling parameters $\{w_\alpha\}$, which have the units of the collective coordinates, are determined in such a way that

$$w_\alpha = \frac{\max\left(|S_\alpha - \langle S_\alpha \rangle|\right)}{\delta s} \quad \forall \alpha \tag{5.1}$$

is satisfied, which is estimated during the preparation runs without adding any Gaussians. The coupling constants k_α allowed the collective coordinates S_α and the associated auxiliary variables s_α to move close to each other. In conjunction with a judicious choice of the mass of the auxiliary variable μ_α, an adiabatic separation of $\{s_\alpha\}$ from the electronic degrees of freedom was maintained throughout all simulations. The values of k_α and μ_α depend on the type of collective coordinate and are listed in SI Table III.

The metadynamics time step Δt^{MTD} is chosen adaptively during the dynamics in such a way that a Gaussian is placed at time t_i once the following condition is fulfilled

$$|s(t) - s(t_i)| = 3/2\,\delta s \ , \tag{5.2}$$

which is necessary to avoid so–called "hill-surfing" problems as discussed in Ref. [19]. Note that steepness in the rough biasing potential may result in heating–up the system, because the auxiliary variables pull on the ionic system and thus accelerate the nuclei as they "roll down a hill". Thus, it is necessary to keep the temperature of the auxiliary variables close to that of the physical system, i.e. the nuclei. Due to the strong nonequilibrium nature of the dynamics of the auxiliary variables $\{s_\alpha\}$, simple velocity scaling is used to restrict the fluctuations of their fictitious kinetic energy corresponding to $\pm\,200\,\text{K}$ with respect to the target temperature, while the nuclear degrees of freedom are controlled by a Nosé–Hoover chain thermostat [11] to establish the canonical ensemble.

SI Figure 2 shows the fluctuations of the collective coordinates $\{S_\alpha(\mathbf{R})\}$ and auxiliary variables $\{s_\alpha\}$ for the distance $d(\text{C}_{\text{COS}} - \text{O}_{\text{COOH}})$ between the carbon in the COS moiety of the isocyanate and one of the carboxylate oxygens, and for the coordination number $c(\text{N}-\text{H})$ of the glycine nitrogen to all non–aliphatic hydrogen atoms. They were monitored during the reaction of thiocarbamate to Leuchs anhydride and during the addition of COS and glycine to thiocarbamate, respectively. It is clear that for the parameters chosen according to SI Table III, the auxiliary variables and collective coordinates move close to each other, as

required for proper sampling. Note that in most of the reactions we have studied, there are some very slow diffusing coordinates which require long simulation times in order to fill the associated spacious free energy basins fully. The atomic distance $d(C_{COS} - O_{COOH})$ shown in SI Figure 2(a) is a representative example, where the coordinate is slow and a large free energy barrier is surmounted (see the depth of the corresponding free energy minimum **3** in SI Figure 3).

SI Figure 3 demonstrates nicely a crucial point of the employed metadynamics technique, namely the efficient exploration of the free energy landscape which can lead to the uncovering of new and unexpected minima and interconnecting pathways, i.e. chemically speaking to the discovery of novel intermediates, products, and reaction mechanisms. In this particular case, minimum C′ corresponds to an isocyanate structure that formed during metadynamics after the ring opening of the Leuchs anhydride, which was not foreseen before we performed this simulation. Subsequent to this, more detailed exploration of this feature of the free energy surface led to the discovery of the efficient indirect route from thiocarbamate via isocyanate to Leuchs anhydride as discussed in the main text.

VI. COMPARISON WITH EXPERIMENTAL DATA

Despite this careful technical approach and despite a massive investment of simulation time and thus of computer resources, issues such as inherent system size limitation due to the supercell approximation, statistical errors caused by metadynamics sampling, systematic errors due to the exchange–correlation density functional, etc. certainly affect the computed free energy minima and barriers. Nevertheless, it can be expected that our data should be reliable *within the chosen setup* thus enabling us to assess the *relative* effect of pressure and temperature along the reaction pathway. In addition, we observe a faithful agreement with respect to available experimental data as well as previous computer simulations of specific aspects as demonstrated in the following.

The relative stabilities of the neutral and zwitterionic species of glycine as well as the activation barriers for their interconversion at *ambient conditions* are consistent with previous *ab initio* data obtained with different techniques [20], as well as with experimental data [21]. In particular, the free energy barrier for the conversion of the zwitterionic into the neutral isomer was calculated to be 53.1 kJ/mol and the experimental value is 59.8 kJ/mol; note

that in the paper all free energies are reported in units of k_BT instead in order to allow for direct comparison of these data at different temperatures. Also the obtained free energy difference between the neutral and zwitterionic forms of 40 kJ/mol is in accord with that of a previous calculation (46.8 kJ/mol) that focused on this particular question [20]; here the experimental value is 30.4 kJ/mol [21]. Note, however, that in contrast to the Helmholtz free energies yielded by our simulations as well as those of Ref. [20] it is the Gibbs free energy that is measured experimentally [21].

Also for the hydrolysis reaction we observe acceptable agreement with experiment. The presented simulation data yield an effective free energy barrier of 140 kJ/mol at ambient conditions, while in experiment [22] a reaction rate of $k = 6.3 \times 10^{-11}\,\mathrm{s}^{-1}$ was obtained for the uncatalyzed hydrolysis of diglycine in sealed quartz tubes at $T = 298\,\mathrm{K}$ and pH 6.8, which corresponds to a free energy of activation of 131 kJ/mol close to our value. Most crucially for the present purpose, also the dependency of the activation energy on thermodynamic conditions is reproduced by our simulations. At $T = 423\,\mathrm{K}$ the measured reaction rate is $k = 9.8 \times 10^{-6}\,\mathrm{s}^{-1}$ [22], i. e. the free energy of activation is about 146 kJ/mol, hence higher than at room temperature which is in line with our finding (173 kJ/mol in HPW vs. 140 kJ/mol in ABW).

---

[1] Lown, D. A.; Thirsk, H. R.Lord Wynne–Jones, *Trans. Faraday Soc.* **1970**, *66*, 51–73, Note: the typographical error in the specific volume of 1.1178 cm^3/g for 200 bar pressure and 225 °C temperature in Table 5 (page 62) is corrected to 1.1778 cm^3/g.

[2] Hamann, D. R. *Phys. Rev. B* **1997**, *55*, R10157–R10160.

[3] Rosso, K. M.; Becker, U; Hochella Jr., M. F. *Am. Mineral.* **1999**, *84*, 1535–1548.

[4] Boehme, C.; Marx, D. *J. Am. Chem. Soc.* **2003**, *125*, 13362–13363.

[5] Pollet, R.; Boehme, C.; Marx, D. *Origins Life Evol. Biospheres* **2006**, *36*, 363–379.

[6] Nair, N. N.; Schreiner, E.; Marx, D. *J. Am. Chem. Soc.* **2006**, *128*, 13815–13826.

[7] Marx, D.; Hutter, J. In *Modern Methods and Algorithms of Quantum Chemistry*; NIC, FZ Jülich, 2000. For downloads see `http://www.theochem-.rub.de/go/cprev.html`.

[8] Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

[9] Vanderbilt, D. *Phys. Rev. B* **1990**, *41*, 7892–7895.

[10] Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.

[11] Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635–2643.

[12] `CPMD` Program Package version 3.11; IBM Corp 1990-2007, MPI für Festkörperforschung Stuttgart 1997-2001. See also `http://www.cpmd.org`.

[13] Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci.* **2002**, *99*, 12562–12566.

[14] Laio, A.; Parrinello, M. In *Computer simulations in condensed matter: From materials to chemical biology*; Berlin Heidelberg, 2006.

[15] Iannuzzi, M.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *90*, 238302–1–4.

[16] Sprik, M. *Chem. Phys.* **2000**, *258*, 139–150.

[17] Eggleston, C. M.; Ehrhardt, J.-J.; Stumm, W. *Am. Mineral.* **1996**, *81*, 1036–1056.

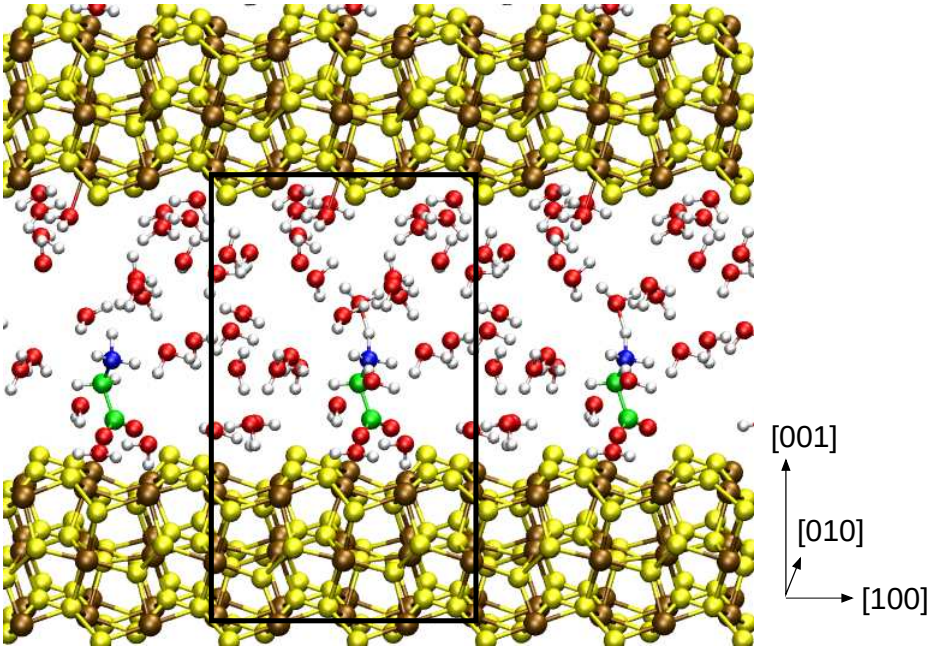
[18] Stirling, A.; Bernasconi, M.; Parrinello, M. *J. Chem. Phys.* **2003**, *118*, 8917–8926.

[19] Ensing, B.; Laio, A.; Parrinello, M.; Klein, M. L. *J. Phys. Chem. B* **2005**, *109*, 6676–6687.

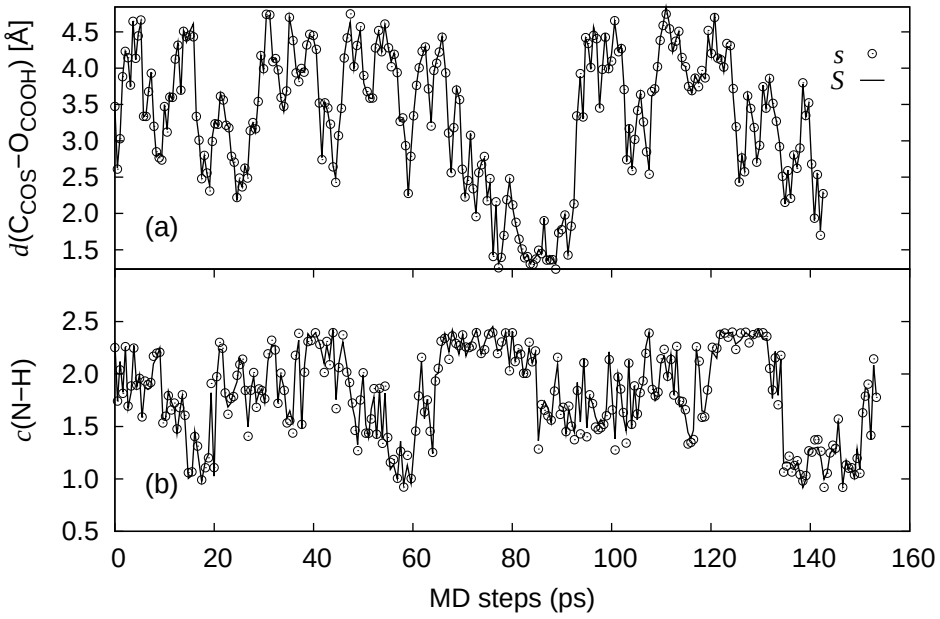
[20] Leung, K.; Rempe, S. B. *J. Chem. Phys.* **2005**, *122*, 184506–1–12.

[21] Slifkin, M. A.; Ali, S. M. *J. Mol. Liquids* **1984**, *28*, 215–221.

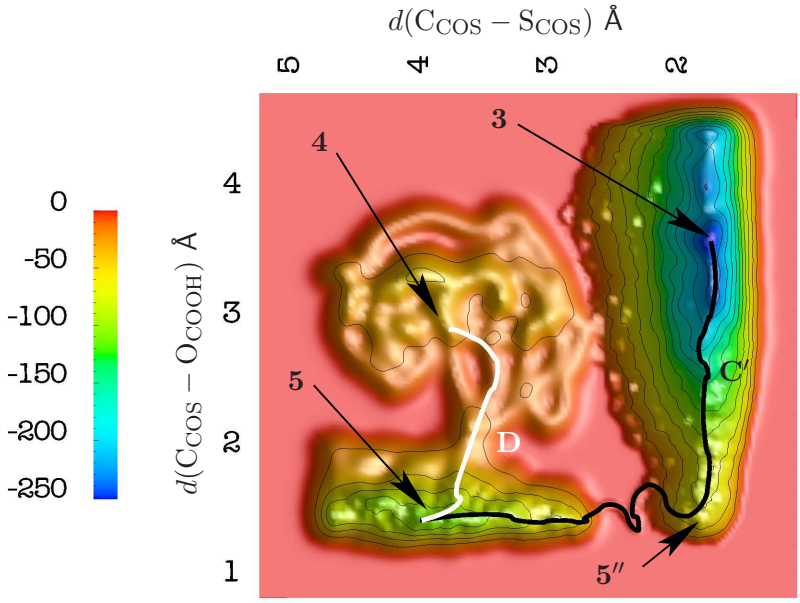
[22] Radzicka, A.; Wolfenden, R. *J. Am. Chem. Soc.* **1996**, *118*, 6105–6109.



SI Figure 1: Supercell (black box) used for the simulation of reactions at the interface of pyrite and hot–pressurized water (PIW) including periodic images in the [010] and [001] directions. Color code of the atoms: hydrogen (white), oxygen (red), carbon (green), nitrogen (blue), sulfur (yellow), iron (brown).



SI Figure 2: Collective coordinate S (solid lines) and corresponding auxiliary variable s (open circles) during the metadynamics simulations of (a) the ring closure of isocyanate to yield Leuchs anhydride in ABW and (b) the addition of COS and glycine to thiocarbamate in HPW. In (a) the distance $d(\mathrm{C_{COS}} - \mathrm{O_{COOH}})$ between the carbon in the COS part of the isocyanate and one of the carboxylate oxygens is shown and in (b) the coordination number $c(\mathrm{N-H})$ of the glycine nitrogen to all non–aliphatic hydrogen atoms in the system.



SI Figure 3: Reconstructed free energy surface in the space of the two collective coordinates $d(C_{COS} - S_{COS})$ and $d(C_{COS} - O_{COOH})$ for the conversion of thiocarbamate to Leuchs anhydride in ABW. The coordinate $d(C_{COS} - S_{COS})$ is the distance between the carbon and sulfur atoms in the COS group of the thiocarbamate **3**, and $d(C_{COS} - O_{COOH})$ is the distance between the carbon atom of the COS moiety to one of the carboxylate oxygen atoms of the thiocarbamate. Labeling of species: **3** corresponds to thiocarbamate, **4** is the isocyanate minimum, **5** denotes the Leuchs anhydride, and **5″** marks a cyclic intermediate structure. The schematic minimum free energy paths that lead to the mechanistic steps connecting these species, i.e. transformations **C′** and **D** in Figure 2 of the main article, are sketched in black and white, respectively, on the underlying free energy landscape. The color bar on the left side defines the relative free energy scale in kJ/mol.

A-B	R^0_{AB}
N-H	1.4
O-H	1.4
S-H	1.4
C-O	1.8

SI Table I: Distance cutoff parameters R^0_{AB} (Å) as used in equations (4.4) and (4.5) for different pairs of species A and B.

Reaction step	Collective coordinate A-B
A	$c\,[\mathrm{N}-\mathrm{H_{na}}]$,
	$c^{\mathrm{tot}}\,[\mathrm{O_{carboxyl}}-\mathrm{H_{na}}]$
B	$d\,[\mathrm{N_{Gly}}-\mathrm{C_{COS}}]$,
	$c\,[\mathrm{N_{Gly}}-\mathrm{H_{na}}]$
C′	$d\,[\mathrm{O_{COOH}}-\mathrm{C_{COSH}}]$,
	$d\,[\mathrm{S_{COSH}}-\mathrm{C_{COSH}}]$
C	$c\,[\mathrm{N}-\mathrm{H_{na}}]$,
	$c\,[\mathrm{S}-\mathrm{H_{na}}]$,
	$d\,[\mathrm{S}-\mathrm{C}]$
D	$d\,[\mathrm{C_{iso}}-\mathrm{O_{COOH}}]$,
	$c\,[\mathrm{N}-\mathrm{H_{na}}]$
E	$d\,[\mathrm{N_{Gly}}-\mathrm{C5_{NCA}}]$,
	$d\,[\mathrm{O1_{NCA}}-\mathrm{C5_{NCA}}]$,
	$c\,[\mathrm{N_{Gly}}-\mathrm{H_{na}}]$
F	$d\,[\mathrm{N}-\mathrm{C}]$,
	$c\,[\mathrm{N}-\mathrm{H_{na}}]$,
	$c^{\mathrm{tot}}\,[\mathrm{O}-\mathrm{H_{na}}]$
G	$d\,[\mathrm{N}-\mathrm{C}]$,
	$c\,[\mathrm{N_{pept}}-\mathrm{H_{na}}]$,
	$c\,[\mathrm{C_{pept}}-\mathrm{O_{solv}}]$

SI Table II: Sets of collective coordinates used for the metadynamics simulation of the individual reaction steps. The nomenclature of the reaction steps is according to Fig. 1 of the main text. The correponding functionals form of the collective coordinates are defined in equations (4.3), (4.4), and (4.5) together with the distance cutoff parameters provided in SI Table I; the associated metadynamics sampling parameters are given in SI Table III. The atomic symbols in square brackets belong to the molecule or functional group specified by the subscript: Gly (glycine), COS (carbonyl sulfide) COOH (carboxyl group), COSH (thiocarboxyl group), NCA (glycine *N*–carboxyanhydride), pept (peptide group), iso (isocyanic group), na (non–aliphatic), and solv (solvent).

Collective Coordinate	k_α (a.u.)	μ_α (a.m.u.)
$d(\mathrm{A}-\mathrm{B})$	0.4	50.0
$c(\mathrm{A}-\mathrm{B})$	2.0	50.0
$c^{\mathrm{tot}}(\mathrm{A}-\mathrm{B})$	2.0	50.0

SI Table III: Parameters k_α and μ_α used for different types of collective coordinates as defined in equations (4.3), (4.4), and (4.5), respectively.